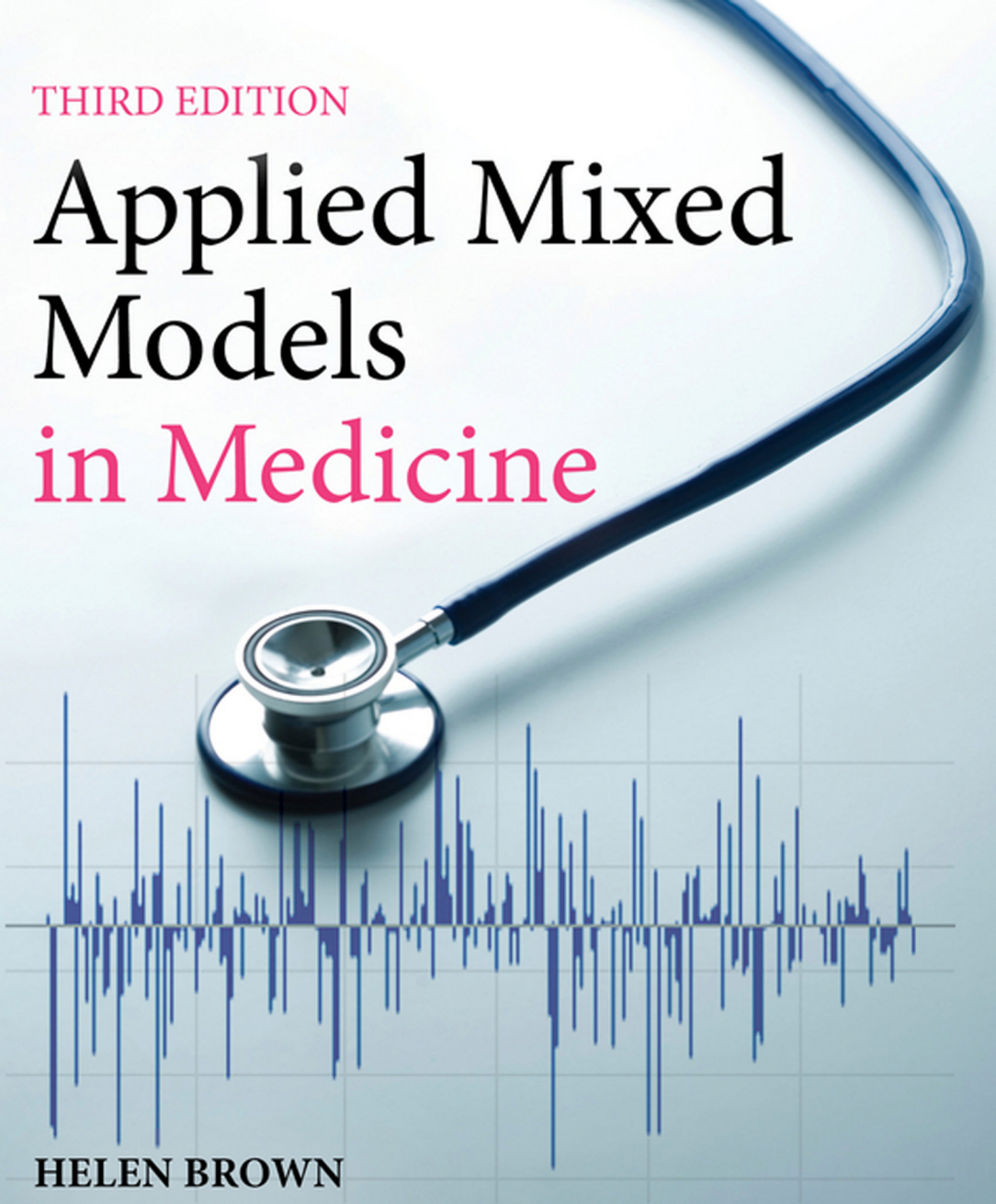
THIRD EDITION

Applied Mixed Models in Medicine



HELEN BROWN
ROBIN PRESCOTT

STATISTICS IN PRACTICE

WILEY

Applied Mixed Models in Medicine

STATISTICS IN PRACTICE

Series Advisors

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Applied Mixed Models in Medicine

Third Edition

Helen Brown
The Roslin Institute
University of Edinburgh, UK

Robin Prescott
Centre for Population Health Sciences
University of Edinburgh, UK

WILEY

This edition first published 2015

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Brown, Helen, 1962- , author.
Applied mixed models in medicine / Helen Brown, Robin Prescott. – Third edition.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-118-77825-8 (cloth)
I. Prescott, Robin, author. II. Title.
[DNLM: 1. Medicine. 2. Statistics as Topic–methods. 3. Models, Statistical. WA 950]
R853.S7
610.72′7–dc23

2014026811

A catalogue record for this book is available from the British Library.

ISBN: 9781118778258

Set in 10/12pt Photina by Laserwords Private Limited, Chennai, India

1 2015

Contents

Preface to third edition xiii

Mixed models notation xvii

About the Companion Website xix

1 Introduction 1
- 1.1 The use of mixed models 1
- 1.2 Introductory example 3
  - 1.2.1 Simple model to assess the effects of treatment (Model A) 4
  - 1.2.2 A model taking patient effects into account (Model B) 6
  - 1.2.3 Random effects model (Model C) 7
  - 1.2.4 Estimation (or prediction) of random effects 12
- 1.3 A multi-centre hypertension trial 12
  - 1.3.1 Modelling the data 13
  - 1.3.2 Including a baseline covariate (Model B) 15
  - 1.3.3 Modelling centre effects (Model C) 16
  - 1.3.4 Including centre-by-treatment interaction effects (Model D) 17
  - 1.3.5 Modelling centre and centre·treatment effects as random (Model E) 18
- 1.4 Repeated measures data 19
  - 1.4.1 Covariance pattern models 19
  - 1.4.2 Random coefficients models 21
- 1.5 More about mixed models 22
  - 1.5.1 What is a mixed model? 22
  - 1.5.2 Why use mixed models? 23
  - 1.5.3 Communicating results 24
  - 1.5.4 Mixed models in medicine 25
  - 1.5.5 Mixed models in perspective 25

1.6 Some useful definitions 27
1.6.1 Containment 28
1.6.2 Balance 29
1.6.3 Error strata 31

2 Normal mixed models 34
2.1 Model definition 34
2.1.1 The fixed effects model 35
2.1.2 The mixed model 37
2.1.3 The random effects model covariance structure 39
2.1.4 The random coefficients model covariance structure 42
2.1.5 The covariance pattern model covariance structure 44
2.2 Model fitting methods 46
2.2.1 The likelihood function and approaches to its maximisation 47
2.2.2 Estimation of fixed effects 50
2.2.3 Estimation (or prediction) of random effects and coefficients 51
2.2.4 Estimation of variance parameters 53
2.3 The Bayesian approach 57
2.3.1 Introduction 58
2.3.2 Determining the posterior density 59
2.3.3 Parameter estimation, probability intervals and *p*-values 60
2.3.4 Specifying non-informative prior distributions 62
2.3.5 Evaluating the posterior distribution 67
2.4 Practical application and interpretation 69
2.4.1 Negative variance components 69
2.4.2 Accuracy of variance parameters 73
2.4.3 Bias in fixed and random effects standard errors 74
2.4.4 Significance testing 75
2.4.5 Confidence intervals 78
2.4.6 Checking model assumptions 79
2.4.7 Missing data 81
2.4.8 Determining whether the simulated posterior distribution has converged 83
2.5 Example 84
2.5.1 Analysis models 84
2.5.2 Results 86
2.5.3 Discussion of points from Section 2.4 88

3 Generalised linear mixed models 113
3.1 Generalised linear models 114
3.1.1 Introduction 114

3.1.2 Distributions 115
3.1.3 The general form for exponential distributions 117
3.1.4 The GLM definition 118
3.1.5 Fitting the GLM 121
3.1.6 Expressing individual distributions in the general exponential form 123
3.1.7 Conditional logistic regression 125
3.2 Generalised linear mixed models 125
3.2.1 The GLMM definition 126
3.2.2 The likelihood and quasi-likelihood functions 127
3.2.3 Fitting the GLMM 129
3.3 Practical application and interpretation 133
3.3.1 Specifying binary data 133
3.3.2 Uniform effects categories 133
3.3.3 Negative variance components 134
3.3.4 Presentation of fixed and random effects estimates 135
3.3.5 Accuracy of variance parameters and potential bias 136
3.3.6 Bias in fixed and random effects standard errors 137
3.3.7 The dispersion parameter 139
3.3.8 Significance testing 140
3.3.9 Confidence intervals 141
3.3.10 Model checking 142
3.3.11 Determining whether the simulated posterior distribution has converged 142
3.4 Example 143
3.4.1 Introduction and models fitted 143
3.4.2 Results 144
3.4.3 Discussion of points from Section 3.3 148

4 Mixed models for categorical data 168
4.1 Ordinal logistic regression (fixed effects model) 168
4.2 Mixed ordinal logistic regression 173
4.2.1 Definition of the mixed ordinal logistic regression model 173
4.2.2 Residual variance matrix 174
4.2.3 Likelihood and quasi-likelihood functions 176
4.2.4 Model fitting methods 177
4.3 Mixed models for unordered categorical data 177
4.3.1 The G matrix 180
4.3.2 The R matrix 180
4.3.3 Fitting the model 180
4.4 Practical application and interpretation 180
4.4.1 Expressing fixed and random effects results 181
4.4.2 The proportional odds assumption 181

4.4.3 Number of covariance parameters 181
4.4.4 Choosing a covariance pattern 182
4.4.5 Interpreting covariance parameters 182
4.4.6 Checking model assumptions 182
4.4.7 The dispersion parameter 182
4.4.8 Other points 183
4.5 Example 183

5 Multi-centre trials and meta-analyses 197
5.1 Introduction to multi-centre trials 197
5.1.1 What is a multi-centre trial? 197
5.1.2 Why use mixed models to analyse multi-centre data? 198
5.2 The implications of using different analysis models 198
5.2.1 Centre and centre·treatment effects fixed 198
5.2.2 Centre effects fixed, centre·treatment effects omitted 199
5.2.3 Centre and centre·treatment effects random 200
5.2.4 Centre effects random, centre·treatment effects omitted 201
5.3 Example: a multi-centre trial 202
5.4 Practical application and interpretation 209
5.4.1 Plausibility of a centre·treatment interaction 209
5.4.2 Generalisation 210
5.4.3 Number of centres 210
5.4.4 Centre size 210
5.4.5 Negative variance components 211
5.4.6 Balance 211
5.5 Sample size estimation 211
5.5.1 Normal data 212
5.5.2 Binary data 215
5.5.3 Categorical data 217
5.6 Meta-analysis 217
5.7 Example: meta-analysis 218
5.7.1 Analyses 219
5.7.2 Results 220
5.7.3 Treatment estimates in individual trials 221

6 Repeated measures data 231
6.1 Introduction 231
6.1.1 Reasons for repeated measurements 231
6.1.2 Analysis objectives 232
6.1.3 Fixed effects approaches 232
6.1.4 Mixed models approaches 234
6.2 Covariance pattern models 234
6.2.1 Covariance patterns 235

6.2.2 Choice of covariance pattern 239
6.2.3 Choice of fixed effects 241
6.2.4 General points 242
6.3 Example: covariance pattern models for normal data 244
6.3.1 Analysis models 245
6.3.2 Selection of covariance pattern 245
6.3.3 Assessing fixed effects 247
6.3.4 Model checking 248
6.4 Example: covariance pattern models for count data 254
6.4.1 Analysis models 255
6.4.2 Analysis using a categorical mixed model 258
6.5 Random coefficients models 262
6.5.1 Introduction 262
6.5.2 General points 264
6.5.3 Comparisons with fixed effects approaches 266
6.6 Examples of random coefficients models 267
6.6.1 A linear random coefficients model 267
6.6.2 A polynomial random coefficients model 270
6.7 Sample size estimation 286
6.7.1 Normal data 286
6.7.2 Binary data 288
6.7.3 Categorical data 288

7 Cross-over trials 289
7.1 Introduction 289
7.2 Advantages of mixed models in cross-over trials 290
7.3 The AB/BA cross-over trial 290
7.3.1 Example: AB/BA cross-over design 292
7.4 Higher order complete block designs 297
7.4.1 Inclusion of carry-over effects 297
7.4.2 Example: four-period, four-treatment cross-over trial 298
7.5 Incomplete block designs 302
7.5.1 Example: Three treatment two-period cross-over trial 303
7.6 Optimal designs 305
7.6.1 Example: Balaam's design 305
7.7 Covariance pattern models 307
7.7.1 Structured by period 308
7.7.2 Structured by treatment 308
7.7.3 Example: four-way cross-over trial 308
7.8 Analysis of binary data 317
7.9 Analysis of categorical data 321
7.10 Use of results from random effects models in trial design 325
7.10.1 Example 325
7.11 General points 326

8 Other applications of mixed models 329
8.1 Trials with repeated measurements within visits 329
8.1.1 Covariance pattern models 330
8.1.2 Example 335
8.1.3 Random coefficients models 341
8.1.4 Example: random coefficients models 343
8.2 Multi-centre trials with repeated measurements 349
8.2.1 Example: multi-centre hypertension trial 349
8.2.2 Covariance pattern models 350
8.3 Multi-centre cross-over trials 355
8.4 Hierarchical multi-centre trials and meta-analysis 356
8.5 Matched case–control studies 357
8.5.1 Example 358
8.5.2 Analysis of a quantitative variable 359
8.5.3 Check of model assumptions 360
8.5.4 Analysis of binary Variables 362
8.6 Different variances for treatment groups in a simple between-patient trial 370
8.6.1 Example 371
8.7 Estimating variance components in an animal physiology trial 374
8.7.1 Sample size estimation for a future experiment 375
8.8 Inter- and intra-observer variation in foetal scan measurements 380
8.9 Components of variation and mean estimates in a cardiology experiment 381
8.10 Cluster sample surveys 384
8.10.1 Example: cluster sample survey 384
8.11 Small area mortality estimates 386
8.12 Estimating surgeon performance 389
8.13 Event history analysis 391
8.13.1 Example 392
8.14 A laboratory study using a within-subject 4×4 factorial design 394
8.15 Bioequivalence studies with replicate cross-over designs 397
8.15.1 Example 402
8.16 Cluster randomised trials 411
8.16.1 Example: A trial to evaluate integrated care pathways for treatment of children with asthma in hospital 412
8.16.2 Example: Edinburgh randomised trial of breast screening 414
8.17 Analysis of bilateral data 418

8.18 Incomplete block designs 425
8.18.1 Introduction 425
8.18.2 Balanced incomplete block (BIB) designs 426
8.18.3 Studies where only comparisons to a particular treatment are of interest 429
8.18.4 Designs to produce lower variances for a specific treatment pair 433
8.18.5 Design to produce lower variances for more than one treatment pair 438

9 Software for fitting mixed models 452
9.1 Packages for fitting mixed models 452
9.2 `PROC MIXED` 454
9.3 Using `SAS` to fit mixed models to non-normal data 476
9.3.1 `PROC GLIMMIX` 477
9.3.2 `PROC GENMOD` 480
9.4 `PROC MCMC` 483

Glossary **489**

References **492**

Index **497**

SIP Series Titles **513**

Preface to third edition

Analysis of variance and regression has for many years been the mainstay of statistical modelling. These techniques usually have as a basic assumption that the residual or error terms are independently and identically distributed. Mixed models are an important approach to modelling, which allows us to relax the independence assumption and take into account more complicated data structures in a flexible way. Sometimes, this interdependence of observations is modelled directly in a mixed model. For example, if a number of repeated measurements are made on a patient, then mixed models allow us to specify a pattern for the correlation between these measurements. In other contexts, such as the cross-over clinical trial, specifying that patient effects are normally distributed, rather than fixed as in the classical approach, induces observations on the same patient to be correlated.

There are many benefits to be gained from using mixed models. In some situations, the benefit will be an increase in the precision of our estimates. In others, we will be able to make wider inferences. We will sometimes be able to use a more appropriate model that will give us greater insight into what underpins the structure of the data. However, it is only the availability of software in versatile packages such as SAS® that has made these techniques widely accessible. It is now important that suitable information on their use becomes available so that they may be applied confidently on a routine basis.

Our intention in this book is to put all types of mixed models into a general framework and to consider the practical implications of their use. We aim to do this at a level that can be understood by applied statisticians and numerate scientists. Greatest emphasis is placed on skills required for the application of mixed models and interpretation of the results. An in-depth understanding of the mathematical theory underlying mixed models is not essential to gain these, but an awareness of the practical consequences of fitting different types of mixed models is necessary. While many publications are available on various aspects of mixed models, these generally relate to specific types of model and often differ in their use of terminology. Such publications are not always readily comprehensible

to the applied statisticians who will be the most frequent users of the methods. An objective of this book is to help overcome this deficit.

Examples given will primarily relate to the medical field. However, the general concepts of mixed models apply equally to many other areas of application, for example, social sciences, agriculture, veterinary science and official statistics. (In the social sciences, mixed models are often referred to as 'multi-level' models.) Data are becoming easier to collect, with the consequence that datasets are now often large and complex. We believe that mixed models provide useful tools for modelling the complex structures that occur in such data.

The third edition of this book retains the structure of the first two, but there are further changes to reflect the continued evolution of SAS. This edition fully incorporates features of SAS up to version 9.3. Compared to what was available at the time of the previous edition, enhancements to SAS include improved graphical facilities. Importantly, there is also a new procedure, PROC MCMC, which facilitates Bayesian analysis. This has led to extensive changes in our coverage of Bayesian methods. SAS 9.3 and later versions now provide output both in text format from the output window and, additionally, as an HTML file in the results viewer. There have been accompanying minor changes in the details of outputs and graphs, such as labelling. Our approach to reporting SAS outputs in this edition has been to change our presentation from earlier editions only when we wish to highlight features that have changed substantially and, importantly, to facilitate the reader's use of mixed models, whatever their version of SAS.

During the drafting of this edition, SAS 9.4 became available. It is not fully incorporated into this book because its new features are focused more on the SAS high performance procedures than on improvements to the SAS/STAT procedures. These high performance procedures 'provide predictive modelling tools that have been specially developed to take advantage of parallel processing in both multithread single-machine mode and distributed multi-machine mode'. Typically, the high performance procedures such as PROC HPLMIXED have a greatly reduced range of options compared to PROC MIXED and, consequently, are peripheral to the aims of this book. We do, however, consider some of the small modifications to improve procedures such as GLIMMIX and MCMC that are available in SAS/STAT® 12.1 and later versions.

Chapter 1 provides an introduction to the capabilities of mixed models, defines general concepts and gives their basic statistical properties. Chapter 2 defines models and fitting methods for normally distributed data. Chapter 3 first introduces generalised linear models that can be used for the analysis of data that are binomial or Poisson or from any other member of the exponential family of distributions. These methods are then extended to incorporate mixed models concepts under the heading of generalised linear mixed models. The fourth chapter examines how mixed models can be applied when the variable to be analysed is categorical. The main emphasis in these chapters, and indeed in the whole book, is on classical statistical approaches to inference, based on significance tests and confidence intervals. However, the Bayesian approach is

also introduced in Chapter 2, since it has several potential advantages and its use is becoming more widespread. Although the overall emphasis of the book is on the application of mixed models techniques, these chapters can also be used as a reference guide to the underlying theory of mixed models.

Chapters 5–7 consider the practical implications of using mixed models for particular designs. Each design illustrates a different feature of mixed models.

Multi-centre trials and meta-analyses are considered in Chapter 5. These are examples of hierarchical data structures, and the use of a mixed model allows for any additional variation in treatment effects occurring between centres (or trials) and hence makes results more generalisable. The methods shown can be applied equally to any type of hierarchical data.

In Chapter 6, the uses of covariance pattern models and random coefficients models are described using the repeated measures design. These approaches take into account the correlated nature of the repeated observations and give more appropriate treatment effect estimates and standard errors. The material in this chapter will apply equally to any situation where repeated observations are made on the same units.

Chapter 7 considers cross-over designs where each patient may receive several treatments. In this design, more accurate treatment estimates are often achieved by fitting patient effects as random. This improvement in efficiency can occur for any dataset where a fixed effect is 'crossed' with a random effect.

In Chapter 8, a variety of other designs and data structures is considered. These either incorporate several of the design aspects covered in Chapters 5–7 or have structures that have arisen in a more unplanned manner. They help to illustrate the broad scope of application of mixed models. This chapter includes two new sections. We have added a section on the analysis of bilateral data, a common structure in some areas of medical research, but one that we had not previously addressed. There is also a substantial new section on incomplete block designs.

Chapter 9 gives information on software available for fitting mixed models. Most of the analyses in the book are carried out using `PROC MIXED` in `SAS`, supplemented by `PROC GENMOD`, `PROC GLIMMIX`, and `PROC MCMC`. This chapter introduces the basic syntax for these procedures. This information should be sufficient for fitting most of the analyses described, but the full `SAS` documentation should be referenced for those who wish to use more complex features. The `SAS` code used for most of the examples is supplied within the text. In addition, the example datasets and `SAS` code may be obtained electronically from www.wiley.com/go/brown/applied_mixed.

This book has been written to provide the reader with a thorough understanding of the concepts of mixed models, and we trust it will serve well for this purpose. However, readers wishing to take a shortcut to the fitting of normal mixed models should read Chapter 1 for an introduction, Section 2.4 for practical details, and the chapter relevant to their design. To fit non-normal or categorical mixed models, Section 3.3 or Section 4.4 should be read in addition to Section 2.4. In an attempt to make this book easier to use, we have presented at the beginning of

the text a summary of the notation we have used, while at the end, we list some key definitions in a glossary.

Our writing of this book has been aided in many ways. The first edition evolved from a constantly changing set of course notes that accompanied a 3-day course on the subject, run regularly over the previous 6 years. The second edition was helped by many individuals who were kind enough to comment on the first edition, including the identification of some errors that had slipped in, and by further participants at our courses who have contributed to discussions and have thereby helped to shape our views. This process has continued with the third edition. We are also grateful to many other colleagues who have read and commented on various sections of the manuscript and especially to our colleagues who have allowed us to use their data. We hope that readers will find the resulting book a useful reference in an interesting and expanding area of statistics.

Helen Brown
Robin Prescott
Edinburgh

Mixed models notation

The notation below is provided for quick reference. Models are defined more fully in Sections 2.1, 3.1 and 4.1.

Normal mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e},$$

$$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}),$$

$$\mathrm{var}(\mathbf{e}) = \mathbf{R},$$

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

Generalised linear mixed model

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e},$$

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta},$$

$$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}),$$

$$\mathrm{var}(\mathbf{e}) = \mathbf{R},$$

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathrm{var}(\boldsymbol{\mu}) + \mathbf{R},$$

$$\approx \mathbf{B}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{B} + \mathbf{R} \quad \text{(a first-order approximation)},$$

where

$\mathbf{y}$ = dependent variable,
$\mathbf{e}$ = residual error,
$\mathbf{X}$ = design matrix for fixed effects,
$\mathbf{Z}$ = design matrix for random effects,
$\boldsymbol{\alpha}$ = fixed effects parameters,
$\boldsymbol{\beta}$ = random effects parameters,
$\mathbf{R}$ = residual variance matrix,
$\mathbf{G}$ = matrix of covariance parameters,
$\mathbf{V}$ = var($\mathbf{y}$) variance matrix,

$\boldsymbol{\mu}$ = expected values,
g = link function,
$\mathbf{B}$ = diagonal matrix of variance terms (e.g. $\mathbf{B} = \text{diag}\{\mu_i(1-\mu_i)\}$ for binary data).

Ordered categorical mixed model

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e},$$

$$\text{logit}(\boldsymbol{\mu}^{[c]}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta},$$

$$\boldsymbol{\beta} \sim \text{N}(\mathbf{0}, \mathbf{G}),$$

var($\mathbf{y}$) is defined as in the GLMM,

where

$$\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}, \ \dots\ , \mu_{n1}, \mu_{n2}, \mu_{n3})',$$
μ_{ij} = probability observation i is in category j,
$$\boldsymbol{\mu}^{[c]} = (\mu_{11}^{[c]}, \mu_{12}^{[c]}, \mu_{13}^{[c]}, \mu_{21}^{[c]}, \mu_{22}^{[c]}, \mu_{23}^{[c]}, \ \dots\ , \mu_{n1}^{[c]}, \mu_{n2}^{[c]}, \mu_{n3}^{[c]})',$$
$$\mu_{ij}^{[c]} = \text{probability}\ \ (y_i \leq j) = \sum_{k=1}^{j} \mu_{ik}.$$

About the Companion Website

This book is accompanied by a companion website:

www.wiley.com/go/brown/applied_mixed

This website includes SAS codes and datasets for most of the examples. In the future, updates and further materials may be added.

1

Introduction

At the start of each chapter, we will 'set the scene' by outlining its content. In this introductory chapter, we start Section 1.1 by describing some situations where a mixed models analysis will be particularly helpful. In Section 1.2, we describe a simplified example and use it to illustrate the idea of a statistical model. We then introduce and compare fixed effects and random effects models. In the next section, we consider a more complex 'real-life' multi-centre trial and look at some of the variety of models that could be fitted (Section 1.3). This example will be used for several illustrative examples throughout the book. In Section 1.4, the use of mixed models to analyse a series of observations (repeated measures) is considered. Section 1.5 broadens the discussion on mixed models and looks at mixed models with a historical perspective of their use. In Section 1.6, we introduce some technical concepts: containment, balance and error strata.

We will assume in our presentation that the reader is already familiar with some of the basic statistical concepts as found in elementary statistical textbooks.

1.1 The use of mixed models

In the course of this book, we will encounter many situations in which a mixed models approach has advantages over the conventional type of analysis, which would be accessible via introductory texts on statistical analysis. Some of them are introduced in outline in this chapter and will be dealt in detail later on.

Example 1: Utilisation of incomplete information in a cross-over trial Cross-over trials are often utilised to assess treatment efficacy in chronic conditions, such as asthma. In such conditions, an individual patient can be tested for response to a succession of two or more treatments, giving the benefit of a 'within-patient' comparison. In the most commonly used cross-over design, just two treatments

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

are compared. If, for generality, we call these treatments A and B, then patients will be assessed either on their response to treatment A, followed by their response to treatment B, or vice versa. If all patients complete the trial, and both treatments are assessed, then the analysis is fairly straightforward. However, commonly, patients drop out during the trial and may have a valid observation from only the first treatment period. These incomplete observations cannot be utilised in a conventional analysis. In contrast, the use of a mixed model will allow all of the observations to be analysed, resulting in more accurate comparisons of the efficacy of treatment. This benefit, of more efficient use of the data, applies to all types of cross-over trial where there are missing data.

Example 2: Cross-over trials with fewer treatment periods than treatments In cross-over trials, for logistical reasons, it may be impractical to ask a patient to evaluate more than two treatments (e.g. if the treatment has to be given for several weeks). Nevertheless, there may be the need to evaluate three or more treatments. Special types of cross-over design can be used in this situation, but a simple analysis will be very inefficient. Mixed models provide a straightforward method of analysis, which fully uses the data, resulting again in more precise estimates of the effect of the treatments.

Example 3: A surgical audit A surgical audit is to be carried out to investigate how different hospitals compare in their rates of postoperative complications following a particular operation. As some hospitals carry out the operation commonly, while other hospitals perform the operation rarely, the accuracy with which the complication rates are estimated will vary considerably from hospital to hospital. Consequently, if the hospitals are ordered according to their complication rates, some may appear to be outliers compared with other hospitals, purely due to chance variation. When mixed models are used to analyse data of this type, the estimates of the complication rates are adjusted to allow for the number of operations, and rates based on small numbers become less extreme.

Example 4: Analysis of a multi-centre trial Many clinical trials are organised on a multi-centre basis, usually because there is an inadequate number of suitable patients in any single centre. The analysis of multi-centre trials often ignores the centres from which the data were obtained, making the implicit assumption that all centres are identical to one another. This assumption may sometimes be dangerously misleading. For example, a multi-centre trial comparing two surgical treatments for a condition could be expected to show major differences between centres. There could be two types of differences. First, the centres may differ in the overall success, averaged over the two surgical treatments. More importantly, there may be substantial differences in the relative benefit of the two treatments across different centres. Surgeons who have had more experience with one operation (A) may produce better outcomes with A, while surgeons with more experience with the alternative operation (B) may obtain better results

with B. Mixed models can provide an insightful analysis of such a trial by allowing for the extent to which treatment effects differ from centre to centre. Even when the difference between treatments can be assumed to be identical in all centres, a mixed model can improve the precision of the treatment estimates by taking appropriate account of the centres in the analysis.

Example 5: Repeated measurements over time In a clinical trial, the response to treatment is often assessed as a series of observations over time. For example, in a trial to assess the effect of a drug in reducing blood pressure, measurements might be taken at two, four, six and eight weeks after starting treatment. The analysis will usually be complicated by a number of patients failing to appear for some assessments or withdrawing from the study before it is complete. This complication can cause considerable difficulty in a conventional analysis. A mixed models analysis of such a study does not require complete data from all subjects. This results in more appropriate estimates of the effect of treatment and their standard errors (SEs). The mixed model also gives great flexibility in analysis, in that it can allow for a wide variety of ways in which the successive observations are correlated with one another.

1.2 Introductory example

We consider a very simple cross-over trial using artificial data. In this trial, each patient receives each of treatments A and B for a fixed period. At the end of each treatment period, a measurement is taken to assess the response to that treatment. In the analysis of such a trial, we commonly refer to treatments being *crossed* with patients, meaning that the categories of 'treatments' occur in combination with those of 'patients'. For the purpose of this illustration, we will suppose that the response to each treatment is unaffected by whether it is received in the first or second period. The table shows the results from the six patients in this trial.

	Treatment		Difference	
Patient	**A**	**B**	**A − B**	**Patient mean**
1	20	12	8	16.0
2	26	24	2	25.0
3	16	17	−1	16.5
4	29	21	8	25.0
5	22	21	1	21.5
6	24	17	7	20.5
Mean	22.83	18.67	4.17	20.75

1.2.1 Simple model to assess the effects of treatment (Model A)

We introduce in this section a very simple example of a statistical model using this data. A model can be thought of as an attempt to describe quantitatively the effect of a number of factors on each observation. Any model we describe is likely to be a gross oversimplification of reality. In developing models, we are seeking ones which are as simple as possible but which contain enough truth to ask questions of interest. In this first simple model, we will deliberately be oversimplistic in order to introduce our notation. We just describe the effect of the two treatments. The model may be expressed as

$$y_{ij} = \mu + t_j + e_{ij},$$

where

$j =$ A or B,
$y_{ij} =$ observation for treatment j on the ith patient,
$\mu =$ overall mean,
$t_j =$ effect of treatment j,
$e_{ij} =$ error for treatment j on the ith patient.

The constant μ represents the overall mean of the observations. $\mu + t_A$ corresponds to the mean in the treatment group A, while $\mu + t_B$ corresponds to the mean in the treatment group B. The constants μ, t_A and t_B can thus be estimated from the data. In our example, we can estimate the value of μ to be 20.75, the overall mean value. From the mean value in the first treatment group, we can estimate $\mu + t_A$ as 22.83, and hence our estimate of t_A is $22.83 - 20.75 = 2.08$. Similarly, from the mean of the second treatment group, we estimate t_B as -2.08. The term t_j can therefore be thought of as a measure of the relative effect that treatment j has had on our outcome variable.

The error term, e_{ij}, or *residual* is what remains for each patient in each period when $\mu + t_j$ is deducted from their observed measurement. This represents random variation about the mean value for each treatment. As such, the residuals can be regarded as the result of drawing random samples from a distribution. We will assume that the distribution is Gaussian or normal, with standard deviation σ, and that the samples drawn from the distribution are independent of each other. The mean of the distribution can be taken as zero, since any other value would simply cause a corresponding change in the value of μ. Thus, we will write this as

$$e_{ij} \sim \mathrm{N}(0, \sigma^2),$$

where σ^2 is the variance of the residuals. In practice, checks should be made to determine whether this assumption of normally distributed residuals is reasonable. Suitable checking methods will be considered in Section 2.4.6. As individual

observations are modelled as the sum of $\mu + t_j$, which are both constants, and the residual term, it follows that the variance of individual observations equals the residual variance:

$$\text{var}(y_{ij}) = \sigma^2.$$

The covariance of any two separate observations y_{ij} and $y_{i'j'}$ can be written as

$$\begin{aligned}\text{cov}(y_{ij}, y_{i'j'}) &= \text{cov}(\mu + t_i + e_{ij}, \quad \mu + t_{i'} + e_{i'j'}) \\ &= \text{cov}(e_{ij}, e_{i'j'}) \text{ (since other terms are constants).}\end{aligned}$$

Since all the residuals are assumed independent (i.e. uncorrelated), it follows that

$$\text{cov}(y_{ij}, y_{i'j'}) = 0.$$

The residual variance, σ^2, can be estimated using a standard technique known as analysis of variance (ANOVA). The essence of the method is that the total variation in the data is decomposed into components that are associated with possible causes of this variation, for example, that one treatment may be associated with higher observations, with the other being associated with lower observations. For this first model, using this technique, we obtain the following ANOVA table:

Source of variation	**Degrees of freedom**	**Sums of squares**	**Mean square**	***F***	***p***
Treatments	1	52.08	52.08	2.68	0.13
Residual	10	194.17	19.42		

Note: F = value for the F test (ratio of mean square for treatments to mean square for residual).
p = significance level corresponding to the F test.

The residual mean square of 19.42 is our estimate of the residual variance, σ^2, for this model. The key question often arising from this type of study is as follows: 'do the treatment effects differ significantly from each other?' This can be assessed by the F test, which assesses the null hypothesis of no mean difference between the treatments (the larger the treatment difference, the larger the treatment mean square and the higher the value of F). The p value of 0.13 is greater than the conventionally used cutoff point for statistical significance of 0.05. Therefore, we cannot conclude that the treatment effects are significantly different. The difference between the treatment effects and the SE of this difference provides a measure of the size of the treatment difference and the accuracy with which it is estimated:

$$\text{difference} = t_A - t_B = 2.08 + 2.08 = 4.16.$$

The SE of the difference is given by the formula

$$\begin{aligned}\mathrm{SE}(t_A - t_B) &= \sqrt{\sigma^2(1/n_A + 1/n_B)} \\ &= \sqrt{(2 \times \sigma^2/6)} = \sqrt{6.47} = 2.54.\end{aligned}$$

Note that a t test can also be constructed from this difference and SE, giving $t = 4.16/2.54 = 1.63$. This is the square root of our F statistic of 2.68 and gives an identical t test p value of 0.13.

1.2.2 A model taking patient effects into account (Model B)

Model A as discussed previously did not utilise the fact that pairs of observations were taken on the same patients. It is possible, and indeed likely, that some patients will tend to have systematically higher measurements than others, and we may be able to improve the model by making allowance for this. This can be done by additionally including patient effects into the model:

$$y_{ij} = \mu + p_i + t_j + e_{ij},$$

where p_i are constants representing the ith patient effect.

The ANOVA table arising from this model is as follows:

Source of variation	Degrees of freedom	Sums of squares	Mean square	*F*	*p*
Patients	5	154.75	30.95	3.93	0.08
Treatments	1	52.08	52.08	6.61	0.05
Residual	5	39.42	7.88		

The estimate of the residual variance, σ^2, is now 7.88. It is lower than in Model A because it represents the 'within-patient' variation, as we have taken account of patient effects. The F test p value of 0.05 indicates that the treatment effects are now significantly different. The difference between the treatment effects is the same as in Model A, 4.16, but its SE is now as follows:

$$\mathrm{SE}(t_A - t_B) = \sqrt{(2 \times \sigma^2/6)} = \sqrt{2.63} = 1.62.$$

(Note that the SE of the treatment difference could alternatively have been obtained directly from the differences in patient observations.)

Model B is perhaps the 'obvious' one to think of for this dataset. However, even in this simple case, by comparison with Model A we can see that the statistical

modeller has some flexibility in his/her choice of model. In most situations, there is no single 'correct' model, and, in fact, models are rarely completely adequate. The job of the statistical modeller is to choose that model which most closely achieves the objectives of the study.

1.2.3 Random effects model (Model C)

In the Models A and B, the only assumption we made about variation was that the residuals were normally distributed. We did not assume that patient or treatment effects arose from a distribution. They were assumed to take constant values. These models can be described as *fixed effects models*, and all effects fitted within them are *fixed effects*.

An alternative approach available to us is to assume that some of the terms in the model, instead of taking constant values, are realisations of values from a probability distribution. If we assumed that patient effects also arose from independent samples from a normal distribution, then the model could be expressed as

$$\begin{aligned} y_{ij} &= \mu + p_i + t_j + e_{ij}, \\ e_{ij} &\sim \mathrm{N}(0, \sigma^2) \\ p_i &\sim \mathrm{N}(0, \sigma_\mathrm{p}^2). \end{aligned}$$

The p_i are now referred to as *random effects*. Such models, which contain a mixture of fixed and random effects, provide an example of a *mixed model*. In this book, we will meet several different types of mixed model, and we describe in Section 1.5 the common feature that distinguishes them from fixed effects models. To distinguish the class of models we have just met from those we will meet later, we will refer to this type of model as a *random effects model*.

Each random effect in the model gives rise to a *variance component*. This is a model parameter that quantifies random variation due to that effect only. In this model, the patient variance component is σ_p^2. We can describe variation at this level (between patients) as occurring within the patient *error stratum* (see Section 1.6 for a full description of the error stratum). This random variation occurs in addition to the residual variation (the residual variance can also be defined as a variance component.)

Defining the model in this way causes some differences in its statistical properties compared with the fixed effects model met earlier.

The variance of individual observations in a random effects model is the sum of all the variance components. Thus,

$$\mathrm{var}(y_{ij}) = \sigma_\mathrm{p}^2 + \sigma^2.$$

This contrasts with the fixed effects models where we had

$$\text{var}(y_{ij}) = \sigma^2.$$

The effect on the covariance of pairs of observations in the random effects model is interesting and perhaps surprising. Since $y_{ij} = \mu + p_i + t_j + e_{ij}$, we can write

$$\begin{aligned}\text{cov}(y_{ij}, y_{i'j'}) &= \text{cov}(\mu + p_i + t_j + e_{ij}, \mu + p_{i'} + t_{j'} + e_{i'j'}) \\ &= \text{cov}(p_i + e_{ij}, p_{i'} + e_{i'j'}).\end{aligned}$$

When observations from different patients are being considered (i.e. $i \neq i'$), because of the independence of the observations, $\text{cov}(y_{ij}, y_{i'j'}) = 0$. However, when two samples from the same patient are considered (i.e. $i = i'$), then

$$\begin{aligned}\text{cov}(y_{ij}, y_{i'j'}) &= \text{cov}(p_i + e_{ij}, p_i + e_{ij'}) \\ &= \text{cov}(p_i, p_i) = \sigma_{\text{p}}^2.\end{aligned}$$

Thus, observations on the same patient are correlated and have covariance equal to the patient variance component, while observations on different patients are uncorrelated. This contrasts with the fixed effects models where the covariance of any pair of observations is zero.

The ANOVA table for the random effects model is identical to that for the fixed effects model. However, we can now use it to calculate the patient variance component using results from the statistical theory that underpins the ANOVA method. The theory shows the expected values for each of the mean square terms in the ANOVA table, in terms of σ^2, σ_{p}^2 and the treatment effects. These are tabulated in the following table. We can now equate the expected value for the mean squares expressed in terms of the variance components to the observed values of the mean squares to obtain estimates of σ^2 and σ_{p}^2.

Source of variation	Degrees of freedom	Sums of squares	Mean square	E(MS)
Patients	5	154.75	30.95	$2\sigma_{\text{p}}^2 + \sigma^2$
Treatments	1	52.08	52.08	$\sigma^2 + 6\Sigma t_i^2$
Residual	5	39.42	7.88	σ^2

Note: E(MS) = expected mean square.

Thus, from the residual line in the ANOVA table, $\hat{\sigma}^2 = 7.88$. In addition, by subtracting the third line of the table from the first we have:

$$2\hat{\sigma}_{\text{p}}^2 = (30.95 - 7.88), \quad \text{and} \quad \hat{\sigma}_{\text{p}}^2 = 11.54.$$

(We are introducing the notation $\hat{\sigma}_{\text{p}}^2$ to denote that this is an estimate of the unknown σ_{p}^2, and $\hat{\sigma}^2$ is an estimate of σ^2.)

In this example, we obtain identical treatment effect results to those for the fixed effects model (Model B). This occurs because we are, in effect, only using within-patient information to estimate the treatment effect (since all information on treatment occurs in the within-patient residual error stratum). Again, we obtain the treatment difference as −4.16 with a SE of 1.62. Thus, in this case, it makes no difference at all to our conclusions about treatments whether we fit patient effects as fixed or random. However, had any of the values in the dataset been missing, this would not have been the case. We now consider this situation.

Dataset with missing values

We will now consider analysing the dataset with two of the observations set to missing.

	Treatment		Difference	
Patient	**A**	**B**	**A − B**	**Patient mean**
1	20	12	8	16.0
2	26	24	2	25.0
3	16	17	−1	16.5
4	29	21	8	25.0
5	22	–	–	22.0
6	–	17	–	17.0
Mean			4.25	

As shown previously, there are two ways we can analyse the data. We can base our analysis on a model where the patient effects are regarded as fixed (Model B) or can regard patient effects as random (Model C).

The fixed effects model For this analysis, we apply ANOVA in the standard way, and the result of that analysis is summarised as follows:

Source of variation	**Degrees of freedom**	**Sums of squares**	**Mean square**	***F***	***p***
Patients	5	167.90	33.58	3.32	0.18
Treatments	1	36.13	36.13	3.57	0.16
Residual	3	30.38	10.12		

In the fitting of Model B, it is interesting to look at the contribution that the data from patient 5 are making to the analysis. The value of 22 gives us information that will allow us to estimate the level in that patient, but it tells us nothing at all

about the difference between the two treatments, nor does it even tell us anything about the effect of treatment A, which was received, because all the information in the observed value of 22 is used up in estimating the patient effect. The same comment applies to the data from patient 6.

Thus, in this fixed effects model, the estimate of the mean treatment difference, $\widehat{t}_{\mathrm{FE}}$, will be calculated only from the treatment differences for patients 1–4 who have complete data:

$$\widehat{t}_{\mathrm{FE}} = 4.25.$$

The variance of $\widehat{t}_{\mathrm{FE}}$ can be calculated from the residual variance, $\widehat{\sigma}^2 = 10.12$, as

$$\mathrm{var}(\widehat{t}_{\mathrm{FE}}) = \widehat{\sigma}^2(1/n_{\mathrm{p}} + 1/n_{\mathrm{p}}) = 10.12 \times (1/4 + 1/4) = 5.06,$$

where n_{p} is the number of observations with data on treatments A and B. The SE of the treatment difference is $\sqrt{5.06} = 2.25$.

The random effects model When patient effects are fitted as random, the variance components cannot be derived in a straightforward way from an ANOVA table since the data are unbalanced. They are found computationally (using `PROC MIXED`, a `SAS` procedure, which is described in more detail in Chapter 9) as

$$\widehat{\sigma}^2_{\mathrm{p}} = 12.63,$$
$$\widehat{\sigma}^2 = 8.90.$$

The treatment difference is estimated from the model to be 4.32, with a SE of 2.01. Thus, the SE is smaller than that of 2.25 obtained in the fixed effects model. This is not only due to a fortuitously lower estimate of σ^2, but also due to the fact that the random effects model utilises information on treatment from both the patient error stratum (between patients) and the residual stratum (within patients). As noted previously, the SE of the estimates is less than that in the fixed effects model, which only uses information from within patients. The use of this extra information compared with the fixed effects model can be referred to as the *recovery* of between-patient information.

In practice, we would recommend that random effects models are always fitted computationally using a procedure such as `PROC MIXED`. However, in our simple example given in this chapter, it may be of help to the understanding of the concept of recovery of information if we illustrate how the treatment estimates can be obtained manually.

Manual calculation In this example, the estimate of the treatment difference for the random effects model may be obtained by combining estimates from the between-patient and within-patient (residual) error strata. It is calculated by a weighted average of the two estimates, with the inverses of the variances of

the estimates used as weights. The within-patient estimate, $\widehat{t}_W$, is obtained as in the fixed effects model from patients 1–4 as 4.25. However, its variance is now calculated from the new estimate of σ^2 as

$$\text{var}(\widehat{t}_W) = \sigma^2(1/n_p + 1/n_p) = 8.90 \times (1/4 + 1/4) = 4.45.$$

The between-patient estimate, $\widehat{t}_B$, is simply the difference between the single values for patients 5 and 6

$$\widehat{t}_B = 22 - 17 = 5$$

and has variance as

$$\text{var}(\widehat{t}_B) = (\sigma^2 + \sigma_p^2) \times (1/1 + 1/1) = (8.90 + 12.63) \times 2 = 43.06.$$

The combined random effects model estimate, $\widehat{t}_{RE}$, is obtained as a weighted average of $\widehat{t}_W$ and $\widehat{t}_B$:

$$\widehat{t}_{RE} = K \times (\widehat{t}_W/\text{var}(\widehat{t}_W) + \widehat{t}_B/\text{var}(\widehat{t}_B)),$$

where

$$K = 1/(1/\text{var}(\widehat{t}_W) + 1/\text{var}(\widehat{t}_B)).$$

For our data,

$$K = 1/(1/4.45 + 1/43.06) = 4.03,$$

giving

$$\widehat{t}_{RE} = 4.03 \times (4.25/4.45 + 5/43.06) = 4.03 \times 1.07 = 4.32.$$

To calculate $\text{var}(\widehat{t}_{RE})$, we use the property $\text{var}(nx) = n^2\text{var}(x)$, so that

$$\text{var}(\widehat{t}_{RE}) = K^2 \times [\text{var}(\widehat{t}_W)/(\text{var}(\widehat{t}_W))^2 + \text{var}(\widehat{t}_B)/(\text{var}(\widehat{t}_B))^2],$$

giving

$$\begin{aligned}\text{var}(\widehat{t}_{RE}) &= K^2 \times (1/\text{var}(\widehat{t}_W) + 1/\text{var}(\widehat{t}_B)) \\ &= K.\end{aligned}$$

Thus, for our data:

$$\text{var}(\widehat{t}_{RE}) = 4.03,$$

and

$$\text{SE}(\widehat{t}_{RE}) = 2.01.$$

These results are identical to those obtained initially using `PROC MIXED`. However, it is not usually quite so simple to combine estimates manually from different error strata. A general formula for calculating fixed effects estimates for all types of mixed model will be given in Section 2.2.2.

The point that we hope has been made clear by the example is the way in which the random effects model has used the information from patients 5 and 6, which would have been lost in a fixed effects analysis.

1.2.4 Estimation (or prediction) of random effects

In the previous model, the patient terms were regarded as random effects. That is, they were defined as realisations of samples from a normal distribution, with mean equal to zero and with variance σ_p^2. Thus, their expected values are zero. We know, however, that patients may differ from one another, and the idea that all have the same expected value is counterintuitive. We resolve this paradox by attempting to determine for each individual patient a *prediction* of the location within the normal distribution from which that patient's observations have arisen. This prediction will be affected by that for all other patients and will differ from the corresponding estimate in the fixed effects model. The predictions will be less widely spread than the fixed effects estimates, and because of this, they are described as *shrunken*. The extent of this shrinkage depends on the relative sizes of the patient and residual variance components. In the extreme case where the estimate of the patient variance component is zero, all patients will have equal predictions. Shrinkage will also be relatively greater when there are fewer observations per patient. It occurs for both balanced and unbalanced data, and the relevant formula is given in Section 2.2.3. Although, on technical grounds, it is more accurate to refer to *predictions* of *random effects categories* (e.g. of individual patients), in this book, we will use the more colloquial form of expression and refer to estimates of patient effects.

In our example, using the complete trial data, the random effects estimates can be obtained computationally using `PROC MIXED`. They are listed as follows along with the fixed effects patient means.

Patient number	**1**	**2**	**3**	**4**	**5**	**6**
Fixed patients	16.0	25.0	16.5	25.0	21.5	20.5
Random patients	17.2	23.9	17.6	23.9	21.3	20.6

We observe that the mean estimates are indeed 'shrunken' towards the grand mean of 20.8. Shrinkage has occurred because patients are treated as a sample from the overall patient population.

1.3 A multi-centre hypertension trial

We now introduce a more complex 'real-life' clinical trial. Measurements from this trial will be used to provide data for several examples in future chapters. Although

it is by no means the only example we will be presenting, by the repeated use of this trial, we hope that the reader will identify more readily with the analyses.

The trial was a randomised, double blind comparison of three treatments for hypertension and has been reported by Hall *et al.* (1991). One treatment was a new drug (A), and the other two (B and C) were standard drugs for controlling hypertension (A = Carvedilol, B = Nifedipine, C = Atenolol). Twenty-nine centres participated in the trial, and patients were randomised in the order of entry. Two pre-treatment and four post-treatment visits were made as follows:

- Visit 1 (week 0): measurements were made to determine whether patients met the eligibility criteria for the trial. Patients who did so received a placebo treatment for 1 week, after which they returned for a second visit.
- Visit 2 (week 1): measurements were repeated, and patients who still satisfied the eligibility criteria were entered into the study and randomised to receive one of the three treatments.
- Visits 3–6 (weeks 3, 5, 7 and 9): measurements were repeated at four post-treatment visits, which occurred at 2-weekly intervals.
- Three hundred and eleven patients were assessed for entry into the study. Of these, 288 patients were suitable and were randomised to receive one of the three treatments. Thirty patients dropped out of the study prior to Visit 6.
- Measurements on cardiac function, laboratory values and adverse events were recorded at each visit. Diastolic blood pressure (DBP) was the primary endpoint, and we will consider its analysis in this section.
- The frequencies of patients attending at least one post-treatment visit at each of the 29 centres are shown in Table 1.1.

1.3.1 Modelling the data

The main purpose of this trial was to assess the effect of the three treatments on the primary endpoint, DBP recorded at the final visit. As in the previous example, we can do this by forming a statistical model. We will now describe several possible models. A simple model (Model A) to assess just the effects of treatment could be expressed as

$$\mathrm{DBP}_i = \mu + t_k + e_i,$$

where

DBP_i = diastolic blood pressure at final visit for patient i,
μ = intercept,
t_k = kth treatment effect (where patient i has received treatment k),
e_i = error term (residual) for the ith patient.

Before the model is fitted, we should be certain that we have the most relevant dataset for our objectives. In this trial, 30 patients dropped out of the study before their final visit. If treatments have influenced whether patients dropped out,

Table 1.1 Number of patients included in analyses of final visits by treatment and centre.

	Treatment			
Centre	**A**	**B**	**C**	**Total**
1	13	14	12	39
2	3	4	3	10
3	3	3	2	8
4	4	4	4	12
5	4	5	2	11
6	2	1	2	5
7	6	6	6	18
8	2	2	2	6
9	0	0	1	1
11	4	4	4	12
12	4	3	4	11
13	1	1	2	4
14	8	8	8	24
15	4	4	3	11
18	2	2	2	6
23	1	0	2	3
24	0	0	1	1
25	3	2	2	7
26	3	4	3	10
27	0	1	1	2
29	1	0	2	3
30	1	2	2	5
31	12	12	12	36
32	2	1	1	4
35	2	1	1	4
36	9	6	8	23
37	3	1	2	6
40	1	1	0	2
41	2	1	1	4
Total	100	91	94	288

Note: Several additional centres were numbered but did not eventually participate in the study.

omitting these patients from the analysis could give rise to biased estimates of treatment effects. We therefore adopt a 'last value carried forward' approach and substitute the last recorded value for the final visit values in these patients (the issue of how to deal with missing data will be considered again in Section 2.4.7.)

1.3.2 Including a baseline covariate (Model B)

Model A was a very simple model for assessing the effect of treatment on DBP. It is usually reasonable to assume that there may be some relationship between pre-treatment and post-treatment values on individual patients. Patients with relatively high DBP before treatment are likely to have higher values after treatment and likewise for patients with relatively low DBPs. We can utilise this information in the model by fitting the baseline (pre-treatment) DBP as an additional effect in Model A:

$$\text{DBP}_i = \mu + b \cdot pre + t_k + e_i,$$

where

$b =$ baseline covariate effect,
$pre =$ baseline (pre-treatment) DBP.

In this case, we will take the values recorded at visit 2 as the baseline values. We could, of course, have considered using either the visit 1 value or the average of the visit 1 and visit 2 values, instead. The visit 2 value was chosen because it measured the DBP immediately prior to randomisation, after 1 week, during which all patients received the same placebo medication. The baseline DBP is measured on a quantitative scale (unlike treatments). Such quantitative variables are commonly described as *covariate effects*, and an analysis based on the above model is often referred to as *analysis of covariance*. The term b is a constant that has to be estimated from our data. There is an implicit assumption in our model that the relationship between the final DBP and the baseline value is linear; Additionally, that within each treatment group, an increase of 1 unit in the baseline DBP is associated with an average increase of b units in the final DBP. Figure 1.1 shows the results from fitting this model to the data (only a sample of data points is shown, for clarity).

This demonstrates that performing an analysis of covariance is equivalent to fitting separate parallel lines for each treatment to the relationship between post-treatment DBP and baseline DBP. The separation between the lines represents the magnitude of the treatment effects. The analysis will be considered in much greater detail in Section 2.5, but we note for now that two of the treatments appear to be similar to one another, while the lowest post-treatment blood pressures occur with treatment C.

The use of a baseline covariate will usually improve the precision of the estimates of the treatment effects. It will also compensate for any differences between the mean levels of the covariate in the treatment groups prior to treatment being received. Of course, our assumption that there is a linear relationship between pre-treatment and post-treatment values may not be true. If this were the case, fitting a baseline covariate could lead to less precise results. However, in practice,

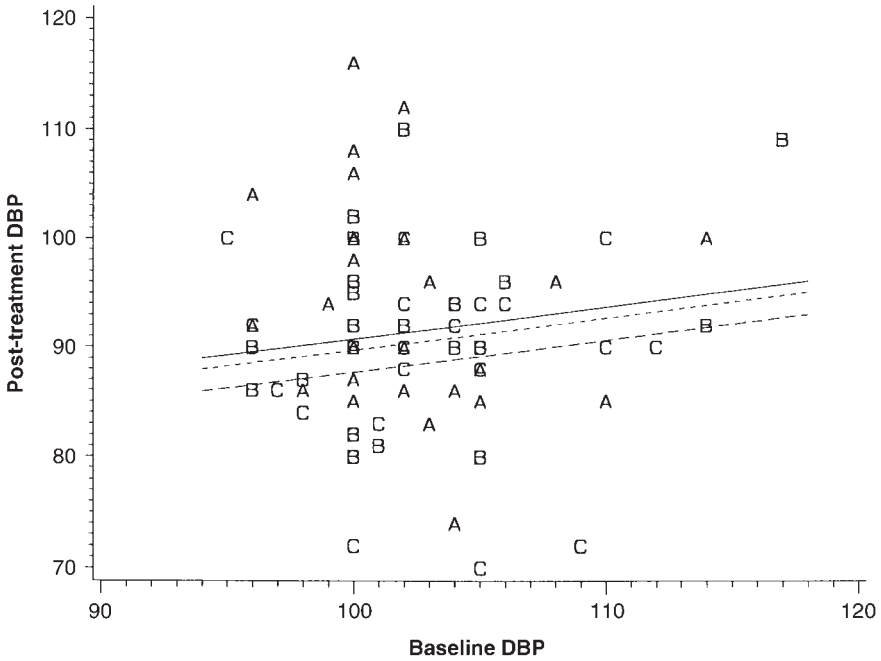


Figure 1.1 Plot to illustrate the analysis of covariance. Treatment: ———A; --------- B; – – – – – C.

the assumption is very frequently justified in medicine, and it has become almost standard to take baseline values into account in the model if they are available.

An alternative way of using baseline values (which we do not recommend) is to analyse the differences between pre-treatment and post-treatment values. However, this generally leads to less accurate results than the 'covariate' approach, particularly when the relationship between pre-treatment and post-treatment values is weak.

1.3.3 Modelling centre effects (Model C)

So far, the model has taken no account of the fact that the data are recorded at different centres. It is possible that values in some centres may tend to be higher than those in other centres. Such differences could be due, for example, to differences in the techniques of personnel across centres. It is also possible that some centres/clinics may recruit patients with differing degrees of severity of hypertension (within the bounds of the study entry criteria) who could, on average, have higher or lower values of DBP. We can allow for these possibilities by adding centre effects to Model B:

$$\mathrm{DBP}_i = \mu + b \cdot pre + t_k + c_j + e_i,$$

where

c_j = the jth centre effect.

Thus, part of the residual term in Model B may now be explained by the centre effects, c_j. If there are differences between the centres, this model will have a smaller residual variance than Model B (i.e. a smaller σ^2). This in turn allows treatment effects to be calculated with greater accuracy.

1.3.4 Including centre-by-treatment interaction effects (Model D)

In Model C, we took account of the fact that there may be an underlying difference in DBP between the centres. We did so in such a way that the effect of a patient being in a particular centre would be additive to the effect of treatment. Another possibility is that the response of patients to treatments may vary between the centres. That is, the effects of centre and treatment are non-additive or that there is an *interaction*. For example, in any multi-centre trial, if some centres tended to have more severely ill patients, it is plausible that the reaction of these patients to the treatments would differ from that of patients at other centres who are less severely ill. We can take this possibility into account in the model by allowing the treatment effects to vary between the centres. This is achieved by adding a centre·treatment *interaction* to Model C. It causes a separate set of treatment effects to be fitted for each centre.

$$\mathrm{DBP}_i = \mu + b \cdot pre + t_k + c_j + (ct)_{jk} + e_i,$$

where

$(ct)_{jk}$ = the kth treatment effect at the jth centre.

Throughout this book, we will refer to such interactions using the notation 'centre·treatment'. When Model D is fitted, the first question of interest is whether the centre·treatment effect is statistically significant. If the interaction term is significant, then we have evidence that the treatment effect differs between the centres. It will then be inadvisable to report the overall treatment effect across the centres. Results will need to be reported for each centre. If the interaction is not significant, centre·treatment may be removed from the model and the results from Model C reported. Further discussion on centre·treatment interactions appears in Chapter 5.

As we will see in more detail in Section 2.5, the centre·treatment effect is non-significant for our data ($p = 0.19$), and the results of Model C can be presented. Centre effects are statistically significant in Model C ($p = 0.004$), and so this model will be preferred to Model B.

From our data, b is estimated to be 0.22, with a SE of 0.11. Thus, if the baseline DBPs of two patients receiving the same treatment differ by 10 mm Hg, we can expect that their final DBPs will differ by only 2.2 mm Hg (0.22×10), as illustrated in Figure 1.1. The relationship is therefore weak, and hence we can anticipate

that the analysis of covariance approach will be preferable to a simple analysis of change in DBP. In fact, the statistical significance of the treatment differences is $p = 0.054$ using the analysis of covariance compared with $p = 0.072$ for the analysis of change.

1.3.5 Modelling centre and centre·treatment effects as random (Model E)

Models A–D can all be described as fixed effects models, and only the residual term is assumed to have a distribution. Alternatively, we could assume that the centre and centre·treatment effects also arose from a distribution. We again write the model as:

$$\mathrm{DBP}_i = \mu + b + t_k + c_j + (ct)_{jk} + e_i,$$

but now we assume that the residual, centre and centre·treatment effects are all realisations of separate distributions, all with zero means:

$$e_i \sim \mathrm{N}(0, \sigma^2),$$
$$c_j \sim \mathrm{N}(0, \sigma_c^2),$$
$$(ct)_{jk} \sim \mathrm{N}(0, \sigma_{ct}^2).$$

Hence, c_j and $(ct)_{jk}$ are now random effects, and b and t_k are fixed effects. This random effects model can be described as *hierarchical* since treatment effects are *contained* within the random centre·treatment effects. The concept of containment will be picked up again in Section 1.6.

Since we have assumed that centre·treatment effects have a distribution, that is that differences between treatments vary randomly across the centres, we can relate our results to the population of potential centres. This is in contrast to Model D, where treatment effects are assumed to be specific to the centres observed.

There are no hard and fast rules about whether effects should be modelled as fixed or random (or indeed whether some effects should be fitted at all). In this case, various approaches are acceptable, but they offer us different interpretations of the results. These various approaches will be discussed in much greater detail in Section 2.5, but for now, we pick up on just one point: the precision with which treatment effects are estimated. We have seen previously that fitting centre and centre·treatment effects as random enables our inferences to apply to a 'population' of centres. There is a price to be paid, however. The SEs of the treatment estimates will be inflated because we allow the treatment effects to vary randomly across centres. Thus, the mean difference in final DBP between treatments A and C is estimated as 2.93 mm Hg, with a SE of 1.41 mm Hg. In contrast, using Model C, the corresponding estimate is 2.99 mm Hg, with a smaller SE of 1.23 mm Hg. Arguments in favour of the random effects model are

the wider scope of the inferences and perhaps a more appropriate modelling of the data. In some circumstances, however, it is adequate to establish treatment differences in a specific set of centres. Statisticians in the pharmaceutical industry, for example, may prefer to avoid the penalty of less precise treatment estimates, with a corresponding reduction in *power* (the probability of obtaining statistically significant treatment differences when treatments do differ in their effect) and will often use a fixed effects model. This discussion point will be taken up again in Chapter 5.

1.4 Repeated measures data

There were four post-treatment visits in the multi-centre hypertension trial introduced in the previous section. However, so far in this chapter, we have chosen only to model measurements made at the final visit, which were of primary interest. An alternative strategy would be to include measurements from all four post-treatment visits in the model. Since measurements are made repeatedly on the same patients, we can describe these types of data as *repeated measures* data. For illustrative purposes, we now assume that the centre has no effect at all on the results and consider which models are appropriate for analysing repeated measures data. The mean levels for the three treatments at all time points are shown in Figure 1.2.

1.4.1 Covariance pattern models

Again, our primary objective is to assess the effect of the treatments on DBP, and we might again consider models which fit treatment and baseline DBP as in Model B in Section 1.3. The models will, of necessity, be more complicated, as we now have four observations per patient. In addition, it is possible that there is an underlying change in DBP over the four post-randomisation visits, and we can allow for this in the model by including a time effect, which we will denote by m. It is also possible that treatment effects may differ across time points, and to allow for this, we can also include a treatment-by-time interaction, (tm). Thus, the jth observation on patient i can be modelled as:

$$\mathrm{DBP}_{ij} = \mu + b \cdot pre + t_k + m_j + (tm)_{jk} + e_{ij},$$

where

$m_j =$ time effect at the jth post-treatment visit,
$(tm)_{jk} =$ the kth treatment effect at the jth post-treatment visit,
$e_{ij} =$ residual term for the ith patient at the jth post-treatment visit.

So far, in developing this model, we have taken no account of the fact that post-treatment measurements taken on the same patient may not be independent

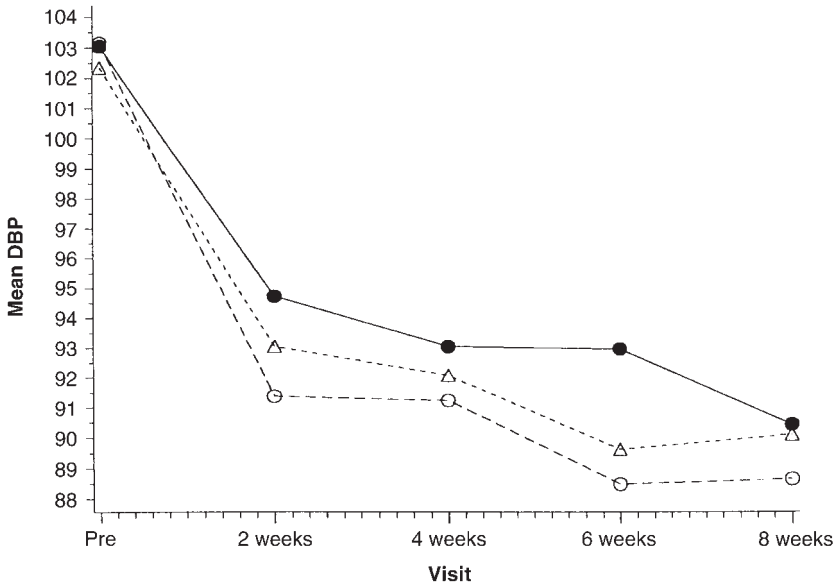


Figure 1.2 Plot of mean DBP by treatment group and visit. Treatment: ————A; ---------- B; – – – – – C.

of one another. A straightforward way to do this would be to assume that there is a constant correlation for all pairs of measurements on the same patient. Then, we could write the correlation between the residuals as

$$\text{corr}(e_{ij}, e_{ij'}) = \rho, \quad j \neq j'.$$

Alternatively, it is possible that the correlation between pairs of measurements decays as they become more widely separated in time. We could then write

$$\text{corr}(e_{ij}, e_{ij'}) = \rho^{|j'-j|}, \quad j \neq j'.$$

In the extreme, we can set a separate correlation for each pair of visits and may write

$$\text{corr}(e_{ij}, e_{ij'}) = \rho_{j,j'}, \quad j \neq j'.$$

A *covariance pattern model* can be used to fit any of these covariance (or correlation) patterns. This type of model forms another class of mixed models. Fitting covariance patterns leads to a more appropriate analysis than occurs when the fact that the repeated observations are correlated is ignored. The covariance parameter estimates may also uncover additional information about the data. They are considered in more detail in Section 6.2, and the analysis of this example is presented in Section 6.3.

1.4.2 Random coefficients models

In the previous section, the pattern of covariance between the repeated observations was modelled. An alternative approach to modelling repeated measures data would be to devise a model that explained arithmetically the relationship between DBP and time. A very simple way to do this would be to include a quantitative time effect (e.g. in measured weeks) as a covariate in the model.

$$\text{DBP}_{ij} = \mu + b \cdot pre + t_k + m \cdot time_{ij} + e_{ij},$$

where

$time_{ij}$ = time of observation j for patient i (weeks),
m = constant representing the change in DBP for unit time (week).

Thus, we obtain a time slope with gradient m, which defines a linear relationship between DBP and time. It is also possible (and indeed likely) that the relationship between DBP and time would vary between patients. To allow for this, we could model a separate regression of DBP on time for each patient. To do this, we fit patient effects to provide the intercept terms for each patient and a patient·time interaction to provide the slopes for each patient.

$$\text{DBP}_{ij} = \mu + b \cdot pre + t_k + p_i + m \cdot time_{ij} + (pm)_i \cdot time_{ij} + e_{ij},$$

where

$(pm)_i$ = difference in slope for the ith patient from the average slope,
p_i = difference from average in the intercept term for the ith patient.

It would seem reasonable to regard the values of patient effects and their slopes against time as arising from a distribution. Thus, patient and patient·time effects can both be fitted as random effects. However, the statistical properties of a model where some of the random effects involve covariate terms (time in this example) differ from ordinary random effects models (where the random effects do not involve any covariates). For this reason, we distinguish these models from ordinary random effects models and refer to them as *random coefficients models*. They form a third class of mixed models.

The statistical properties of random coefficients models are similar in many respects to random effects models. The residuals again are assumed to be independent and to have a normal distribution, with zero mean:

$$\text{var}(e_{ij}) = \sigma^2.$$

The main statistical difference from ordinary random effects models arises from the fact that when we fit a straight line, the estimates of the slope and the intercept are not independent. Thus, the patient effects (intercepts) and patient·time effects (slopes) are correlated within each patient. We therefore need to extend the approach met earlier, where separate normal distributions were used for

each random effect. We do this by use of the bivariate normal distribution. As well as terms for the means of both effects (which, as usual, are zero) and the variance components σ^2_{p} and σ^2_{pm} for patients and patient·time, this incorporates a covariance parameter $\sigma_{\mathrm{p,pm}}$. We denote the bivariate normal distribution as

$$\begin{pmatrix} p_i \\ pm_i \end{pmatrix} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}),$$

where

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{\mathrm{p}} & \sigma_{p,pm} \\ \sigma_{p,pm} & \sigma^2_{pm} \end{pmatrix}.$$

Thus, repeated measures data can be modelled using two alternative types of mixed model. Either the pattern of covariance between the repeated observations is modelled using a covariance pattern model or the relationship with time can be modelled using a random coefficients model. The latter approach is usually more appropriate if the repeated measurements do not occur at fixed intervals or when the relationship with time is of particular interest.

1.5 More about mixed models

In Sections 1.2–1.4, we used examples to introduce various concepts and types of mixed models. In this section, we pull together some of the ideas introduced earlier and define them more concisely. We also discuss some general points about mixed models. Finally, we present a perspective of mixed models, giving an outline of the history of their development.

1.5.1 What is a mixed model?

We have already met a number of models that have been described as mixed models, but it may not be clear what unites them. The key distinguishing feature of mixed models compared with fixed effects models is that they are able to model data in which the observations are not independent. To express this more positively, we say that mixed models are able to model the covariance structure of the data.

A simple type of mixed model is the *random effects model*, which was introduced in Sections 1.2 and 1.3. Here, certain effects in the model are assumed to have arisen from a distribution and thus give rise to another source of random variation in addition to the residual variation. These effects are referred to as *random effects*. For example, when patient effects were fitted in the trial introduced in Section 1.2, random variation occurred both between patients and as residual variation. Any number of random effects can be specified in a model; for example, in a multi-centre trial (as in Section 1.3), both centre and centre·treatment effects can be fitted as random, giving rise to two additional sources of variation.

In *random coefficients models*, a covariate effect is allowed to vary randomly. For example, in the repeated measures hypertension data considered in Section 1.4, interest might centre on the rate of change of DBP measured over the four treatment visits in the three arms of the trial. The random coefficients model allows this rate of change (or slope) to vary randomly between patients. This is achieved technically by fitting patients and the patient·slope interaction as random, and these effects are referred to as *random coefficients*.

The *covariance pattern model*, introduced in Section 1.4, is a third type of mixed model that directly models a pattern of correlations between observations. For example, in repeated measures trials, interest is focused on several observations of the response variable made over a period, and we can allow for the correlations (or, equivalently, covariances) between these observations. Suitable mixed models lead to more appropriate estimates of fixed effects and can investigate the nature of these covariances.

Random effects models, random coefficients models and covariance pattern models form three categories of mixed models. Mixed models can also be defined with combinations of random effects, random coefficient effects and covariance patterns. The choice will depend on the application and the objectives of the analysis.

1.5.2 Why use mixed models?

To stimulate further interest, we now mention some potential advantages that can be gained by using a mixed model. In some situations, a mixed model may simply be the most plausible model for a given data structure. For example, it is clearly desirable to take account of correlations between measurements in repeated measures data. In other circumstances, the choice is less obvious between a fixed effects model and a mixed model. Factors influencing the decision will depend partly on the structure of the data. For example, in a multi-centre trial (as in Section 1.3), the decision depends mainly on the interpretation to be put on the results. When centre and centre·treatment effects are fitted as fixed, inference can only formally be applied to the centres observed, but if they are fitted as random, inference can be applied with more confidence to a wider population of centres.

Some potential advantages that can be gained by using a mixed model are as follows:

- *Fitting covariance pattern models leads to more appropriate fixed effects estimates and SEs*. This type of model is of particular use for analysing repeated measures data. An important advantage is that the presence of missing data does not pose the major problems for analysis that can occur with a traditional analysis. The covariance parameter estimates may also uncover additional information about the data.
- *Results from a mixed model may be more appropriate to the required inference when the data structure is hierarchical*. For example, by fitting centre·treatment effects

as random in a multi-centre trial analysis (as in Section 1.3), treatment effects are allowed to vary randomly across centres, and the treatment SE increases to allow for this. Inference can then be applied to the full population of centres. However, if centre and centre·treatment effects were fitted as fixed, treatment effects would be specific to the centres observed, and inference should only be applied to these centres.

- *In a cross-over trial, estimates of treatment effects can become more accurate in datasets where there are missing data* (as in Section 1.2). The degree of benefit from using a mixed model in this situation will depend on the amount of missing data. If the original trial design was balanced and only occasional values were missing, there would be little to be gained. However, if several values were missing, treatment estimates could become notably more accurate.
- *In a random effects model, estimates of random effects are 'shrunken' compared with their fixed effects counterparts*. That is, their mean values are closer to the overall mean than if they were fitted as fixed. This helps to avoid the potential problem of extreme parameter estimates occurring due to chance when the estimates are based on small numbers. For example, in Section 1.1, we introduced an example on surgical audit. If failure rates from a particular type of operation were measured at several hospitals, a model fitting hospitals as fixed would produce unreliable failure rates for hospitals performing a small number of operations. Sometimes, these would appear as outliers compared with other hospitals, purely due to chance variation. A model fitting hospitals as random would estimate failure rates that were shrunken towards the overall failure rate. The shrinkage is greatest for hospitals performing fewer operations because less is known about them, and so misleading outliers are avoided.
- *Different variances can be fitted in a mixed model for each treatment group*. Such different variances for the treatment groups often arise in clinical trials comparing active treatments with a placebo, but they are rarely accounted for in fixed effects analyses.
- *Problems caused by missing data when fitting fixed effects models do not arise in mixed models*, provided that missing data can be assumed missing at random. This applies particularly in repeated measures trials, as noted previously, and in cross-over trials.

Although we have listed several advantages to mixed models, there is a potential disadvantage. This is that more distributional assumptions are made, and approximations are used to estimate certain model parameters. Consequently, the conclusions are dependent on more assumptions being valid, and there will be some circumstances where parameter estimates are biased. These difficulties are addressed in Section 2.4.

1.5.3 Communicating results

Statistical methods have been defined as those which elucidate data affected by a multiplicity of causes. A problem with methods of increasing complexity can be

difficulty in communicating the results of the analysis to the practitioner. There is the danger of obfuscating rather than elucidating. Estimation methods for mixed models are more complex than those used for fixed effects models, and results can therefore be more difficult to justify to non-statistical colleagues. It is not usually realistic to describe the exact methodology. However, a satisfactory explanation can often be given by emphasising the key point that mixed models take account of the covariance structure or interdependence of the data, whereas more conventional fixed effects methods assume that all observations are independent. Mixed models may therefore provide results that are more appropriate to the study design. A (hypothetical) statistical methods section in a medical journal might read:

> *The trial was analysed using a mixed model (see Brown and Prescott, 2015) with centres and the centre·treatment interaction fitted as random, so that possible differences in the size of the treatment effect across centres could be assessed.*

1.5.4 Mixed models in medicine

Frequently, there are advantages to be gained from using mixed models in medical applications. Data in medical studies are often clustered; for example, data may be recorded at several centres, hospitals or general practices. This design can be described as hierarchical, and wider inferences can be made by fitting the clustering effect as random. Repeated measures designs are also often used in medicine, and it is not uncommon for some of the observations to be missing. There are then advantages to be gained from using a mixed models analysis, which makes allowance for the missing data. Another consideration is that it is ethically desirable to use as few patients as possible, and therefore any improvements in the accuracy of treatment estimates gained by using a mixed model are particularly important. Although several examples of using mixed models in medicine have appeared in the literature for some time (e.g. Brown and Kempton, 1994), their use is still in the process of becoming routine.

1.5.5 Mixed models in perspective

It is interesting to see the application of mixed models in its historical context. In doing so, we will have to use occasional technical terms that have not yet been introduced in this book. They will, however, be met later on, and readers for whom some of the terms are unfamiliar may wish to return to this section after reading subsequent chapters.

The idea of attributing random variation to different sources by fitting random effects is not new. Fisher (1925), in his book *Statistical Methods for Research Workers*, outlined the basic method for estimating variance components by equating

the mean squares from an ANOVA table to their expected values (as described in Section 1.2). However, this method was only appropriate for balanced data. Yates (1940) and Henderson (1953) showed how Fisher's technique could be extended to unbalanced data, but their method did not always lead to unique variance components estimates. Hartley and Rao (1967) showed that unique estimates could be obtained using the method of maximum likelihood (see Section 2.2.1 for details on maximum likelihood). However, the estimates of the variance components are generally biased downwards because the method assumes that the fixed effects are known, rather than being estimated from the data. This problem of bias was overcome by Patterson and Thompson (1971) who proposed a method known as residual maximum likelihood (REML) (see Section 2.2.1), which automatically adjusted for the degrees of freedom corresponding to estimated fixed effects, as does ANOVA for balanced data. Many of the methods we describe in this book will be based on the REML method. Likelihood-based methods have only been adopted slowly because they are computationally intensive, and this has limited their use until recently.

In the past 30 years, there have been developments in parallel, in the theory and practice of using the different types of mixed model that we described earlier. Random coefficients problems have sometimes in the past been handled in two stages: first, by estimating time slopes for each patient and then by performing an analysis of the time slopes (e.g. Rowell and Walters, 1976). An early theoretical article describing the fitting of a random coefficients model in a single stage, as we will do in this book, is by Laird and Ware (1982). We consider random coefficients models again in Section 6.5.

Covariance pattern models have developed largely from time series models. Jennrich and Schluchter (1986) described the use of different covariance pattern models for analysing repeated measures data and gave some indication of how to choose between them. These models are considered more in detail in Section 6.2.

Random effects models have been frequently applied in agriculture. They have been used extensively in animal breeding to estimate heritabilities and predict genetic gain from breeding programmes (Meyer, 1986; Thompson, 1977). They have also been used for analysing crop variety trials. For example, Talbot (1984) used random effects models to estimate variance components for variety trailing systems carried out across several centres and years for different crops and was thus able to compare their general precision and effectiveness. The adoption of these models in medicine has been much slower, and a review of applications in clinical trials was given by Brown and Kempton (1994). Since then, there has been an increasing acceptability of these methods, not only by medical statisticians, but also by the regulatory authorities. The Food and Drug Administration (FDA) website contains, for example, recommended code using `SAS` to fit mixed models to multi-period cross-over trials to establish bioequivalence (www.fda.gov). Analyses of such designs are considered in Section 8.15, and other cross-over designs are considered in Chapter 7.

More recently, mixed models have become popular in the social sciences. However, they are usually described as multi-level or hierarchical models, and the terminology used for defining the models differs from that used in this book. This reflects parallel developments in different areas of application. However, the basic concept of allowing the data to have a covariance structure is the same. Two books published in this area are *Multilevel Statistical Models, Fourth Edition* by Goldstein (2010) and *Random Coefficients Models* by Longford (1993).

Perhaps, the biggest change in the use of mixed models in recent years has been the increasing use of Bayesian methods. Historically, the dual problems of computational power and available software have been a factor in restricting the use of the Bayesian approach to analysis. While this approach is based on a different philosophy, it will often lead to superficially similar results to a conventional random effects model when used with uninformative priors. The increasing availability of good software to implement the Bayesian approach and, in particular, the implementation in `SAS` of `PROC MCMC` will undoubtedly lead to its wider use in future. There has also been a shift in terminology to make the methods more acceptable to statisticians who may distrust Bayesian methods by referring to them as simulation methods. Indeed, with flat priors, you are obtaining a simulation of the full likelihood. The Bayesian approach to modelling is considered in Section 2.3.

The expansion of interest in mixed models is illustrated by its wider coverage in undergraduate and postgraduate courses in statistics and the accompanying increase in books on the topic. These include *Linear Mixed Models for Longitudinal Data by* Verbeke and Molenberghs (2000), *Generalized, Linear, and Mixed Models* by McCulloch *et al.* (2008), *Linear Mixed Models: A Practical Guide Using Statistical Software* by West *et al.* (2006), and *Mixed Models: Theory and Applications with R* by Demidenko (2013).

1.6 Some useful definitions

We conclude this introductory chapter with some definitions. The terms we are introducing in this chapter will recur frequently within subsequent chapters, and the understanding of these definitions and their relevance should increase as their applications are seen in greater detail. The terms we will introduce are containment, balance and error strata. In the analyses we will be presenting, we usually wish to concentrate on estimates of treatment effects. With the help of the definitions we are introducing, we will be able to distinguish between situations where the treatment estimates are identical whether fixed effects models or mixed models are fitted. We will also be able to identify the situations where the treatment estimates will coincide with the simple average calculated from all observations involving that treatment. The first term we need to define is containment.

1.6.1 Containment

Containment occurs in two situations. First, consider the repeated measures data encountered in Section 1.4. In that hypertension trial, DBP was recorded at four visits after treatment had been started. In the analysis of that study, the residual variance will reflect variation *within* patients at individual visits. However, in this trial, the patients receive the same treatment throughout, and so all the observations on a patient will reflect the effect of that one treatment on the patient. It can therefore perhaps be appreciated intuitively that it is the variation in response *between* patients, which is appropriate for assessing the accuracy of the estimates of treatment effects rather than the residual or 'within-patient' variation. We can see this more dramatically with a totally artificial set of data which might have arisen from this trial.

		Post-treatment visits			
Patient	**Treatment**	**1**	**2**	**3**	**4**
1	A	80	80	80	80
2	B	85	85	85	85
3	B	85	85	85	85
4	A	91	91	91	91

In this situation, there is no within-patient variation, and the residual variance is zero. Thus, if the residual variance were used in the determination of the precision of treatment estimates, we would conclude that these data showed convincingly that treatment B produced lower DBPs than treatment A. Common sense tells us that this conclusion is ridiculous with these data and that between-patient variation must form the basis for any comparison.

Here, we say that treatment effects are contained within-patient effects.

The second situation where we can meet containment can also be illustrated with data from the hypertension trial, this time concentrating on the multi-centre aspect of the design. In Section 1.3, we actually met containment for the first time when dealing with Model E, and both centre effects and centre·treatment effects were fitted as random. We say in this context that the treatment effects are contained within centre·treatment effects. In fact, there is no requirement for the centre·treatment effects to be random for the definition of containment to hold. Thus, similarly, in Model D, where the centre·treatment effects were regarded as fixed, we can still refer to the treatment effects as being contained within centre·treatment effects. It applies in general to any data with a hierarchical

structure in which the fixed effects (treatment) appears in interaction terms with other effects.

1.6.2 Balance

In many statistical textbooks that discuss the concept of balance, it is never defined but, rather, left to the intuitive feel of the reader to determine whether an experimental design is balanced. Some authors (e.g. Searle *et al.*, 1992) have defined balance as occurring when there are equal numbers of observations per *cell*. Cells are formed by all possible combinations of the levels of all the effects in the model, otherwise known as the crossing between all effects fitted in the model. For example, if we fit centre effects and treatment effects in the analysis of a multi-centre trial, and we suppose that there are four centres and two treatments, then each of the eight combinations of centre and treatment requires the same number of patients to achieve balance.

When there is balance according to this definition, the estimate of a fixed effects mean will equal the mean of all the observations at that fixed effects level. To make this clearer, if we call one of the treatments A, then the estimate of the mean response to treatment A will simply be the average of all of the observations for all patients who received treatment A. In general, this will not happen when there is imbalance. Consider the dataset illustrated in the following section. If all of the observations are present, then the estimated means for treatments A and B are 85.0 and 95.0, respectively, corresponding to their means.

Centre	Treatment A	Treatment B
1	90	100
	80	90
2	90	100
	80	(90)

If the figure in brackets is missing, however, so that there is no longer balance, then the mean treatment estimates will be 85.0 and 97.0 compared with their means of 85.0 and 96.7.

Although the condition of equal numbers in all cells is sufficient for the fixed effects mean estimates to equal their 'raw' means, it is not a necessary condition. In the multi-centre trial, for example, as long as we do not fit centre·treatment effects, it does not matter if the numbers differ across centres, provided the

treatments are allocated evenly within the centres. The following dataset produces treatment mean estimates that equal their raw means.

Centre	Treatment A	Treatment B
1	90	100
	80	
2	85	95
	85	90
	80	
	80	

Another anomaly is the cross-over trial, which is always unbalanced by the Searle *et al.* definition if period effects are fitted, as well as patient and treatment effects. This leads to empty cells because we cannot have both treatments given in the same period to any patient. Nevertheless, in a simple two-period, cross-over trial, if every patient receives every treatment, equal numbers of patients receive each sequence of treatments, and no covariates are fitted, the treatment mean estimates will equal their raw means.

We suggest, therefore, an alternative definition of balance, whereby the fixed effects means will equal their raw means whenever data are balanced but not (in general) when they are unbalanced. Balance occurs for a fixed effect when both of the following conditions are met:

- Within each category of the fixed effect (e.g. treatment), observations occur in equal proportions among categories of every other effect, which is fitted at the same containment level (see the previous section).
- If the fixed effect (e.g. treatment) is contained within a random effect (e.g. centre·treatment), then an equal number of observations are required in each category of the containing effect.

Balance across random effects

It is of importance in this book to identify the situations in which the fixed effects means (usually treatments) will differ depending on whether a fixed effects model or a mixed model is used. When balance, as defined previously, is achieved, then the fixed effects mean estimates will equal the raw means, whether a fixed effects model or a mixed model has been applied. There are other situations when the fixed effects mean estimates will not equal their raw means, but the same estimates will be obtained whether the fixed effects approach or mixed models approach is followed. This occurs when both of the following conditions apply, and we have a situation that we define as balance across random effects:

- Within each category of the specific effect (e.g. treatment), observations are allocated in equal proportions among categories of every random effect (e.g. patient), which is fitted at the same containment level.
- If the effect (e.g. treatment) is contained within a random effect (e.g. centre·treatment), then an equal number of observations are required in each category of the containing effect.

An example of the subtle distinction between these two definitions is provided by the cross-over trial example. If there were an equal number of patients on the AB and BA sequence of treatments, with no missing values, then our definition of balance would be satisfied, as described earlier. If there were no missing values, but the numbers differed between the AB and BA sequences, then there would be balance over random effects. This is true because the only random effect is patients and within each category of the containing effect (i.e. within individual patients), each treatment occurs once, and hence the definition is satisfied. Thus, the treatment estimates will be identical whether the patient effect is fitted as fixed or random, but these estimates will (in general) differ from the raw means.

This definition has been applied in the context of one particular type of mixed model; namely, the random effects model. In random coefficients models, the random coefficient blocking effect (usually patients) can be substituted for 'random effect' in the definition. In covariance pattern models, the blocking effect within which the covariance pattern is defined (again usually patients) can be substituted for 'random effect'.

Assessing balance

It can sometimes be difficult to gain an immediate feel for when balance is achieved from these definitions. The three following common situations are easily classified:

- If any observations are missing, then imbalance across random effects occurs (except for simple parallel group situations).
- If a continuous effect is fitted, then imbalance will occur (unless identical means for the effect happen to occur within each fixed effects category). However, balance across the random effects may still be achieved.
- If an equal number of observations occur in every cell and no continuous covariate is fitted, then all fixed effects will be balanced.

1.6.3 Error strata

In the random effects model, an error stratum or error level is defined by each random effect and by the residual. For example, if patients are fitted as random in a cross-over trial, there are error strata corresponding to the patients and to the residual. The *containment stratum* for a particular fixed effect is defined by the

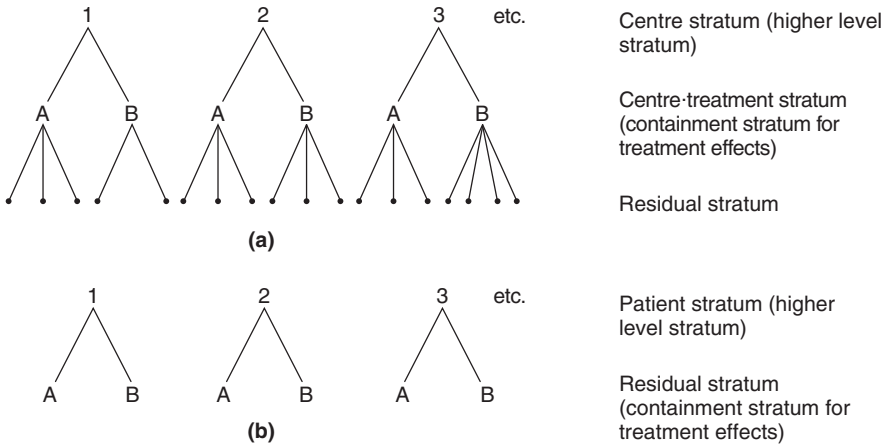


Figure 1.3 (a) Error strata for a multi-centre trial analysis fitting centre and centre·treatment effects as random; (b) error strata for a cross-over trial analysis fitting patient effects as random. A = treatment A; B = treatment B.

residual stratum, unless the effect is contained within a random effect in a random effects model or a blocking effect (see Section 6.2) in a random coefficients or covariance pattern model, in which case it is that of the containing effect. For example, in a repeated measures study, treatments are contained within patients, and thus the patient error stratum forms the containment stratum for treatments. Usually, an effect has only one containment stratum, and examples in this book will be restricted to this more usual situation. However, situations could be conceived where this is not the case. For example, if clinics and GPs were recorded in a trial and GP·treatment and clinic·treatment effects were fitted as random, then both of these effects would form containing strata for the treatment effect.

Higher level strata are defined by any random effects that are contained within the containment stratum. For example, in a multi-centre trial in which centre and centre·treatment effects are fitted as random, the centre·treatment stratum forms the containment stratum for treatment effects, and the centre stratum forms a higher level stratum (see Figure 1.3(a)). In a cross-over trial, the containment stratum for treatment effects is the residual stratum, and the patient stratum is a higher level stratum (see Figure 1.3(b)). Whenever higher level strata are present and data are not balanced across random effects, a fixed effect will be estimated using information from these strata, as well as from the containment stratum (i.e. information is *recovered* from the higher level strata).

Thus, in a cross-over trial with missing values, information is recovered from the patient level, as we saw in Section 1.2. The same occurs with missing values in a repeated measures trial where a covariance pattern is fitted. In random coefficients models, information is recovered from the patient level except in highly unusual

circumstances of equal numbers of observations at the same set of time points for all patients.

In random coefficients and covariance pattern models, error strata are not defined quite as easily because correlations occur between the random coefficients or residuals. However, random coefficients and blocking effects have a similar role to error strata, although their properties are not quite the same.

2

Normal mixed models

In this chapter, we discuss in more detail the mixed model with normally distributed errors. We will refer to this as the 'normal mixed model'. Of course, this does not imply that values of the response variables follow normal distributions because they are, in fact, mixtures of effects with different means. In practice, though, if a variable appears to have a normal distribution, the assumption of normal residuals and random effects is often reasonable.

In the examples introduced in Sections 1.1–1.4, we defined several mixed models using a notation chosen to suit each situation. In Section 2.1, we define the mixed model using a general matrix notation, which can be used for all types of mixed model. Matrix notation may at first be unfamiliar to some readers, and it is outwith the scope of this book to teach matrix algebra. A good introductory guide is *Matrices for Statistics, Second Edition* by Healy (2000). Once grasped, though, matrix notation can make the overall theory underlying mixed models easier to comprehend. Mixed models methods based on classical statistical techniques are described in Section 2.2, and in Section 2.3, the Bayesian approach to fitting mixed models will be introduced. These two sections can be omitted by readers who do not desire a detailed understanding of the more theoretical aspects of mixed models. In Section 2.4, some practical issues related to the use and interpretation of mixed models are considered, and a worked example illustrating several of the points made in Section 2.4 is described in Section 2.5. For those who wish a more in-depth understanding of the theory underlying mixed models, the textbook *Mixed Models: Theory and Applications with R, Second Edition* by Demidenko (2013) is recommended.

2.1 Model definition

In this section, the mixed model is defined using a general matrix notation that provides a compact means to specify all types of mixed model. We start by defining the fixed effects model, and then extend this notation to encompass the mixed model.

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

2.1.1 The fixed effects model

All fixed effects models can be specified in the general form

$$y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_p x_{ip} + e_i,$$
$$\text{var}(e_i) = \sigma^2.$$

For example, in Section 1.2, Model B was presented as

$$y_{ij} = \mu + p_i + t_j + e_{ij}.$$

This model used a subscript i to denote results from the ith patient and a subscript j to denote results on the jth treatment, in the context of a cross-over trial. In the general model notation, however, every observation is denoted separately with a single subscript. Thus, y_1 and y_2 could represent the observations from patient 1, y_3 and y_4 the observations from patient 2, and so on. The α terms in the general model will correspond to p_1, p_2, p_3, p_4, p_5 and p_6 and to t_1 and t_2 and are constants giving the size of the patient and treatment effects. The terms $x_{i1}, x_{i2}, \ldots, x_{i8}$ are used in this example to indicate the patient and treatment to which the observation y_i belongs, and in this case will take the values one or zero. If y_1 is the observation from patient 1 who receives treatment 1, x_{11} then will equal one (corresponding to α_1, which represents the first patient effect), $x_{12}-x_{16}$ will equal zero (as this observation is not from patients 2 to 6), x_{17} will equal one (corresponding to α_7, representing the first treatment effect) and x_{18} will equal zero. A further example to follow shortly should clarify this notation further.

The above model fits $p+1$ fixed effects parameters, $\alpha_1-\alpha_p$, and an intercept term, μ. If there are n observations, then these may be written as

$$\begin{aligned}
y_1 &= \mu + \alpha_1 x_{11} + \alpha_2 x_{12} + \cdots + \alpha_p x_{1p} + e_1,\\
y_2 &= \mu + \alpha_1 x_{21} + \alpha_2 x_{22} + \cdots + \alpha_p x_{2p} + e_2,\\
&\vdots\\
y_n &= \mu + \alpha_1 x_{n1} + \alpha_2 x_{n2} + \cdots + \alpha_p x_{np} + e_n;\\
\text{var}(e_1) &= \sigma^2,\\
&\vdots\\
\text{var}(e_n) &= \sigma^2.
\end{aligned}$$

These can be expressed more concisely in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e},$$
$$\mathbf{V} = \text{var}(\mathbf{y}) = \sigma^2\mathbf{I},$$

where

$\mathbf{y} = (y_1, y_2, y_3, \ldots, y_n)'$ = observed values,
$\boldsymbol{\alpha} = (\mu, \alpha_1, \alpha_2, \ldots, \alpha_p)'$ = fixed effects parameters,
$\mathbf{e} = (e_1, e_2, e_3, \ldots, e_n)'$ = residuals,
σ^2 = residual variance,
$\mathbf{I}$ = $n \times n$ identity matrix.

The parameters in $\boldsymbol{\alpha}$ may encompass several variables. In the above example, they covered patient effects and treatment effects. Both of these are qualitative or categorical variables, and we will refer to such effects as categorical effects. They are also sometimes referred to as factor effects. More generally, categorical effects are those where observations will belong to one of several classes. There may also be several covariate effects (such as age or baseline measurement) contained in $\boldsymbol{\alpha}$. These relate to variables that are measured on a quantitative scale. Several parameters may be required to model a categorical effect, but just one parameter is needed to model a covariate effect.

$\mathbf{X}$ is known as the *design matrix* and has the dimension $n \times p$ (i.e. n rows and p columns). It specifies values of fixed effects corresponding to each parameter for each observation. For categorical effects, the values of zero and one are used to denote the absence and presence of effect categories, and for covariate effects, the variable values themselves are used in $\mathbf{X}$.

We will exemplify the notation with the following data, which are the first nine observations in a multi-centre trial of two treatments to lower blood pressure.

Centre	Treatment	Pre-treatment systolic BP	Post-treatment systolic BP
1	A	178	176
1	A	168	194
1	B	196	156
1	B	170	150
2	A	165	150
2	B	190	160
3	A	175	150
3	A	180	160
3	B	175	160

The observation vector $\mathbf{y}$ is formed from the values of the post-treatment systolic blood pressure:

$$\mathbf{y} = (176, 194, 156, 150, 150, 160, 150, 160, 160)'.$$

If pre-treatment blood pressure and treatment were fitted in the analysis model as fixed effects (ignoring centres for the moment), then the design matrix would be

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \mu & \alpha_1 & \alpha_2 & \alpha_3 \end{matrix} \\ \begin{pmatrix} 1 & 178 & 1 & 0 \\ 1 & 168 & 1 & 0 \\ 1 & 196 & 0 & 1 \\ 1 & 170 & 0 & 1 \\ 1 & 165 & 1 & 0 \\ 1 & 190 & 0 & 1 \\ 1 & 175 & 1 & 0 \\ 1 & 180 & 1 & 0 \\ 1 & 175 & 0 & 1 \end{pmatrix} \end{array},$$

where the columns of the design matrix correspond to the parameters

$$\begin{aligned} \mu &= \text{intercept}, \\ \alpha_1 &= \text{pre-treatment blood pressure parameter}, \\ \alpha_2, \alpha_3 &= \text{parameters for treatments A and B}. \end{aligned}$$

We note in this case that the design matrix, $\mathbf{X}$, is overparameterised. This means that there are linear dependencies between the columns, for example, we know that α_3 will be zero if $\alpha_2 = 1$ and one if $\alpha_2 = 0$. $\mathbf{X}$ could alternatively be specified omitting the α_3 column to correspond with the number of parameters actually modelled. However, the overparameterised form is used here since it is used for specifying contrasts by SAS procedures such as `PROC MIXED` (this procedure will be used to analyse most of the examples in this book).

$\mathbf{V}$ is a matrix containing the variances and covariances of the observations. In the usual fixed effects model, variances for all observations are equal, and no observations are correlated. Thus, $\mathbf{V}$ is simply $\sigma^2\mathbf{I}$.

2.1.2 The mixed model

The mixed model extends the fixed effects model by including random effects, random coefficients and/or covariance terms in the residual variance matrix. In this section, the general notation will be given, and in the following three sections, the specific forms of the covariance matrices for each type of mixed model will be specified.

Extending our fixed effects model to incorporate random effects (or coefficients), the mixed model may be specified as

$$\begin{aligned} y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_p x_{ip} \\ + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_q z_{iq} + e_i \end{aligned}$$

for a model fitting p fixed effects parameters and q random effects (or coefficients) parameters. It will be recalled from Chapter 1 that random effects are assumed to follow a distribution, whereas fixed effects are regarded as fixed constants. The model can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e},$$

where $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\alpha}$ and $\mathbf{e}$ are as defined in the fixed effects model, and

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)' = \text{random effect/coefficient parameters.}$$

$\mathbf{Z}$ is a second design matrix with dimension $n \times q$ giving the values of random effects corresponding to each observation. It is specified in exactly the same way as $\mathbf{X}$ was for the fixed effects, except that an intercept term is not included. If centres were fitted as random in the multi-centre example given previously, the $\boldsymbol{\beta}$ vector would then consist of three parameters, β_1, β_2 and β_3, corresponding to the three centres, and the $\mathbf{Z}$ matrix would be

$$\mathbf{Z} = \begin{array}{c} \begin{matrix} \beta_1 & \beta_2 & \beta_3 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Alternatively, if both the centre and the centre·treatment effects were fitted as random, then the vector of random effects parameters, $\boldsymbol{\beta}$, would consist of the three centre parameters, plus six centre·treatment interaction parameters β_4, β_5, β_6, β_7, β_8 and β_9. The $\mathbf{Z}$ matrix would then be

$$\mathbf{Z} = \begin{array}{c} \begin{matrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 & \beta_7 & \beta_8 & \beta_9 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Again, note that this matrix is overparameterised due to linear dependencies between the columns. It could alternatively have been written using four columns: $3-1=2$ for the centre effects and $(3-1)\times(2-1)=2$ for the centre·treatment effects.

Covariance matrix, V

We saw in the fixed effects model that all observations have equal variances, and the observations are uncorrelated. This leads to the $\mathbf{V}$ matrix being diagonal. When random effects are fitted, we saw in Section 1.2 that this results in correlated observations. In the context of the cross-over trial, we saw that observations on the same patient were correlated (with covariance equal to the patient variance component), while those on different patients were uncorrelated. We now generalise this result, using the matrix notation.

The covariance of $\mathbf{y}$, $\text{var}(\mathbf{y})=\mathbf{V}$, can be written as

$$\mathbf{V}=\text{var}(\mathbf{X\alpha}+\mathbf{Z\beta}+\mathbf{e}).$$

Since we assume that the random effects and the residuals are uncorrelated,

$$\mathbf{V}=\text{var}(\mathbf{X\alpha})+\text{var}(\mathbf{Z\beta})+\text{var}(\mathbf{e}).$$

Since $\boldsymbol{\alpha}$ describes the fixed effects parameters, $\text{var}(\mathbf{X\alpha})=\mathbf{0}$. Also, $\mathbf{Z}$ is a matrix of constants. Therefore,

$$\mathbf{V}=\mathbf{Z}\text{var}(\boldsymbol{\beta})\mathbf{Z}'+\text{var}(\mathbf{e}).$$

We will let $\mathbf{G}$ denote $\text{var}(\boldsymbol{\beta})$, and since the random effects are assumed to follow normal distributions, we may write $\boldsymbol{\beta}\sim\text{N}(\mathbf{0},\mathbf{G})$. Similarly, we write $\text{var}(\mathbf{e})=\mathbf{R}$, the residual covariance matrix, and $\mathbf{e}\sim\text{N}(\mathbf{0},\mathbf{R})$. Hence,

$$\mathbf{V}=\mathbf{ZGZ}'+\mathbf{R}.$$

In the following three sections, we will define the structure of the $\mathbf{G}$ and $\mathbf{R}$ matrices in random effects models, random coefficients models and covariance pattern models.

2.1.3 The random effects model covariance structure

The G matrix

The dimension of $\mathbf{G}$ is $q\times q$, where q is equal to the total number of random effects parameters.

In random effects models, $\mathbf{G}$ is always diagonal (i.e. random effects are assumed uncorrelated). If just centre effects were fitted as random in the simple multi-centre

example with three centres, then **G** would have the form

$$\mathbf{G} = \begin{pmatrix} \sigma_c^2 & 0 & 0 \\ 0 & \sigma_c^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{pmatrix},$$

where σ_c^2 is the centre variance component. If both centre and centre·treatment effects were fitted as random, then **G** would have the form

$$\mathbf{G} = \begin{pmatrix} \sigma_c^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 \end{pmatrix},$$

where σ_{ct}^2 is the centre·treatment variance component.

The R matrix

The residuals are uncorrelated in random effects models and $\mathbf{R} = \sigma^2\mathbf{I}$:

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}.$$

The V matrix

We showed earlier that the variance matrix, **V**, has the form $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$. $\mathbf{ZGZ}'$ specifies the covariance due to the random effects. If just centre effects are

fitted as random, then we obtain

$$\mathbf{ZGZ'} = \begin{pmatrix} \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \end{pmatrix}.$$

This matrix could be obtained by the laborious process of matrix multiplication but it always has the same form. It has a *block diagonal* form with the size of blocks corresponding to the number of observations at each random effects category. The total variance matrix, $\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$, is then

$$\mathbf{V} = \begin{pmatrix} \sigma_c^2 + \sigma^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma^2 \end{pmatrix}.$$

This also has a block diagonal form with the covariances for observations at the same centre equal to the random effects variance component, σ_c^2, and variance terms on the diagonal equal to the sum of the centre and residual variance components, $\sigma_c^2 + \sigma^2$. (We note that this corresponds to the results from the cross-over trial example introduced in Section 1.2, where the random effect was patient rather than centre.) If both centre and centre·treatment effects had been fitted as random, then

$$\mathbf{ZGZ'} = \begin{pmatrix} \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 \end{pmatrix}$$

and

$$\mathbf{V} = \begin{pmatrix} \theta & \sigma_c^2+\sigma_{ct}^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2+\sigma_{ct}^2 & \theta & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \theta & \sigma_c^2+\sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2+\sigma_{ct}^2 & \theta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \theta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta & \sigma_c^2+\sigma_{ct}^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2+\sigma_{ct}^2 & \theta & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \theta \end{pmatrix},$$

where $\theta = \sigma_c^2 + \sigma_{ct}^2 + \sigma^2$. Thus, $\mathbf{V}$ again has a block diagonal form with a slightly more complicated structure. The centre·treatment variance component is added to the covariance terms for observations at the same centre and with the same treatment.

2.1.4 The random coefficients model covariance structure

The statistical properties of random coefficients models were described in the repeated measures example introduced in Section 1.4. We will define their covariance structure in terms of the general matrix notation we have just introduced for mixed models. Random coefficients models will be discussed in more detail in Section 6.5.

The following data will be used to illustrate the covariance structure. They represent measurement times for the first three patients in a repeated measures trial of two treatments.

Patient	Treatment	Time (days)
1	A	t_{11}
1	A	t_{12}
1	A	t_{13}
1	A	t_{14}
2	B	t_{21}
2	B	t_{22}
3	A	t_{31}
3	A	t_{32}
3	A	t_{33}

If patient and patient·time effects were fitted as random coefficients, then there would be six random coefficients. We will change notation from Chapter 1 for ease

of reading to define these as $\beta_{p,1}$, $\beta_{pt,1}$, $\beta_{p,2}$, $\beta_{pt,2}$, $\beta_{p,3}$ and $\beta_{pt,3}$, allowing an intercept (patient) and slope (patient·time) to be calculated for each of the three patients. The **Z** matrix would then be

$$\mathbf{Z} = \begin{array}{c} \begin{matrix} \beta_{p,1} & \beta_{pt,1} & \beta_{p,2} & \beta_{pt,2} & \beta_{p,3} & \beta_{pt,3} \end{matrix} \\ \begin{pmatrix} 1 & t_{11} & 0 & 0 & 0 & 0 \\ 1 & t_{12} & 0 & 0 & 0 & 0 \\ 1 & t_{13} & 0 & 0 & 0 & 0 \\ 1 & t_{14} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & t_{21} & 0 & 0 \\ 0 & 0 & 1 & t_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & t_{31} \\ 0 & 0 & 0 & 0 & 1 & t_{32} \\ 0 & 0 & 0 & 0 & 1 & t_{33} \end{pmatrix} \end{array}.$$

The R matrix

As in random effects models, the residuals are uncorrelated, and the residual covariance matrix is

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}.$$

The G matrix

In a random coefficients model, the patient effects (intercepts) are correlated with the random patient·time effects (slopes). Correlation occurs only for coefficients on the same patient (i.e. between $\beta_{p,i}$ and $\beta_{pt,i}$), and coefficients on different patients are uncorrelated. Thus, the **G** matrix would be

$$\mathbf{G} = \begin{pmatrix} \sigma^2_p & \sigma_{p,pt} & 0 & 0 & 0 & 0 \\ \sigma_{p,pt} & \sigma^2_{pt} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_p & \sigma_{p,pt} & 0 & 0 \\ 0 & 0 & \sigma_{p,pt} & \sigma^2_{pt} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_p & \sigma_{p,pt} \\ 0 & 0 & 0 & 0 & \sigma_{p,pt} & \sigma^2_{pt} \end{pmatrix},$$

where σ^2_{p} and σ^2_{pt} are the patient and patient·time variance components, and $\sigma_{\text{p,pt}}$ is the covariance between the random coefficients. Note that **G** has dimension 6×6 because the model includes six random coefficients.

The V matrix

Again, **V** is obtained as $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$, where $\mathbf{ZGZ}'$ specifies the covariance due to the random coefficients and for our data has the form

$$\mathbf{ZGZ}' = \begin{pmatrix} v_{1,11} & v_{1,12} & v_{1,13} & v_{1,14} & 0 & 0 & 0 & 0 & 0 \\ v_{1,12} & v_{1,22} & v_{1,23} & v_{1,24} & 0 & 0 & 0 & 0 & 0 \\ v_{1,13} & v_{1,23} & v_{1,33} & v_{1,34} & 0 & 0 & 0 & 0 & 0 \\ v_{1,14} & v_{1,24} & v_{1,34} & v_{1,44} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v_{2,11} & v_{2,12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v_{2,12} & v_{2,22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & v_{3,11} & v_{3,12} & v_{3,13} \\ 0 & 0 & 0 & 0 & 0 & 0 & v_{3,12} & v_{3,22} & v_{3,23} \\ 0 & 0 & 0 & 0 & 0 & 0 & v_{3,13} & v_{3,23} & v_{3,33} \end{pmatrix},$$

where $v_{i,jk} = \sigma^2_{\text{p}} + (t_{ij} + t_{ik})\sigma_{\text{p, pt}} + t_{ij}t_{ik}\sigma^2_{\text{pt}}$.

Thus, $\mathbf{ZGZ}'$ has a block diagonal form, with the size of blocks corresponding to the number of observations on each patient. It is added to the diagonal $\mathbf{R} = \sigma^2\mathbf{I}$ to form the total covariance matrix, $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$, which will also have a block diagonal form. It may appear that covariances will increase with time and that a different origin for time would lead to different results. However, **V** is invariant to time origin and, although the covariance parameters alter, we still obtain the same overall results (see further discussion in Section 6.5 and examples in Section 6.6).

Note that the covariance structure in random coefficients models is induced by the random coefficients. This differs from covariance pattern models shown in the following sections where covariance parameters in the **R** (or occasionally **G**) matrix are chosen to reflect a particular pattern in the data.

2.1.5 The covariance pattern model covariance structure

In the repeated measures example in Section 1.4, the idea of modelling the covariances between observations was introduced. In this section, we show how covariance patterns fit into the general mixed models definition using matrix notation. In covariance pattern models, the covariance structure of the data is not defined by specifying random effects or coefficients but by specifying a pattern for the covariance terms directly in the **R** (or, occasionally, **G**) matrix. Observations within a chosen *blocking variable* (e.g. patients) are allowed to be correlated, and a pattern for their covariances is specified. This pattern is usually chosen to depend on a

variable such as time or the visit number. **R** will have a block diagonal form and can be written as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{R}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_5 & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_6 & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_7 & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \end{pmatrix}.$$

The submatrices, $\mathbf{R}_i$, are covariance blocks corresponding to the ith blocking effect (the ith patient, say). They have dimension equal to the number of repeated measurements on each patient. The **0** represent matrix blocks of zeros, giving zero covariances for observations on different patients. We now give two examples of **R** matrices using a small hypothetical dataset. We assume that the first three patients in a repeated measures trial attended at the following visits:

Patient	Visit
1	1
1	2
1	3
2	1
2	2
2	3
2	4
3	1
3	2

Then, using patients as the blocking effect, an **R** matrix where a separate correlation is allowed for each pair of visits (this can be described as a 'general' covariance pattern) is given by

$$\mathbf{R} = \begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{12} & \sigma_2^2 & \theta_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{13} & \theta_{23} & \sigma_3^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \theta_{12} & \theta_{13} & \theta_{14} & 0 & 0 \\ 0 & 0 & 0 & \theta_{12} & \sigma_2^2 & \theta_{23} & \theta_{24} & 0 & 0 \\ 0 & 0 & 0 & \theta_{13} & \theta_{23} & \sigma_3^2 & \theta_{34} & 0 & 0 \\ 0 & 0 & 0 & \theta_{14} & \theta_{24} & \theta_{34} & \sigma_4^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_1^2 & \theta_{12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{12} & \sigma_2^2 \end{pmatrix}.$$

Alternatively, a simpler pattern assuming a constant correlation between each visit pair (known as the 'compound symmetry' pattern) is given by

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix},$$

where $\rho =$ the correlation between observations on the same patient.

Commonly, in the analysis of repeated measures data, no random effects are fitted, in which case, the variance matrix $\mathbf{V} = \mathbf{R}$. Otherwise, $\mathbf{R}$ is added to $\mathbf{ZGZ'}$ to form the full variance matrix for the data, $\mathbf{V}$. Other ways to define covariance patterns in the $\mathbf{R}_i$ matrix blocks will be considered in Section 6.2.

Covariance patterns in the G matrix

It is also possible, although less usual, to fit a covariance pattern in the $\mathbf{G}$ matrix so that the random effects are correlated within a blocking effect. For example, consider a repeated measures trial in which each patient is assessed at a number of visits and where several measurements are made at each visit. One may wish to model the correlation between visits, within patients, as well as modelling the correlation between observations at the same visit. To achieve this, it is necessary to specify covariance patterns in the $\mathbf{G}$ matrix as well as for the $\mathbf{R}$ matrix. We will return to this type of covariance structure in the example given in Section 8.1 (Model 3).

2.2 Model fitting methods

In this section, the numerical methods for fitting mixed models will be described. This material is not essential to those who only wish to apply mixed models without gaining a theoretical understanding, and we will assume some knowledge of likelihood in our presentation. We showed in Chapter 1 that in some simple circumstances the random effects model could be fitted by using an ANOVA table. However, in general, a more sophisticated method is required to fit the mixed model. In Section 2.2.1, the likelihood function for the mixed model is specified, and different methods for maximising it are introduced. The different approaches to maximising the likelihood lead to different estimates for the model parameters and their standard errors. The model fitting process has three distinctive components: estimating fixed effects (i.e. $\boldsymbol{\alpha}$), estimating random effects (i.e. $\boldsymbol{\beta}$), and

estimating variance parameters (i.e. variance components or covariance terms). In Sections 2.2.2–2.2.4, we will see how the various fitting methods apply to each of these components, respectively.

2.2.1 The likelihood function and approaches to its maximisation

The mixed model can be fitted by maximising the *likelihood function* for values of the data. The likelihood function, L, measures the likelihood of the model parameters given the data and is defined using the density function of the observations. In models where the observations are assumed independent (e.g. fixed effects models), the likelihood function is simply the product of the density functions for each observation. However, observations in a mixed model are not independent, and the likelihood function therefore needs to be based on a multivariate density function for the observations. The likelihood for the variance parameters and the fixed effects can be defined using the multivariate normal distribution for $\mathbf{y}$ (the term 'variance parameters' here encompasses all parameters in the $\mathbf{G}$ and $\mathbf{R}$ matrices, that is, variance components and the covariance parameters). As random effects have expected values of zero and therefore do not affect the mean, this distribution has a mean vector $\mathbf{X}\alpha$ and a covariance matrix $\mathbf{V}$. The likelihood function based on the multivariate normal density function is then

$$L = \frac{\exp\left[-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\alpha})\right]}{(2\pi)^{(1/2)n}|\,\mathbf{V}\,|^{1/2}}.$$

In practice, the log likelihood function is usually used in place of the likelihood function, since it is simpler to work with, and its maximum value coincides with that of the likelihood. The log likelihood is given by

$$\log(L) = K - \frac{1}{2}\left[\log|\,\mathbf{V}\,| + (\mathbf{Y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\alpha})\right],$$

where

$K = -\frac{1}{2}n\log(2\pi) =$ (a constant that can be ignored in the maximization process),

$n =$ number of observations.

The values of the model parameters that maximise the log likelihood can then be determined. We now introduce briefly several approaches to fitting the mixed model, which are all based (directly or indirectly) on maximising the likelihood function. As we will see, the methods are not all equivalent and can lead to different estimates of the model parameters. Following this introduction to the methods, we then look in more detail at separate aspects of the fitting process: estimation of fixed effects, estimation of random effects and estimation of variance parameters.

Maximum likelihood (ML)

This method is based on the concept of maximising the log likelihood with respect to the variance parameters while treating the fixed effects, $\boldsymbol{\alpha}$, as constants. Having obtained the variance parameter estimates, the fixed effects estimates are then obtained by treating the variance parameters as fixed and finding the values of $\boldsymbol{\alpha}$ that maximise the log likelihood. This method has the effect of producing variance parameter estimates that are biased downwards to some degree. This can be illustrated with a very simple example. Suppose we have a simple random sample, $x_1, x_2, \ldots, x_n$, and wish to estimate the mean and variance. If $\widehat{\mu}$ is the sample mean, then the maximum likelihood (ML) variance estimator would be $\sum_i (x_i - \widehat{\mu})^2 / n$ rather than the unbiased estimator $\sum_i (x_i - \widehat{\mu})^2 / (n-1)$. The bias is the greatest when a small number of degrees of freedom (DF) are used for estimating the variance parameters.

Residual maximum likelihood (REML)

Residual maximum likelihood (REML) (sometimes referred to as restricted maximum likelihood) was first suggested by Patterson and Thompson (1971). In this approach, the parameter $\boldsymbol{\alpha}$ is eliminated from the log likelihood so that it is defined only in terms of the variance parameters. We outline the method in the following section.

First, we obtain a likelihood function based on the residual terms, $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$. This contrasts with the likelihood initially defined, which is based directly on the observations, $\mathbf{y}$. You will notice that these residuals differ from the ordinary residuals, $\mathbf{e} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}} - \mathbf{Z}\widehat{\boldsymbol{\beta}}$, in that $\mathbf{Z}\widehat{\boldsymbol{\beta}}$ is not deducted. Some authors (including those of REML) also refer to the $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$ as residuals. This is not unreasonable, since they can be considered as error terms that include all sources of random variation (residual and random effects). In this section, we will refer to $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$ as the *full residuals* in order to differentiate them from the ordinary residuals. The full residuals, $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$, are, in fact, a linear combination of the $\mathbf{y}$ as we will see in Section 2.2.2 where we show how to produce the estimates, $\widehat{\boldsymbol{\alpha}}$. It can also be shown that $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\alpha}}$ are independent (see Diggle *et al.,* 1994, Section 4.5), and therefore the joint likelihood for $\boldsymbol{\alpha}$ and the variance parameters, $\boldsymbol{\gamma}$, can be expressed as a product of the likelihoods based on $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\alpha}}$:

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) = L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}) L(\boldsymbol{\alpha}; \widehat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}).$$

Thus, the likelihood for $\boldsymbol{\gamma}$ based on $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}$ is given by

$$L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}) = L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) / L(\boldsymbol{\alpha}; \widehat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}).$$

Now, from the above equations,

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) \propto |\mathbf{V}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\right)$$

and $\hat{\boldsymbol{\alpha}}$ has a multivariate normal distribution, with mean and variance given by the ML estimates, which will be obtained in Section 2.2.2. Hence,

$$L(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}) \propto |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1/2} \exp\left(\frac{1}{2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'\mathbf{X}\mathbf{V}^{-1}\mathbf{X}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\right).$$

Taking the ratio of these two likelihoods, we obtain the REML as

$$L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) \propto |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1/2}|\mathbf{V}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})\right),$$

and the REML log likelihood as

$$\log(L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})) = K - \frac{1}{2}[\log|\mathbf{V}| - \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1} + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})].$$

Although $\hat{\boldsymbol{\alpha}}$ still appears, it does so as a function of the variance parameters ($\hat{\boldsymbol{\alpha}}$ is derived in Section 2.2.2 as $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$). The parameter $\boldsymbol{\alpha}$ does not appear. Note that the difference between the REML log likelihood and the ordinary log likelihood is caused by the extra term $\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1}$, which is the log of the determinant of var($\hat{\boldsymbol{\alpha}}$). The REML is equivalent to having integrated $\boldsymbol{\alpha}$ out of the likelihood for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, and for this reason, REML is sometimes referred to as a 'marginal' method. Because the REML takes account of the fact that $\boldsymbol{\alpha}$ is a parameter and not a constant, the resulting variance parameter estimates are unbiased. As with ML, $\boldsymbol{\alpha}$ is then estimated by treating the variance parameters as fixed and finding the values of $\boldsymbol{\alpha}$ that maximise the REML log likelihood.

Iterative generalised least squares (IGLS)

This method can be used iteratively to fit a mixed model, and the results will be the same as those obtained using ML. This approach obtains estimates of the fixed effects parameters, $\boldsymbol{\alpha}$, by minimising the product of the full residuals weighted by the inverse of the variance matrix, $\mathbf{V}^{-1}$. The residual product is given by $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})$.

The variance parameters are obtained by setting the matrix of products of the full residuals $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})$ equal to the variance matrix, $\mathbf{V}$, specified in terms of the variance parameters. This gives

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})' = \mathbf{V}$$

and leads to a set of $n \times n$ simultaneous equations (one for each element in the $n \times n$ matrices) that can be solved iteratively for the variance parameters (n = number of observations). The equations do not take account of the fact that $\boldsymbol{\alpha}$ will be estimated and is not known. Therefore, the resulting variance parameter estimates are biased downwards and are the same as the ML estimates.

An adaption to iterative generalised least squares (IGLS) that leads to the unbiased REML variance parameter estimates is *restricted iterative generalised least*

squares (RIGLS). It is described by Goldstein (1989) who notes that because $\hat{\boldsymbol{\alpha}}$ is estimated and not known,

$$\mathrm{E}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathrm{var}(\mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}',$$

which leads to an alternative set of $n \times n$ equations to solve for the variance parameters

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'.$$

Since the observed full residuals, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}$, depend on the fixed effects parameter estimates, iteration is required between the fixed effects and the variance parameter estimates to obtain either the IGLS or RIGLS solution. Further detail on this method can be found in Goldstein (2010).

Variance parameter bias

We have stated that estimates of variance parameters are biased in ML (and IGLS) and unbiased in REML (and RIGLS). We believe that lack of bias is an important property, and therefore REML will be used to analyse most of the examples in this book. Fixed effects estimates are unlikely to differ greatly between ML and REML analyses, but their standard errors will always be biased downwards if the variance parameters are biased (because they are calculated as weighted sums of the variance parameters). This will be most noticeable when the DF used to estimate the variance parameters are small. Fixed effects estimates are also subject to additional sources of bias as explained in the following section.

2.2.2 Estimation of fixed effects

Maximum likelihood and REML

The fixed effects solution can be obtained by maximising the likelihood by differentiating the log likelihood with respect to $\boldsymbol{\alpha}$ and setting the resulting expression to zero. This leads to a solution that is expressed in terms of the variance parameters:

$$\mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) = \mathbf{0}.$$

Rearrangement gives

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\,\mathbf{y},$$

and the variance of $\hat{\boldsymbol{\alpha}}$ is obtained as

$$\begin{aligned}\mathrm{var}(\hat{\boldsymbol{\alpha}}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\,\mathrm{var}(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.\end{aligned}$$

This formula is based on the assumption that $\mathbf{V}$ is known. Because $\mathbf{V}$ is, in fact, estimated, it can be shown that there will be some downward bias in var($\widehat{\boldsymbol{\alpha}}$). However, this is usually very small, and approximate corrections for the bias can be made (see Section 2.4.3). Note that this occurs even when the variance component estimates are themselves unbiased. In Section 2.3, we will see that the Bayesian approach avoids having to make this assumption so that the problem of bias does not arise.

Iterative generalised least squares

The same ML solution for $\boldsymbol{\alpha}$ can alternatively be obtained using *generalised least squares*. With this approach, the product of the full residuals, weighted by the inverse of the variances, $(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})$, is minimised by differentiation with respect to $\boldsymbol{\alpha}$:

$$\frac{\delta(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})}{\delta\boldsymbol{\alpha}} = -2\mathbf{X}'\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\alpha}).$$

By setting this expression to zero, we find that the residual product is minimised when $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\alpha}$, giving the solution for $\boldsymbol{\alpha}$ as

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\,\mathbf{y},$$

again with variance

$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{V}^{-1}\,\mathbf{X})^{-1}.$$

This solution is sometimes referred to as the *generalised least squares* solution.

In unweighted least squares, where $\mathbf{V} = \sigma^2\mathbf{I}$, the solution will be $\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, with variance $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$. This is the solution obtained from fitting fixed effects models using *ordinary least squares (OLS)*; for example, by using `PROC GLM` in `SAS`. The difference in the mixed models estimate is due to the use of the inverse variance matrix, $\mathbf{V}^{-1}$, in the formula for $\widehat{\boldsymbol{\alpha}}$. When the data are unbalanced, it is this difference that causes information on the fixed effects to be combined from different error strata.

2.2.3 Estimation (or prediction) of random effects and coefficients

In general, random effects and coefficients are defined to have normal distributions with zero means, and the specific values they take must be thought of as realisations of a sample from a distribution. Thus, their expected values are, by definition, zero. Nonetheless, as we saw earlier, in Section 1.2 in the context of patient effects in a cross-over trial, it is possible to obtain predictions of them. We will now outline how these predictions are obtained.

Maximum likelihood and REML

To predict the random effects, $\boldsymbol{\beta}$ (for simplicity, we use the term random effects to refer to either random effects or coefficients), we define a likelihood function in terms of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ ($\boldsymbol{\gamma}$ is the vector of variance parameters). This can be written as the product of the likelihoods for: the fixed parameters and variance parameters in the $\mathbf{R}$ matrix ($\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}_{\mathbf{R}}$) given the data with $\boldsymbol{\beta}$, for the moment, treated as fixed ($\mathbf{y} \mid \boldsymbol{\beta}$); and that of the variance parameters in the $\mathbf{G}$ matrix ($\boldsymbol{\gamma}_{\mathbf{G}}$) given the random effects $\boldsymbol{\beta}$:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\,;\, \mathbf{y}) = L(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{R}};\, \mathbf{y} \mid \boldsymbol{\beta}) L(\boldsymbol{\gamma}_{\mathbf{G}}\,;\, \boldsymbol{\beta}),$$

where

$\boldsymbol{\gamma}_{\mathbf{R}} =$ variance parameters in the $\mathbf{R}$ matrix,
$\boldsymbol{\gamma}_{\mathbf{G}} =$ variance parameters in the $\mathbf{G}$ matrix.

Using multivariate normal distributions for $\mathbf{y} \mid \boldsymbol{\beta}$ and $\boldsymbol{\beta}$, we have

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) \propto |\mathbf{R}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})\right)$$
$$\times |\mathbf{G}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}\right),$$

giving the corresponding log likelihood as

$$\log(L) = -\frac{1}{2}[\log|\mathbf{R}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})$$
$$+ \log|\mathbf{G}| + \boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}] + K.$$

The ML solution for $\boldsymbol{\beta}$ can be obtained by differentiating this log likelihood with respect to $\boldsymbol{\beta}$ and setting the resulting expression to zero:

$$\delta \log(L) / \delta \boldsymbol{\beta} = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}) - \mathbf{G}^{-1}\boldsymbol{\beta}$$
$$= -(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\boldsymbol{\beta} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}).$$

Setting to zero gives

$$\widehat{\boldsymbol{\beta}}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}) = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}),$$
$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \quad \text{(A)}$$

As discussed in Chapter 1, estimates are 'shrunken' compared with what they would have been if fitted as fixed. Note that since the estimates are centred about zero, the intercept estimate would need to be added in order to obtain mean random effects estimates. In random effects models, the $\mathbf{R}$ matrix is diagonal, $\mathbf{R} = \sigma^2\mathbf{I}$, and we can alternatively write

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} / \sigma^2)^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}).$$

Compared with the OLS solution for a fixed effects model, $\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, we notice the additional term, $\mathbf{G}^{-1} / \sigma^2$, in the denominator. It is this term that causes the estimates to be shrunken towards zero.

Also, an alternative, more compact, form for $\widehat{\boldsymbol{\beta}}$ can be obtained from the solution given earlier (A) using matrix manipulation and recalling that $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$:

$$\widehat{\boldsymbol{\beta}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}).$$

The variance of $\widehat{\boldsymbol{\beta}}$ can be obtained as

$$\text{var}(\widehat{\boldsymbol{\beta}}) = \mathbf{GZ}'\mathbf{V}^{-1}\mathbf{ZG} - \mathbf{GZ}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZG}.$$

As with var($\widehat{\boldsymbol{\alpha}}$), this formula is based on the assumption that $\mathbf{V}$ is known. Because $\mathbf{V}$ is, in fact, estimated, there will be some downward bias in var($\widehat{\boldsymbol{\beta}}$), although this is usually small (see Section 2.4.3). Again, the Bayesian approach (Section 2.3) avoids having to make this assumption, and the problem of bias does not arise.

Iterative generalised least squares

We note that $\widehat{\boldsymbol{\beta}}$ is not obtained directly from the usual least squares equations used by IGLS. However, once estimates for the variance parameters are obtained using IGLS, the above formulae can then be applied to obtain $\widehat{\boldsymbol{\beta}}$.

2.2.4 Estimation of variance parameters

In this section, we consider the numerical procedures used to apply ML and least squares based methods for estimating variance parameters. These are usually embedded in the statistical packages used to perform the analysis, and so knowledge of their details is not necessary for application. We present them in this section for completeness.

Maximum likelihood and REML

Both of these methods work by obtaining variance parameter estimates that maximise a likelihood function. A solution cannot be specified by a single equation as it was for the fixed and random effects because the derivatives of the log likelihood with respect to the variance parameters are non-linear. An iterative process such as the widely applied *Newton–Raphson algorithm* is therefore required. This works by repeatedly solving a quadratic approximation to the log likelihood function. We will now illustrate this algorithm by first showing how a quadratic function is solved in terms of its first and second derivatives, and then showing how this solution is used to define the iterative Newton–Raphson method for maximising a likelihood function.

Solving a quadratic To express the solution of a quadratic function of $\boldsymbol{\theta}$ in terms of its first and second derivatives, we first write the function in matrix notation in the general quadratic form as

$$f(\boldsymbol{\theta}) = a + \mathbf{b}'\boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\theta}'\mathbf{C}\boldsymbol{\theta}.$$

Note that the first derivative of $f(\boldsymbol{\theta})$ will be a vector, and the second derivative will be a square matrix. For example, if $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, then the first derivative of $f(\boldsymbol{\theta})$ is a vector of length 3, and the second derivative of the log likelihood is a 3×3 matrix. The first derivative is given by

$$f'(\boldsymbol{\theta}) = \mathbf{b} + \mathbf{C}\boldsymbol{\theta},$$

and the second derivative by

$$f''(\boldsymbol{\theta}) = \mathbf{C}.$$

The solution for $\boldsymbol{\theta}$, which maximises $f(\boldsymbol{\theta})$, $\widehat{\boldsymbol{\theta}}$, is obtained by setting the first derivative to zero:

$$f'(\boldsymbol{\theta}) = \mathbf{b} + \mathbf{C}\boldsymbol{\theta} = \mathbf{0},$$

to give

$$\widehat{\boldsymbol{\theta}} = -\mathbf{C}^{-1}\mathbf{b}.$$

By adding and deducting an arbitrary value, $\boldsymbol{\theta}_i$ say, this solution to the quadratic function can then be expressed as

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_i - \mathbf{C}^{-1}(\mathbf{b} + \mathbf{C}\boldsymbol{\theta}_i),$$

and can be rewritten in terms of $\boldsymbol{\theta}_i$ and the first and second derivatives of $f(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_i$:

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_i - f''^{-1}(\boldsymbol{\theta}_i) f'(\boldsymbol{\theta}_i)$$

We now show that this mathematical trick is the key to the use of the Newton–Raphson algorithm.

Newton–Raphson iteration We are seeking the value $\boldsymbol{\theta}$ to maximise the likelihood function $f(\boldsymbol{\theta})$. If we start with an initial approximate solution $\boldsymbol{\theta}_1$, we make the assumption that the function will be approximately quadratic to obtain an improved approximation $\boldsymbol{\theta}_2$, using the formula obtained previously. The process is then repeated using $\boldsymbol{\theta}_2$ as the approximate solution, to obtain an improved approximation $\boldsymbol{\theta}_3$. Although the function will not, in general, be quadratic, in the region of the ML solution, the quadratic approximation is usually quite good, and the Newton–Raphson iterative procedure will usually converge appropriately. Convergence is obtained when parameter values change very little between successive iterations. The iterative process can be defined by

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - f''^{-1}(\boldsymbol{\theta}_i) \times f'(\boldsymbol{\theta}_i),$$

where $f'(\boldsymbol{\theta}_i)$ and $f'(\boldsymbol{\theta}_i)$ are the actual (unapproximated) derivatives of $f(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_i$. The matrix of second derivatives, $f''(\boldsymbol{\theta})$, is often referred to as the *Hessian matrix*. In mixed models, $f(\boldsymbol{\theta})$ is taken to be the log likelihood expressed in terms of the variance parameters, $\boldsymbol{\theta}$.

The need to evaluate the derivatives at each iteration can make the Newton–Raphson algorithm computationally intensive. Computation can be made easier

by using a matrix known as the information matrix in place of $f''(\boldsymbol{\theta}_i)$ in the iterative process. The information matrix is the expected value of the Hessian matrix, and it is easier to compute than the Hessian matrix because some of the correlation terms are zero. When it is used, the process can be referred to as the *method of scoring* or *Fisher scoring*. This method has been shown to be more robust to poor starting values than the Newton–Raphson algorithm. Within SAS, PROC MIXED uses Fisher scoring for the first iteration and then Newton–Raphson for the remaining iterations, as the default fitting method.

Covariances of variance parameters An indication of the precision of the variance parameters can be obtained from an estimate of their variance and their degree of correlation from the covariances. However, this estimate is based on standard asymptotic (large sample) theory. The asymptotic covariances of the variance parameters are given by the negative of the expected values of second partial derivatives of the log likelihood (see, e.g. Searle *et al.*, 1992, Section 3.7):

$$\hat{\text{var}}(\theta_i) = -\text{E}\{\delta^2 \log(L)/\delta\theta_i\delta\theta_i\},$$
$$\hat{\text{cov}}(\theta_i, \theta_j) = -E\{\delta^2 \log(L)/\delta\theta_i\delta\theta_j\}.$$

Since the resulting covariances are based on asymptotic theory and are related to the estimated variance parameter values themselves, they should be interpreted cautiously. Also, remember that the distribution of variance parameters is not usually symmetrical.

In SAS, asymptotic standard errors can be obtained by specifying the COVTEST option in the PROC MIXED statement, while the use of the ASYCOV option will produce the asymptotic variance matrix of the covariance parameter estimates. The CL option will produce confidence limits for the covariance parameter estimates based on a chi-squared distribution by using the ratio of the covariance parameter estimate to its standard error. Thus, the warning for cautious interpretation should be applied to any of these options.

Iterative generalised least squares

This method estimates the variance parameters by setting the full residual products equal to the variance matrix, $\mathbf{V}$, specified in terms of the variance parameters and solving the resulting equations. This leads to a set of $n \times n$ simultaneous equations that can be solved iteratively for the variance parameters ($n =$ number of observations). In ordinary IGLS that gives variance parameters that are biased downwards, the equations are as follows:

$$\mathbf{V} = (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'.$$

To illustrate the structure of the equations more clearly, we will show their form for a small hypothetical dataset. We assume that the first two patients in a repeated measures trial attended at the following visits:

Patient	Visit
1	1
1	2
1	3
2	1
2	2

The equations in a model using a separate covariance term for each pair of visits (i.e. with a 'general' covariance structure) are given by

$$\begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & 0 & 0 \\ \theta_{12} & \sigma_2^2 & \theta_{23} & 0 & 0 \\ \theta_{13} & \theta_{23} & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \theta_{12} \\ 0 & 0 & 0 & \theta_{12} & \sigma_2^2 \end{pmatrix}$$

$$= \begin{pmatrix} (y_1-\mu_1) & (y_1-\mu_1) & (y_1-\mu_1) & (y_1-\mu_1) & (y_1-\mu_1) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_2-\mu_2) & (y_2-\mu_2) & (y_2-\mu_2) & (y_2-\mu_2) & (y_2-\mu_2) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_3-\mu_3) & (y_3-\mu_3) & (y_3-\mu_3) & (y_3-\mu_3) & (y_3-\mu_3) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_4-\mu_4) & (y_4-\mu_4) & (y_4-\mu_4) & (y_4-\mu_4) & (y_4-\mu_4) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_5-\mu_5) & (y_5-\mu_5) & (y_5-\mu_5) & (y_5-\mu_5) & (y_5-\mu_5) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \end{pmatrix}.$$

In this simple example, equating corresponding terms from the left-hand side and right-hand side of this equation gives

$$\theta_{13} = (y_1-\mu_1)(y_3-\mu_3),$$
$$\theta_{23} = (y_2-\mu_2)(y_3-\mu_3).$$

There are two equations for θ_{12}, and we may obtain an estimate from their average:

$$\theta_{12} = [(y_1-\mu_1)(y_2-\mu_2) + (y_5-\mu_5)(y_4-\mu_4)]/2.$$

Thus, in this artificially simple dataset, the covariance terms are calculated over the average of the observed covariances for just one or two subjects. In a genuine dataset, the covariances will be estimated from the average of the observed covariances from many more subjects. The variance terms can be estimated in a similar way.

The approach extends to other covariance patterns. Suppose that with the same artificial dataset we wish to fit a simpler correlation pattern, with a constant covariance between each visit pair (i.e. compound symmetry pattern), then

$$\begin{pmatrix} \sigma^2 & \theta & \theta & 0 & 0 \\ \theta & \sigma^2 & \theta & 0 & 0 \\ \theta & \theta & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & \theta \\ 0 & 0 & 0 & \theta & \sigma^2 \end{pmatrix}$$

$$= \begin{pmatrix} (y_1-\mu_1) & (y_1-\mu_1) & (y_1-\mu_1) & (y_1-\mu_1) & (y_1-\mu_1) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_2-\mu_2) & (y_2-\mu_2) & (y_2-\mu_2) & (y_2-\mu_2) & (y_2-\mu_2) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_3-\mu_3) & (y_3-\mu_3) & (y_3-\mu_3) & (y_3-\mu_3) & (y_3-\mu_3) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_4-\mu_4) & (y_4-\mu_4) & (y_4-\mu_4) & (y_4-\mu_4) & (y_4-\mu_4) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \\ (y_5-\mu_5) & (y_5-\mu_5) & (y_5-\mu_5) & (y_5-\mu_5) & (y_5-\mu_5) \\ \times(y_1-\mu_1) & \times(y_2-\mu_2) & \times(y_3-\mu_3) & \times(y_4-\mu_4) & \times(y_5-\mu_5) \end{pmatrix}.$$

The solution is then given by averaging the observed covariances over all pairs of time points so that

$$\begin{aligned} \theta = \; & [(y_1-\mu_1)(y_2-\mu_2) + (y_4-\mu_4)(\mu_5-\mu_5) \\ & + (y_1-\mu_1)(y_3-\mu_3) + (y_2-\mu_2)(\mu_3-\mu_3)]/4, \end{aligned}$$

and

$$\sigma^2 = \sum_{i=1}^{5} (y_i-\mu_i)^2 / 5.$$

In random effects and coefficients models, each linear equation may involve more than one parameter. Simple averaging will then not be sufficient to obtain the parameter estimates, and standard methods for solving sets of linear equations can be applied.

For RIGLS that gives unbiased variance parameters (as in REML), the equations are as follows:

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'.$$

Further details on IGLS can be found in Goldstein (2003).

2.3 The Bayesian approach

A Bayesian approach to fitting a mixed model provides an interesting alternative to the classical methods we have already described, which are based on maximising

the likelihood function. Some statisticians have difficulties in accepting the philosophy of the Bayesian approach and will not be willing to use such an analysis. We feel that such an attitude is misguided, and a Bayesian approach to mixed models is introduced because it has several potential advantages over maximum likelihood methods. It must be recognised, though, that the application of Bayesian methods in medicine is relatively uncommon and may be unfamiliar to many of the potential 'consumers' of the analysis. There may therefore be a greater communication problem in reporting such analyses.

Bayesian methods have also been hampered in the past by the fact that they can require very large amounts of computer power and time. However, this has now become less of a restriction, and their use has become much more widespread. In the context of mixed models, Bayesian methods have been developed most fully for use with random effects models, and we therefore concentrate primarily on these models. As some of the ideas underlying Bayesian modelling are quite different from those underlying classical statistical approaches, we will first spend some time giving a brief introduction to Bayesian concepts. An example of a Bayesian analysis will be given in Section 2.5.

2.3.1 Introduction

In a Bayesian analysis, the distribution of the model parameters is obtained and then used to obtain parameter estimates (e.g. of the mean treatment effects). This contrasts with the fitting methods we have considered so far, which are based on finding values of parameters that optimise the likelihood function. The distribution of the model parameters is obtained by combining the likelihood function with a *prior* distribution for the parameters to obtain what is called the *posterior* distribution. The prior distribution can be either informative (based on prior knowledge of the parameter) or non-informative (containing no prior information on the parameter). While there can be good reason to use informative priors, particularly when they use only basic knowledge of the context of the problem (e.g. range for human temperature is 30–45°C or a variance parameter is positive), the results obtained will then not be wholly dependent on the data and lead to an altered interpretation. A Bayesian analysis using informative priors might therefore be considered a distinct subset of Bayesian methods. In this section, we will only consider the use of non-informative priors.

Bayesian methods are often considered to be quite different from Maximum likelihood methods (e.g. ML and REML). However, when a non-informative prior is used in a Bayesian analysis, the posterior density has a similar shape to the likelihood function. In fact, when a flat prior is used, the posterior has an identical shape to the likelihood. Thus, the main difference between a Bayesian analysis (with non-informative priors) and a ML approach is that the posterior density (which can be similar to the likelihood function) is fully evaluated, whereas in ML, only the parameter values that *maximise* the likelihood are obtained.

The advantage of fully evaluating the posterior density is that exact posterior standard deviations and probability intervals can be obtained from the posterior distributions for each model parameter. These statistics are analogous to the standard errors and confidence intervals derived using ML. We will also show how posterior distributions can be used to yield exact 'Bayesian' *p*-values, which are analogous to *p*-values resulting from classical significance tests (see Section 2.3.3). We believe that the potential to obtain such exact statistics is a major advantage of using a Bayesian approach over ML where parameter standard errors, confidence intervals and *p*-values are computed using formulae that assume that the variance parameters are known (see Sections 2.2.2 and 2.4.3).

A potential disadvantage with using a Bayesian approach is that the techniques used to obtain the posterior density usually rely on simulation, and it can be difficult to define exactly when a simulated distribution for the parameters has converged to the true distribution. In contrast, it is usually much easier to conclude that a likelihood function has been maximised. The problems associated with defining convergence when simulation techniques are used are currently the subject of research interest.

We now describe Bayesian terminology in more detail. This will provide a background to the choices that need to be made when setting up models. However, some users may only require a more basic understanding and may wish to omit the details of how the posterior distribution is simulated. For a more in-depth introduction, the text *Bayesian Inference in Statistical Analysis* by Box and Tiao (1973) or *Bayesian Methods for Data* Analysis by Carlin and Louis (2008) is recommended. In addition, the text *In All Likelihood* by Pawitan (2001) provides a helpful background and comparison between Bayesian analysis, likelihood-based analysis and the frequentist approach, and *Bayesian Approaches to Clinical Trials and Health-Care Evaluation* by Spiegelhalter *et al.* (2004) illustrates the use of Bayesian methods specifically in clinical trials.

2.3.2 Determining the posterior density

In a Bayesian model, it is assumed that parameters have a prior distribution, which reflects knowledge (or lack of it) about the parameters before the analysis commences, and that this distribution can be modified to take account of observed data to form a posterior distribution and that the posterior distribution can be used to make inferences about the model parameters. The posterior density function for the model parameters is proportional to the product of the likelihood function and the prior density of all the model parameters, $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta};\ \mathbf{y}) = p(\boldsymbol{\theta})L(\boldsymbol{\theta};\ \mathbf{y})\,/\,K,$$

where

$$K = \int p(\boldsymbol{\theta})L(\boldsymbol{\theta};\ \mathbf{y})\mathrm{d}\boldsymbol{\theta}.$$

The denominator integral is necessary to ensure that the posterior integrates to one. The likelihood function can be based on any distribution, and hence Bayesian methods can be applied to data with either normal or non-normal distributions. We note that by using a flat prior, $p(\boldsymbol{\theta}) \propto c$ ($c =$ constant), the posterior density becomes

$$p(\boldsymbol{\theta};\ \mathbf{y}) = L(\boldsymbol{\theta};\ \mathbf{y}) / K,$$

which is the *standardized likelihood function*. A Bayesian analysis is then equivalent to evaluating the likelihood function over its full parameter space. However, as we will see later, somewhat surprisingly, a flat prior cannot always be regarded as non-informative.

A mixed model will usually contain several parameters. For example, consider a two-way, cross-over trial model fitting treatments and periods as fixed and patients as random. We let

$\mu =$ intercept,
$t =$ treatment difference,
$p =$ period difference,
$s_1, s_2, \ldots, s_{n-1} =$ subject effects (n subjects),
$\sigma^2 =$ residual variance component,
$\sigma_s^2 =$ subject variance component.

(Note that the redundant parameters for the second treatment and period, and the final subject, are omitted in this parameterisation.)

We assume that estimation of the subject effects is not of interest. Taking the Bayesian approach, a prior distribution, $p(\mu,\ t,\ p, \sigma^2, \sigma_s^2)$ say, is first specified for the model parameters. The posterior distribution of $\mu,\ t,\ p, \sigma^2$ and σ_p^2 is then obtained using the product of the prior distribution and the likelihood function:

$$p(\mu,\ t,\ p, \sigma^2, \sigma_s^2; \mathbf{y}) = p(\mu, t, p, \sigma^2, \sigma_s^2) \times L(\mu, t, p, \sigma^2, \sigma_s^2; \mathbf{y}) / K,$$

where K is the standardising constant used to ensure that the distribution integrates to one.

$$K = \int\int\int\int\int p(\mu,\ t,\ p, \sigma^2, \sigma_s^2) \times L(\mu, t, p, \sigma^2, \sigma_s^2; \mathbf{y}) \delta\mu\ \delta t\ \delta p\ \delta\sigma^2 \delta\sigma_s^2.$$

The posterior distribution can then be used to estimate each of the model parameters. More detail on how this is done is given in the next section.

2.3.3 Parameter estimation, probability intervals and *p*-values

The full posterior distribution $p(\mu,\ t,\ p, \sigma^2, \sigma_s^2; \mathbf{y})$ is not often useful in itself for making inferences about individual parameters. To make inferences about one

parameter, t say, the posterior is usually integrated over all the other parameters (μ, p, σ^2 and σ_s^2) to form a *marginal posterior distribution* for t:

$$p_t(t;\mathbf{y}) = \int\int\int\int p(\mu,\ t,\ p, \sigma^2, \sigma_s^2;\mathbf{y})\mathrm{d}\mu\ \mathrm{d}p\ \mathrm{d}\sigma^2\mathrm{d}\sigma_s^2.$$

This distribution can then be used to calculate estimators for the treatment parameter. With this approach, we note that the problem of biased fixed and random effects standard errors will not occur as it does in ML (see Section 2.4.3). This is because variances can be obtained from exact parameter distributions, rather than by using formulae that assume that the variance components are known.

Parameter estimation

There are no strict rules about which estimator should be used for a given parameter. For location parameters (i.e. fixed and random effects), the posterior distribution can often be conveniently summarised by the posterior mean and its variance. The square root of var(t) gives the standard deviation of t. Although this is analogous to the standard error of t given by ML, Bayesian statisticians usually prefer to quote standard deviations of parameters rather than standard errors of means.

The mean value is not usually the most appropriate estimate for a variance component because the marginal distribution is not symmetrical. The median or mode are better choices when judged by the 'average closeness to the true value' measured, say, by the mean squared error. Another estimator sometimes used is the posterior mean of the square root of the variance component. However, our own preference is to use the median. Box and Tiao (1973, Appendix A5.6) discuss the choice of variance component estimators in more detail. However, in many applications, the variance component estimate is not of particular interest and may not even need to be obtained.

Probability intervals

Exact probability intervals for model parameters can be calculated directly from their marginal posterior distributions. These may either be computed such that the two tails of the distribution have equal density 'equal tail interval'. For example, for a 95% probability interval, each tail would have a density of 0.025. An alternative approach is to obtain the interval such that:

- The probability density of points within the interval is higher than all points outside it.
- The total probability for the region is equal to a specified $1 - \alpha$ (e.g. $\alpha = 0.05$ for a 95% probability interval), referred to as a 'highest posterior density (HPD)' interval. It is usually more difficult to obtain than the equal tail interval.

Statisticians have well documented differences of opinion about how to draw two-tailed inferences from asymmetrical distributions, but our preference is for the equal tail area approach. This ensures that a two-tailed test at say the 5% level of significance is equivalent to two one-tailed tests at the 2.5% level. This has the advantage of conforming to the symmetry principle that we should be equally surprised by differences from the Null hypothesis in either direction. It also adheres to Regulatory guidance that states 'the approach of setting type I errors for one-sided tests at half the conventional type I error used in two-tailed tests is preferable in regulatory settings. This provides consistency with the two-sided confidence intervals that are generally appropriate for estimating the possible size of the difference between two treatments' (Section 5.5 in Statistical Principles for Clinical Trials, ICH Harmonised Tripartite Guideline E9, 1998).

Probability intervals are analogous to the confidence intervals calculated by maximum likelihood methods. However, unlike confidence intervals, they are calculated directly from the posterior distribution and do not rely on estimates of parameter standard errors.

p-values

The concept of the significance test as it appears in 'classical' statistics does not fall within the Bayesian philosophy. In classical statistics, a hypothesis is tested by constructing an appropriate test statistic and obtaining the distribution of this statistic under a null hypothesis (e.g. the treatment difference is zero). Acceptance of the null hypothesis is then obtained from the position of the test statistic within this 'null' distribution. Specifically, we calculate the probability under the null hypothesis of obtaining values of the test statistic that are as, or more, extreme than the observed value. Probabilities below 0.05 are often seen as sufficient evidence to reject a null hypothesis. The closest equivalent using Bayesian methods is achieved through the use of probability intervals. The value of α for the probability interval that has zero on the boundary can be used to provide a 'Bayesian' *p*-value to examine the plausibility that a parameter is zero. This is equivalent to a two-sided 'classical' *p*-value. However, it has the advantage of being exact, and there are no potential inaccuracies in obtaining a test statistic (based on standard error estimates) or the DF for its distribution.

2.3.4 Specifying non-informative prior distributions

We have given little indication so far of the distributional form that non-informative priors should take. The requirement is that a non-informative prior for a parameter should have minimal influence on the results obtained for that parameter. The theoretical background of how to set non-informative priors is not easily accessible to those without a background in Bayesian statistics; so we will outline the methods that can be used in the following section. Prior to that,

though (excusing the pun), we will take the pragmatic approach and simply describe some distributions that have been suggested to provide non-informative priors for mixed models.

For the fixed effects (μ, t and p in the cross-over model), there are (at least) two suitable non-informative priors:

- uniform distribution $(-\infty, \infty)$, $p(\theta) \propto c$, that is a flat prior,
- normal $(0, K)$, where K is a very large number.

We note that as K tends to ∞, the normal distribution tends to the uniform $(-\infty, \infty)$ distribution. For the practitioner, there is the question of how big a number is very large? This depends on the scale on which observations are being recorded. Recording distances in millimetres gives larger numbers than recording in kilometres. The choice of K should be so that its square root is at least an order of magnitude larger than any of the observations.

For the variance components (σ^2 and σ_s^2 in the cross-over model), any of the following distributions may provide suitable non-informative priors:

- uniform $(-\infty, \infty)$, $p(\theta) \propto c$,
- reciprocal distribution, $p(\theta) \propto c/\theta$ (c = constant),
- inverse gamma distribution (K, K), where K is a very small number.

In this book, we will not describe the inverse gamma distribution, other than to note that it is a two-parameter distribution and that as the parameters tend to zero, the distribution tends to the reciprocal distribution. The practical guidance to the choice of K is again that it should be at least an order of magnitude smaller than the observations.

In practice, it often makes little difference to the results obtained from the posterior distribution, whichever of these priors is chosen. An exception is when the true value of a variance component is close to zero. Under these circumstances, the posterior distribution arising from the uniform prior will differ depending on whether the variance component is constrained to be positive. We note, though, that many statisticians would be unlikely to choose the uniform prior for variance components because it is known that variance components cannot be negative. However, it is this prior that leads to a posterior density that is exactly proportional to the likelihood.

We now introduce in more detail a general approach to the setting of non-informative priors. This section will be of greatest interest to those readers who wish to extend their knowledge of the Bayesian approach.

Setting a non-informative prior

At first sight, it might appear that simply using a flat distribution, $p(\theta) \propto c$, would provide a non-informative prior for a parameter, θ. However, this is not always the case. To obtain a non-informative prior for θ, it is first necessary to find a transformation of θ, $h(\theta)$ say, for which the likelihood has the same shape when plotted

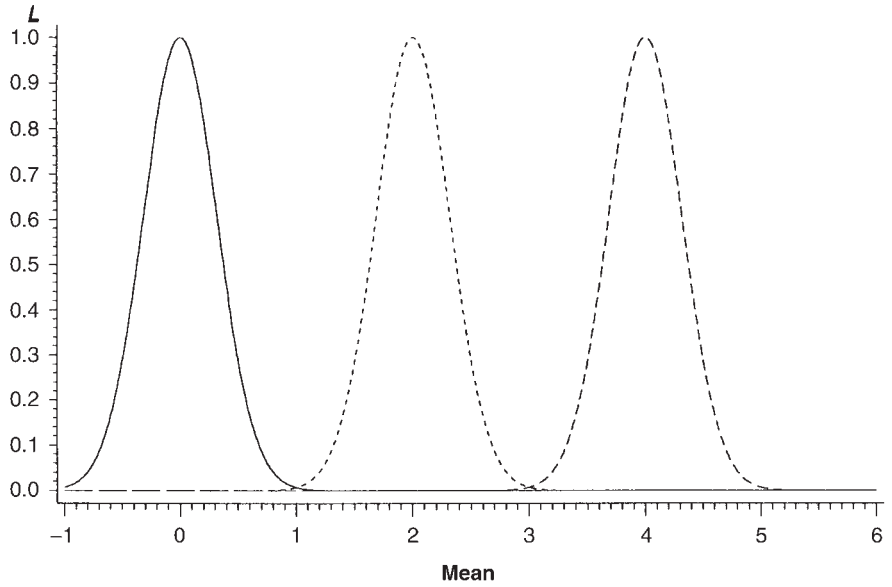


Figure 2.1 Likelihood vs. mean. Mean: _____ 0; -------- 2; – – –4. (*L* is proportional to likelihood).

against it regardless of any changes to the data. Since the shape of the likelihood distribution is then unaffected by the value of $h(\theta)$, a flat density for $h(\theta)$, $p(h(\theta)) \propto c$ will give a non-informative prior for $h(\theta)$. From this, a non-informative prior distribution of θ can be obtained as (see Box and Tiao, 1973, Section 1.3).

$$\begin{aligned} p(\theta) &= p(h(\theta)) \, |\, \mathrm{d}h(\theta) / \mathrm{d}\theta \,| \\ &\propto c \, |\, \mathrm{d}h(\theta) / \mathrm{d}(\theta) \,|. \end{aligned}$$

Consider a sample from the normal distribution, $\mathrm{N}(\mu, \sigma^2)$. If $L(\mu; \sigma, \mathbf{y})$ is plotted against μ, the same shape always arises regardless of the data; only the location is dependent on the data (see Figure 2.1). Thus, in this case, h is the identity function, $h(\mu) = \mu$, and we obtain $p(\mu) \propto c \, |\, \mathrm{d}h(\mu)/\mathrm{d}\mu \,| = c$, a flat prior.

However, when $L(\sigma; \mu, \mathbf{y})$ is plotted against σ, we find that its shape varies with the data (see Figure 2.2). Thus, h is not now the identity function, and we need to consider alternative functions. It turns out that plotting $L(\sigma; \mu, \mathbf{y})$ against $\log(\sigma)$ gives likelihoods that have the same shape regardless of the data values (see Figure 2.3). Thus, $h(\sigma) = \log(\sigma)$, and we obtain $p(\sigma) \propto c \, |\, \mathrm{d} \log(\sigma)/\mathrm{d}\sigma \,| = c/\sigma$.

In general, it is not necessary to go to the trouble of plotting the likelihood against various transformations of a parameter to find the function h. h can instead be

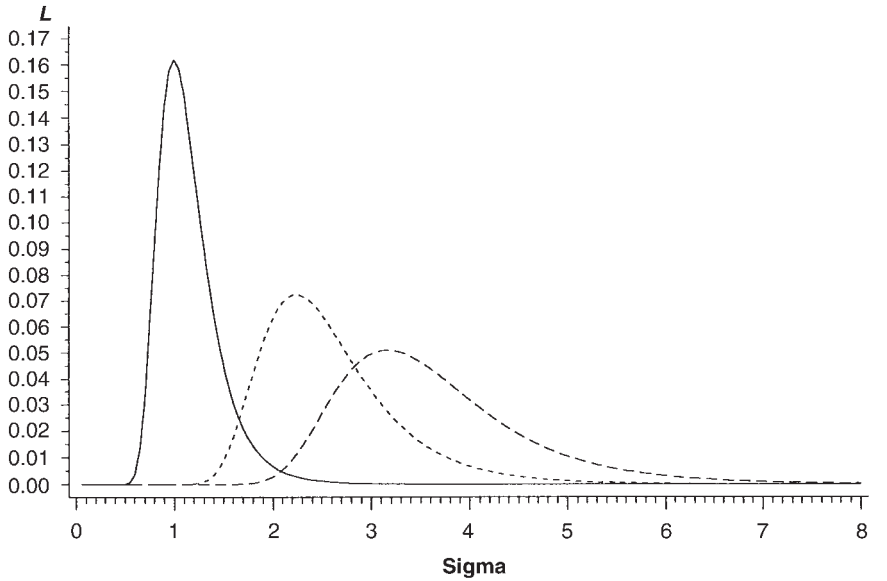


Figure 2.2 Likelihood vs. sigma. Variance: _____ 1; -------- 5; – – –10. (*L* is proportional to likelihood).

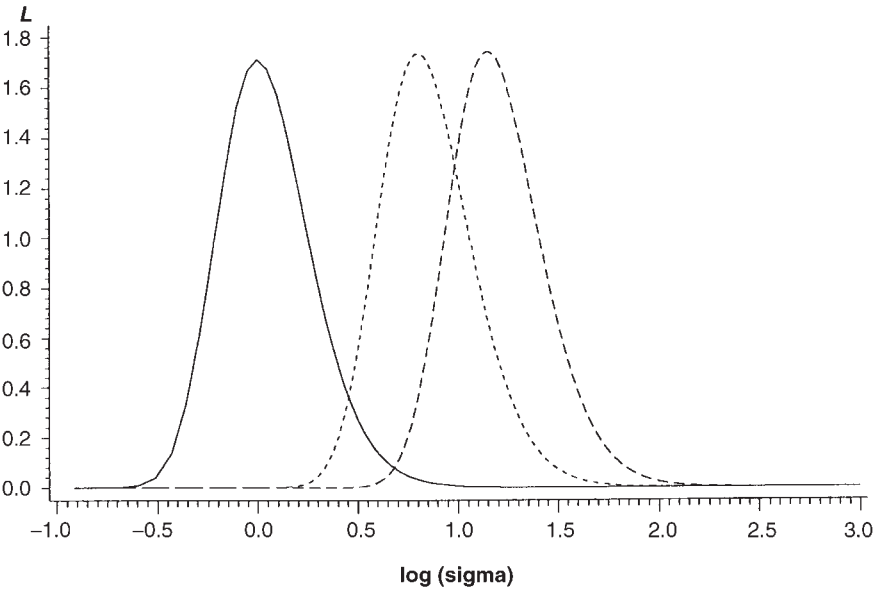


Figure 2.3 Likelihood vs. log (sigma). Variance: _____ 1; ------- 5; – – –10. (*L* is proportional to likelihood).

obtained by writing the likelihood for θ in the form $g(h(\theta) - t(\mathbf{y}))$. For example, the likelihood for μ for the normal sample can be written as

$$L(\mu; \sigma, \mathbf{y}) \propto (1/\sigma^n) \exp(-n(\mu - \overline{y})^2 / 2\sigma^2).$$

This gives $g(z) = \exp(-z^2/2\sigma^2)$, $h(\mu) = \mu$ and $t(\mathbf{y}) = -\overline{y}$.

The likelihood for σ is obtained by rearranging the likelihood function as

$$L(\sigma; \mu, \mathbf{y}) \propto \exp\{-n[\log(\sigma) - \log(s)] - n/2 \exp[-2(\log(\sigma) - \log(s))]\}$$

to give $g(z) = \exp\left[-nz - \frac{1}{2}\exp(-2z)\right]$, $h(\sigma) = \log(\sigma)$ and $t(\mathbf{y}) = \log(s)$.

We have assumed here that μ and σ are uncorrelated, although this is not in fact the case for the normal distribution. When there are several model parameters, prior specification is often simplified by assuming independence. Thus, in the cross-over model introduced in Section 2.3.2, we could write

$$p(\mu, t, p, \sigma^2, \sigma_s^2) = p_\mu(\mu) \times p_t(t) \times p_p(p) \times p_{\sigma^2}(\sigma^2) \times p_{\sigma_s^2}(\sigma_s^2).$$

Alternatively, a joint prior that takes account of the correlations between parameters can be obtained using a method proposed by Jeffreys (Jeffreys, 1961). Jeffreys' method is also helpful in situations where the likelihood cannot be arranged in the required form, $g(h(\sigma) - t(\mathbf{y}))$. It is the default prior used for the variance components when the `PRIOR` statement in `PROC MIXED` is used to carry out a Bayesian analysis. Details of Jeffreys' method and further discussion on setting non-informative priors can be found in Box and Tiao (1973, Section 1.3), Tanner (1996), Section 2.2.1) or Carlin and Louis (2008).

Properties for prior distributions

We now define some of the properties that are considered when setting priors and discuss their relevance in mixed models.

Proper priors Ideally, the prior distribution should integrate to one, and it is then described as a proper prior. A flat prior, $p(\theta) \propto c$, is not proper because the integral $\int_{-\infty}^{\infty} p(\theta) \mathrm{d}\theta$ does not exist no matter how small c is. Likewise, the non-informative prior suggested for variance components, $p(\sigma) \propto c/\sigma$, is also improper. However, the normal distributions with zero mean and very large variances suggested for fixed and random effects and the inverse gamma distributions with very small parameters for variance components are integrable and hence can be described as 'proper'. It is for this reason that these priors are sometimes preferred. However, as we noted earlier, when the spreads of these distributions are taken to their extremes, we obtain $\mathrm{N}(0, \infty) = \text{uniform}(-\infty, \infty)$ and $\mathrm{IG}(0, 0) = \text{reciprocal distribution}$ ($p(\theta) \propto c/\theta$), which are improper priors.

In practice, it is not always important for a prior to be proper. Provided the integral of the likelihood over all the model parameters is finite, there is not a problem.

In mixed models applications, the integral will often be finite even when improper priors are used. However, one situation where this may not be the case, which we will meet later, is when there are uniform fixed or random effects categories in generalised linear mixed models (these will be defined in Section 3.3.2).

Conjugacy When a prior distribution is conjugate, it leads to a posterior distribution that is of the same type as the prior. For example, a particular distribution called the beta distribution could be chosen as the prior for a binomial parameter, and this would lead to a posterior distribution that is also a beta distribution. The beta distribution is then described as conjugate to the binomial distribution. Likewise, conjugate prior distributions can be obtained for all other distributions from the exponential family.

When a model uses more than one parameter, a joint conjugate distribution should ideally be used. However, in practice, independence is often assumed between the parameters, and the joint prior is taken as the product of the conjugate priors for each parameter (obtained assuming the other parameters are fixed). For example, in setting a joint prior density for the normal distribution parameterised by μ and σ^2, the product of a normal prior density (conjugate for μ) and an inverse gamma prior density (conjugate for σ^2) could be taken if independence between μ and σ^2 was assumed.

We have already noted previously that distributions parameterised to have very large (but not infinite) spreads are often used to create proper non-informative priors. These distributions in fact are chosen as the conjugate distributions for each parameter. Hence, using normal prior distributions for fixed and random effects parameters is expected to lead to normal posterior distributions for the parameters. Likewise, inverse gamma prior distributions for variance components are expected to lead to inverse gamma posterior distributions. However, this is based on assuming independence of the parameters in setting the priors. In practice, the model parameters are not independent, and so these posterior distributions will not be obtained exactly.

2.3.5 Evaluating the posterior distribution

Evaluation of the posterior distribution and using it to obtain marginal posterior distributions for individual parameters rely heavily on integration. However, in most situations, algebraic integration is not possible, and a numerical method of integration is required.

The most popular methods of evaluating the posterior now rely on simulation as a means of performing the integration. Such methods can be described as *Monte Carlo* methods, and with increased availability of computer power over recent years, they have become much more feasible. Random samples from the joint distribution of all the model parameters are obtained. Each sample provides a set of values of the model parameters, (e.g. μ, t, p, σ^2 and σ_s^2 in our cross-over example). If we are interested in the marginal distribution of t, say, then we simply

ignore the other parameter values and consider only the randomly sampled values for t. If we take a sufficiently large number of samples from the joint posterior distribution, then we will be able to characterise the marginal distribution of t, to whatever level of detail we wish. Of course, as well as estimating the marginal distributions (usually our main purpose), we can use the full set of values of our model parameters to evaluate the full posterior distribution.

It is not usually possible to define a process for sampling directly from the posterior density posterior distribution, and various sampling approaches have been devised to overcome this difficulty. Some approaches are iterative so that samples are taken from the posterior distribution conditioned on the previous set of sampled values. Iterative approaches are often referred to as *Markov chain Monte Carlo (MCMC)* methods, the description 'Markov chain' being used because parameter values are sampled from a distribution depending only on parameter values sampled at the preceding iteration. Often, a simpler 'proposal distribution' is sampled in place of the true posterior distribution, and samples are accepted with probability proportional to the ratio of the true posterior density and the proposal density. We will now outline an MCMC algorithm known as the Metropolis algorithm, which is used by `PROC MCMC`.

The Metropolis Algorithm

Often, it is not easy to define a process for sampling directly from the true posterior density, $p(\boldsymbol{\theta}; \mathbf{y})$ ($\boldsymbol{\theta}$ = the vector of model parameters). One way to get around this difficulty is to define an alternative density for the model parameters, $g(\boldsymbol{\theta})$ (the *proposal density*), that is easier to sample. (This is often the multivariate normal or multivariate t-distribution). Parameters are sampled from the proposal density but only some are accepted. In practice, the method works by:

- Generate arbitrary initial values $\boldsymbol{\theta}_0$.
- Take a sample, $\boldsymbol{\theta}_{\text{new}}$, from the proposal distribution, $g(\boldsymbol{\theta}_0)$.
- Accept $\boldsymbol{\theta}_{\text{new}}$ as $\boldsymbol{\theta}_1$ only if a uniform variate, u, sampled from Uniform(0,1) is less than $\min\{\text{p}(\boldsymbol{\theta}_{\text{new}}; \mathbf{y})/\text{p}(\boldsymbol{\theta}_0; \mathbf{y}), 1\}$. (Thus, the new value for $\boldsymbol{\theta}$ is always accepted if it has a higher posterior density than the old value of $\boldsymbol{\theta}$, and the probability of acceptance becomes smaller as the ratio of the densities becomes smaller).
- Continue sampling, $\boldsymbol{\theta}_{\text{new}}$, from the proposal distribution, $g(\boldsymbol{\theta}_{\text{i}})$, and accept it as $\boldsymbol{\theta}_{\text{i}+1}$ only if a uniform variate, u, sampled from Uniform(0,1) is less than $\min\{\text{p}(\boldsymbol{\theta}_{\text{new}}; \mathbf{y})/\text{p}(\boldsymbol{\theta}_{\text{i}}; \mathbf{y}), 1\}$.

A large number of samples are taken, and their frequencies are expected to provide the full conditional distribution of the parameters. However, in practice, the procedure is not usually quite this simple. Although it is known that the simulated posterior will eventually converge to the true posterior as more and more samples are taken, it should be examined to obtain reassurance that it has converged based on the sample taken. This may be considered in two ways. Firstly by examining

plots of the simulated values, for example, a plot of their values across the iterations (often called a 'trace plot') and secondly by carrying out tests of convergence. More information on some of the diagnostic statistics and tests available, on interpreting diagnostic plots, and on appropriate action when convergence has not been obtained, will be given in Section 2.4.8, and the example in Section 2.5 will provide an illustration. However, we note that often there is not 100% certainty that the posterior distribution has been adequately sampled and has converged.

Using the simulated posterior to estimate parameters

The simulated samples provide the joint distribution for the model parameters. As we described earlier, the marginal distribution for any parameter is obtained by simply using the sampled values for that parameter. The sampled values can be used directly to estimate the mean, standard deviation, probability intervals and, if required, Bayesian *p*-values for selected parameters or their differences.

2.4 Practical application and interpretation

So far in this chapter, we have considered how to specify and fit mixed models. In this section, we look in some depth at points relating to the practical use of mixed models and their interpretation: negative variance components estimates (2.4.1); variance parameter accuracy (2.4.2); bias in fixed effects standard errors (2.4.3); significance testing (2.4.4); confidence intervals (2.4.5); model checking (2.4.6); and handling missing data (2.4.7). Readers may find that the material presented becomes most helpful once they have gained some experience of applying mixed models and have a need for considering particular aspects more closely. The worked example in Section 2.5 will illustrate many of the points made, and readers may find it helpful to read this section in conjunction with the example. Additional practical points relating specifically to covariance pattern and random coefficients models will be made in Sections 6.2 and 6.5.

2.4.1 Negative variance components

Variance components, by their definition, are non-negative. Nevertheless, the methods for estimating variance components will sometimes produce negative values that are not permissible. Such an estimate will usually be an underestimate of a variance component whose true value is small or zero. The chances of obtaining a negative variance component estimate when the true variance component is positive are increased when:

- The ratio of the true variance component to the residual is small.
- The number of random effects categories is small (e.g. a small number of centres in a multi-centre trial with centre effects fitted as random).

- The number of observations per random effects category is small (e.g. a two-period, cross-over trial with patients fitted as random).

It is not usually straightforward to calculate the probability with which a negative variance component will occur. However, it is possible when balance is achieved across the random effects (see Section 1.6), and there are an equal number of observations per random effects category. The probability of obtaining a negative variance component estimate is then obtained by reference to the F distribution:

$$P(\text{negative variance component}) = P(F_{\text{DF1,DF2}} > 1 + n\gamma),$$

where

$n =$ observations per category,
$\gamma =$ true variance component/residual variance,
DF1 = residual DF,
DF2 = effect DF.

The graphs in Figure 2.4 show how the probability of obtaining a negative variance component is affected by γ, the number of random effects categories, and the number of observations per category. When there are only a few observations per category (e.g. there are only two observations per patient in two-way, cross-over trials), there is a reasonable chance that a negative variance component may occur as an underestimate of a true positive variance component. However, the variance components can be constrained to be non-negative, which would lead to the patient variance component estimate becoming zero. This, in turn, will modify the residual variance estimate, relative to the unconstrained situation, in that the negative patient variance component will be absorbed, resulting in a lower residual variance. The residual variance from the unconstrained random effects model will be the same as fitting a fixed effects model.

How to handle a negative variance component

If a variance component is negative, the usual action would be either to remove the corresponding random effect from the model or to fix the variance component at zero (`PROC MIXED` sets negative variance components to zero by default). Either of these models will lead to the same parameter estimates for the fixed and random effects; however, the DF for significance tests will differ between the models. In some situations, the option of setting a negative variance component to zero, but retaining its DF, may be preferable. This would seem appropriate when an aspect of the study design is modelled by the random effect. For example, cross-over trials are designed to allow for patient effects, and so the DF corresponding to patients might be retained even if the patient variance component is fixed at zero. Similarly, if a multi-centre trial analysis produced a negative centre variance component, but a positive centre·treatment variance component, the centre effect DF might

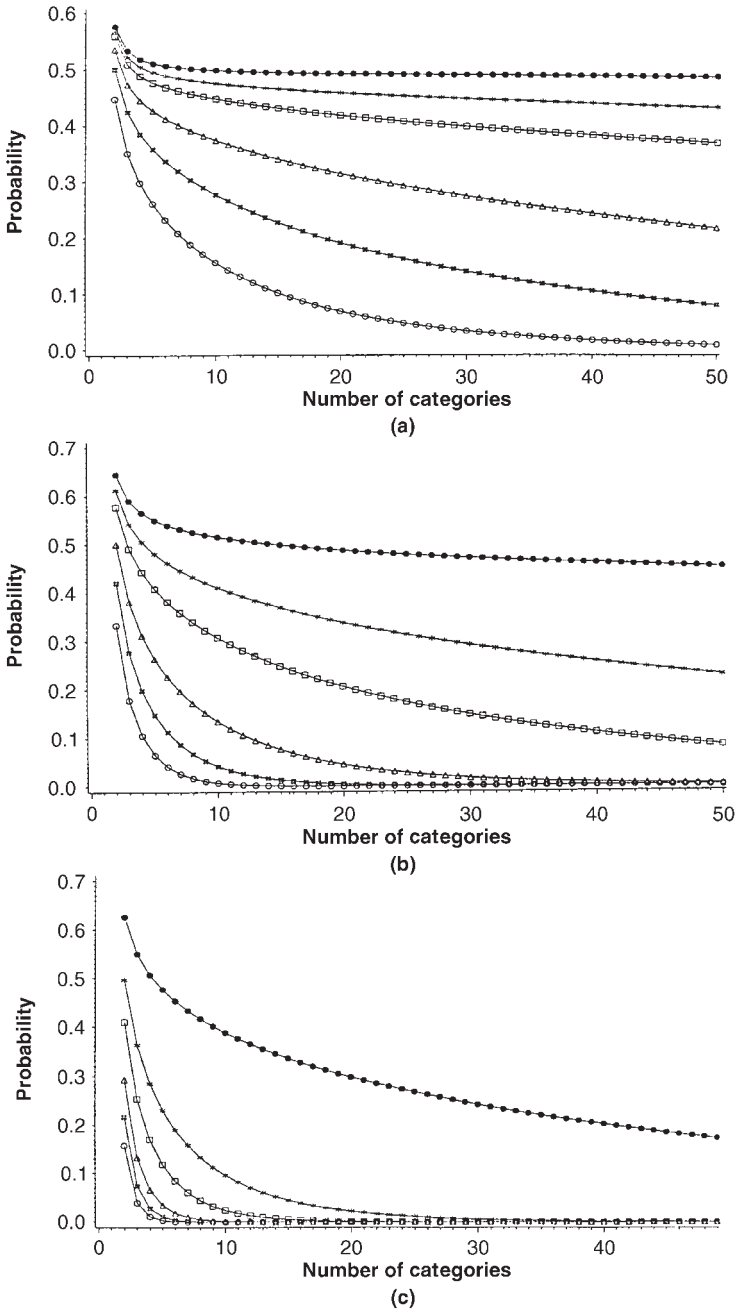


Figure 2.4 Probability of obtaining a negative variance component. (a) Observations per category = 2; (b) observations per category = 5; (c) observations per category = 25. Gamma: —·—·—0.01; ××××0.05; □□□□0.10; △△△△0.25; ✱✱✱✱0.50; O-O-O1.00.

be retained. However, if the random effect does not form part of the study design, then there is more reason to justify removing it from the model, hence excluding its DF.

We noted in the previous section that setting a negative variance component estimate to zero will lead to a different residual variance estimate to an equivalent fixed effects model. This is because a fixed effects model effectively allows negative variance components to occur (they are indicated whenever F is less than one). We will consider now the effect that this has on a cross-over trial with complete data. When the patient variance component estimate is positive, identical treatment effect estimates and standard errors will be obtained regardless of whether patients are fitted as random or fixed. However, when the patient variance component estimate is negative and set to zero in the random effects model, the residual variance will be smaller than that obtained in the fixed effects model. This, in turn, will lead to smaller treatment standard errors. We believe that the residual estimate from the random effects model is therefore preferable, since the intention of fitting patients is only to improve the precision of the treatment estimate, that is, we assume negative correlation within patients is implausible.

Modelling negative correlation

Usually, a negative variance component is an underestimate of a small or zero variance component. However, occasionally, it can indicate negative correlation between observations within the same random effects category. This is an unlikely scenario in most clinical trials; for example, it would be hard to imagine a situation where observations taken on the same patient could be more variable than those taken on different patients. However, in the following veterinary example, negative correlation is more feasible. Imagine an animal feeding experiment where animals are grouped in cages. In this case, it is possible that the greediest animals in a cage eat more than other animals, from a finite food supply, causing animal weight to become more variable within cages than between cages. In a model fitting cage effects as random, this would lead to a negative variance component for cage effects, indicating negative correlation between animal weights in the same cage.

To model this negative correlation, the model can be redefined as a covariance pattern model. In this model, the random effects (e.g. cages) are omitted, but correlation within the random effects is modelled by including covariance parameters in the residual variance matrix, $\mathbf{R}$. Thus, negative as well as positive correlation is allowed within the random effects (cages). We illustrate this model redefinition using the multi-centre data used to describe the mixed model in Section 2.1. Recall that a random effects model was specified with centre effects fitted as random and that the variance matrix, $\mathbf{V}$, was given by

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

$$= \begin{pmatrix} \sigma_c^2+\sigma^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2+\sigma^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2+\sigma^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2+\sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2+\sigma^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2+\sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2+\sigma^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2+\sigma^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2+\sigma^2 \end{pmatrix}.$$

When the model is redefined as a covariance pattern model, centre effects are excluded from the model, but covariance is allowed in the **R** matrix between observations at the same centre. A constant covariance can be obtained by using a compound symmetry covariance pattern, which expresses the **V** matrix as

$$\mathbf{V} = \mathbf{R} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix},$$

where ρ = the correlation between patients at the same centre.

Thus, **V** has an identical form to the random effects model except that it is parameterised differently. A negative covariance of observations at the same centre, $\rho\sigma^2$, is now permissible.

The Bayesian approach

When the Bayesian approach is used, negative variance components estimates are usually avoided by choosing a prior distribution for the variance components that is restricted to have positive values only. However, we have found that this can sometimes cause peaks in the posterior distribution for the variance components close to zero. Therefore, use of an estimator such as the median or expected value for variance parameters will be preferable to using the mode.

2.4.2 Accuracy of variance parameters

It is important that variance parameters are estimated with a reasonable accuracy because of their effect on the calculation of fixed effects and their standard errors.

The accuracy of the variance parameters is dependent on the number of DF used to estimate them. Although there are no hard and fast rules, it would seem inadvisable to fit an effect as random if less than about five DF were available for estimation (e.g. a multi-centre trial with five or less centres).

When an insufficient number of DF are available to estimate a variance parameter accurately, an alternative to resorting to a fixed effects model would be to utilise variance parameter estimates from a similar previous study. An approach that is specifically allowed for in `PROC MIXED` is to fix the variance parameters in the new analysis at their previous values. The fixed effects, $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, are then calculated using a fixed $\mathbf{V}$ matrix. However, this has the weakness of not utilising information on the variance parameters contained in the current study. A more natural approach, using both the previous variance parameter estimates and information in the current study, would be to use an informative prior for the variance parameters in a Bayesian analysis. This can be achieved by using the previous posterior distribution of the variance parameters as the prior distribution in the current analysis.

2.4.3 Bias in fixed and random effects standard errors

Fixed and random effects standard errors are calculated using a formula that is based on a known $\mathbf{V}$ (e.g. $\text{var}(\hat{\boldsymbol{\alpha}}) = (\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}$ for fixed effects). When data are balanced, the standard errors will not be biased. However, because $\mathbf{V}$ is, in fact, estimated, it is known that in most situations we meet in clinical trials, there will be some downward bias in the standard errors. Bias will occur when the data are not balanced across random effects, and effects are estimated using information from several error strata. In most situations, the bias will be small. It is most likely to be relevant when

- the variance parameters are imprecise;
- the ratio of the variance parameters to the residual variance is small; and
- there is a large degree of imbalance in the data.

However, there is not a simple way to determine how much bias there will be in a given analysis. Results from simulation studies for particular circumstances have been reported in the literature (e.g. Yates, 1940; Kempthorne, 1952; McLean and Sanders, 1988; Nabugoomu and Allen, 1994; Kenward and Roger, 1997). Although information from these studies is not yet comprehensive enough to allow any firm rules to be defined, they indicate that the bias may be 5% or more if the number of random effects categories relating to a variance parameter is less than about 10, *and* the ratio of the variance parameter to the residual variance is less than one. In these situations, a mixed model may not always be advisable unless an adjustment to the standard error is made.

Various adjustments for the bias have been suggested (e.g. Kacker and Harville, 1984; Kenward and Roger, 1997; Kenward and Roger, 2009).

Kenward and Roger's (1997) adjustment is available in `PROC MIXED`, but it occurs, surprisingly, as a DF option within the `MODEL` statement. Use of `DDFM=KENWARDROGER (or DDFM=KR)` inflates the estimated variance–covariance matrix of the fixed and random effects as described in their article, following which Satterthwaite-type DF are then computed based on these variances (see Section 2.4.4). Since the last edition a modified adjustment, `DDFM=KR(LINEAR)`, has become available in `PROC MIXED`. This is preferable for certain types of covariance pattern. More detail will be given in Section 6.2.4.

For models fitting covariance patterns in the **R** matrix (e.g. repeated measures models), an alternative 'robust' variance estimator using the observed correlations between residuals known as the *'empirical' variance estimator* (Liang and Zeger, 1986) has been suggested. It is calculated by

$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\text{cov}(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1},$$

where $\text{cov}(\mathbf{y})$ can be taken as $(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\alpha}})(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\alpha}})'$. This estimator takes into account the observed covariance in the data, and it is claimed that this may help alleviate some of the small sample bias. It does, however, cause the fixed effects variances to reflect observed covariances in the data rather than those specified by the covariance pattern modelled, and there is evidence against its use with small samples (see Long and Ervin, 2000). We discuss this further in Section 6.2.4. The empirical variance is calculated in `SAS` by using the `EMPIRICAL` option in the `PROC MIXED` statement.

Note that when the Bayesian approach is used (Section 2.3), exact standard deviations are obtained directly from the posterior distributions for each parameter and the problem of bias does not arise.

2.4.4 Significance testing

Testing fixed effects, random effects and random coefficients

Significance tests for fixed effects, random effects and random coefficients can be carried out using tests based on *F* or *t* distributions, as we will show. A test can be defined using a contrast; for example, $\mathbf{C} = \mathbf{L}'\widehat{\boldsymbol{\alpha}} = \mathbf{0}$ for fixed effects or $\mathbf{C} = \mathbf{L}'\widehat{\boldsymbol{\beta}} = \mathbf{0}$ for random effects/coefficients. For simple contrasts, **L** will just have one column. For example, in a trial comparing three treatments A, B and C, a pairwise comparison of treatments A and B is given by

$$\mathbf{L}'\widehat{\boldsymbol{\alpha}} = (0 \quad 1 \quad -1 \quad 0)\widehat{\boldsymbol{\alpha}} = \widehat{\alpha}_{\text{A}} - \widehat{\alpha}_{\text{B}}.$$

The first term corresponds to the intercept effect and the other three to the treatment effects (we assume treatments are the only fixed effects fitted). Alternatively, in multiple contrasts (e.g. to test overall equality of treatments), **L** will have several

columns. For example, equality of the three treatments might be tested using the multiple contrast

$$\mathbf{L}'\widehat{\boldsymbol{\alpha}} = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \widehat{\boldsymbol{\alpha}} = \begin{pmatrix} \widehat{\alpha}_{\mathrm{A}} - \widehat{\alpha}_{\mathrm{B}} \\ \widehat{\alpha}_{\mathrm{A}} - \widehat{\alpha}_{\mathrm{C}} \end{pmatrix}.$$

The F test statistic for testing the null hypothesis that the contrast is zero is calculated from a statistic known as the Wald statistic, which is given by

$$\begin{aligned} W &= (\mathbf{L}'\widehat{\boldsymbol{\alpha}})'(\mathrm{var}(\mathbf{L}\widehat{\boldsymbol{\alpha}}))^{-1}(\mathbf{L}'\widehat{\boldsymbol{\alpha}}) \\ &= (\mathbf{L}'\widehat{\boldsymbol{\alpha}})'(\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\alpha}}) \end{aligned}$$

for fixed effects. Thus, the Wald statistic can be thought of as the square of the contrast divided by its variance. For random effects and coefficients, $\widehat{\boldsymbol{\beta}}$ and $\mathrm{var}(\widehat{\boldsymbol{\beta}})$ are used in place of $\widehat{\boldsymbol{\alpha}}$ and $\mathrm{var}(\widehat{\boldsymbol{\alpha}})$ in these formulae. Note that if the Kenward–Roger adjustment is not used for $\mathrm{var}(\widehat{\boldsymbol{\alpha}})$, then $\mathrm{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}$ and W can be written as

$$W = (\mathbf{L}'\widehat{\boldsymbol{\alpha}})'(\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\alpha}}).$$

Asymptotically, W follows a chi-squared distribution with DF equal to the DF of $\mathbf{L}$ (number of linearly independent rows of $\mathbf{L}$). If $\mathbf{L}$ has a single row, then the significance test can also be presented as z, the contrast divided by its standard error, in which case z has a normal distribution. However, these distributions are derived on the assumption that there is no variation in the denominator term, $\mathrm{var}(\mathbf{L}'\widehat{\boldsymbol{\alpha}})$. They are equivalent to F or t tests with an infinite denominator DF. Thus, they will only be accurate if the DF of all error strata from which the effect is estimated are high (e.g. in an unbalanced cross-over trial, a high patient and residual DF are required).

A better option is to use the Wald F statistic, which is calculated by

$$F_{\mathrm{DF1,DF2}} = W/\mathrm{DF1},$$

where DF1 is the contrast DF (number of linearly independent rows of $\mathbf{L}$), and DF2 is the denominator DF corresponding to the DF of the contrast variance, $\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L}$. This statistic takes account of the fact that $\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L}$ is estimated and not known. The Wald t statistic for tests of single contrasts is given by

$$t_{\mathrm{DF2}} = (F_{1,\mathrm{DF2}})^{1/2} = W^{1/2}.$$

Wald F and t tests are produced by default in `PROC MIXED`.

The denominator DF for F tests This corresponds to those of the variance of the contrast, $\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L}$, and should reflect all the error strata from which the fixed effects have been estimated. When a fixed effect is balanced across random effects (see Section 1.6), it is estimated from only one error stratum, and the DF are simply those of this error stratum. However, when fixed effects are estimated from

two or more error strata, it is less straightforward to calculate the appropriate DF, and usually an approximation is used. A well-known approximation is given by Satterthwaite (1946):

$$\mathrm{DF} = 2(\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L})^2 \,/\, \mathrm{var}(\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L}).$$

It is equal to twice the variance of the contrast divided by the variance of the variance of the contrast. A further approximation is usually required to calculate $\mathrm{var}(\mathbf{L}'\mathrm{var}(\widehat{\boldsymbol{\alpha}})\mathbf{L})$. Giesbrecht and Burns (1985) show how Satterthwaite's approximation can be calculated to test single contrasts. `SAS` uses a generalisation of this technique to calculate Satterthwaite's approximation for single or multiple contrasts (obtained by using the option `DDFM=SATTERTH` in `PROC MIXED`). Alternatively, the inflated variance matrix suggested by Kenward and Roger (1997) (see Section 2.4.3) can be used in the Satterthwaite approximation by using option `DDFM=KR` in `PROC MIXED`. If software is not available for calculating the correct denominator DF, a conservative strategy would be to use an F test taking the lowest DF of the error strata used for estimating the contrast as the denominator DF. These DF are usually less than Satterthwaite's approximation to the true DF. For example, if in an unbalanced cross-over trial the patient DF were five and the residual DF were 10, then treatment effects could be tested using F tests with a denominator DF of five.

The text by Kenward and Roger (1997) or Elston (1998) is recommended reading for those who wish for more detailed knowledge on this subject.

Testing variance parameters

The significance of a variance parameter can be tested by using a likelihood ratio test to compare the likelihoods of models including (L_1) and excluding (L_2) the parameter. It is a standard result that if, under the null hypothesis, the additional terms in the model have no effect, the differences in the log likelihoods are distributed as $1/2\chi_1^2$. Hence,

$$2[\log(L_1) - \log(L_2)] \sim \chi_1^2.$$

The notation $\sim \chi_1^2$ is used to show that the likelihood ratio test statistic has a chi-squared distribution with one degree of freedom (DF). In general, χ_n^2 denotes a chi-squared distribution with n DF.

More discussion of methods for testing variance parameters can be found in the text *Linear Mixed Models for Longitudinal Data* by Verbeke and Molenberghs (2000). The authors suggest that because variance parameters are truncated to be positive, a 50:50 mixture of χ_0^2 and χ_1^2 distributions is more appropriate for the test than the χ_1^2 distribution stated previously. The χ_0^2 distribution is not a distribution in the usual sense and is described by Verbeke and Molenberghs as 'the distribution which gives probability mass 1 to the value 0'. It is used to represent the part of the distribution under the null hypothesis that would be appropriate

for truncated negative variance components. The mixture distribution results in a *p*-value given by

$$p = \frac{1}{2}p(\chi_0^2) + \frac{1}{2}p(\chi_1^2) = \frac{1}{2}p(\chi_1^2)$$

and has the effect of halving the *p*-value obtained by the standard likelihood ratio test, hence making the test more liberal. However, often, the decision to include random effects and their corresponding variance component in the model will be decided by the underlying structure of the data and the expected sources of random variability rather than on the results of a significance test.

In covariance pattern models, there is often a choice of covariance structures available, and interest usually lies with testing whether a particular covariance pattern causes a significant improvement over another pattern, rather than with testing a single variance parameter. Likelihood ratio tests can again be applied when the models are 'nested'. Further detail will be given in Section 6.2.4. In random coefficients models, the inclusion of a random coefficient (e.g. random slopes for patients) will lead to more than one additional covariance parameter because there is a correlation between the random effects and slopes, which will increase the DF used for the chi-squared tests. Further detail will be given in Section 6.6.2.

The Bayesian approach

When a Bayesian approach is used, exact 'Bayesian' *p*-values can be obtained for all model parameters (fixed effects, random effects and variance parameters) from the posterior distribution, and there are no potential inaccuracies in obtaining a test statistic or the DF for its distribution. A *p*-value to test the null hypothesis that a parameter is zero is obtained as the probability of being outside the probability interval, which has zero on the boundary (see Section 2.3.3). This corresponds to a two-sided, 'classical' *p*-value. However, note that such tests are not available for variance components if a prior distribution with a positive range has been used (e.g. the reciprocal or inverse gamma distribution).

2.4.5 Confidence intervals

The reporting of confidence intervals relating to the fixed effects of interest is, of course, of great importance. The usual classical approach can be applied, but it is important to ensure that appropriate DF are used (see Section 2.4.4):

$$\text{Lower 95 \% confidence limit} = \text{mean effect} - t_{\text{DF},0.975} \times \text{SE},$$
$$\text{Upper 95 \% confidence limit} = \text{mean effect} + t_{\text{DF},0.975} \times \text{SE}.$$

Alternatively, probability intervals can be obtained from a Bayesian analysis (see Section 2.3.3).

2.4.6 Checking model assumptions

In this section, model checking methods are considered for the random effects model where it is assumed that the residuals, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}} - \mathbf{Z}\hat{\boldsymbol{\beta}}$, and also the random effects, $\boldsymbol{\beta}$, are normally distributed about zero and are uncorrelated. In random coefficients models and covariance pattern models, the residuals and random coefficients are correlated, and model checking methods addressing this feature will be considered in Sections 6.2 and 6.5. We now consider the normality assumption separately for the residuals and random effects.

Residuals

The normality of the residuals may be checked using residual and normal plots in the same way as for fixed effects models. A plot of the residuals against their corresponding predicted values can be used to

- provide a rough check of normality of residuals,
- check whether the residual variance is constant across observations,
- look for outliers.

The predicted values corresponding to the residuals are taken as $\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}}$. We choose the predicted values rather than the expected values, $\mathbf{X}\hat{\boldsymbol{\alpha}}$, because one of the features we wish to check against in the plot is whether the size of the residual is associated with the magnitude of the underlying value, which should, of course, reflect both fixed and random effects.

The assumption of normality can be checked more carefully using normal probability plots (i.e. plotting the ordered residuals against their values expected from the standard normal distribution given their ranks, see Snedecor and Cochran, 1989, Section 4.13). If the data are normally distributed, then the residuals will roughly form a straight line on the normal plot. If the plotted data deviate markedly from a straight line, then it is likely that the data are not normally distributed. For example, a bow shape indicates a skewed distribution, and a sigmoid shape indicates a symmetrical but non-normal distribution.

Homogeneity of the residuals can be further assessed by comparing the variances of each set of residuals between fixed effects categories (e.g. between treatments).

An illustration of the use of residual plots is given in the worked example in Section 2.5.

If a general lack of normality is indicated in the residuals, a transformation of the data can be considered. Alternatively, if a residual plot indicates outlying values, then checks should first be made to determine any possible reasons for them being extreme. Plots of standardised residuals (e.g. studentised or Pearson residuals) can help to assess whether the observation is a genuine outlier or is outlying by chance. These residuals take account of the fact that the observed residuals have differing variances given by the diagonals of $(\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}$. When a value is clearly wrong (e.g.

recording error, punching error, machine error), it should be corrected if possible or else removed from the analysis. An outlier will not necessarily have as much effect on the parameter estimate; so, if there is no clear reason for removing it, its influence should be assessed. This can be carried out by calculating influence statistics (e.g. Cook's D statistic). `PROC MIXED` (Version 9) contains options for producing influence statistics. By reanalysing the data with the outlier removed, we can determine whether parameter estimates alter noticeably. If the estimates are similar, then the outliers can be retained. If they differ, then it would seem sensible to remove them, provided there is a convincing external reason for doing so (e.g. measurement suspected to be inaccurate because a blood sample was clotted; centre did not follow protocol). If there is no apparent reason for the outlier, then two alternative analyses and sets of conclusions may need to be presented. We also note that another alternative would be to construct a robust method to reduce the influence of the outlier; however, we will not be considering these methods. Huber and Ronchetti (2009) give an introduction to robust methods.

Random effects

Checking the normality of the random effects is less straightforward than for the residuals. It has been suggested that fixed effects and variance components estimates are unlikely to be sensitive to non-normality of the random effects (Verbeke and Molenberghs, 2000, Section 7.8); therefore, for most of the examples we consider it will not be important to check the normality of random effects. However, when random effect predictions (e.g. estimates of the treatment effect at individual centres in a multicentre trial) are of interest, it is important to be aware that they are, in some situations, sensitive to any misspecification of their distribution. For example, Verbeke (1995) and Verbeke and Lesaffre (1996) examine a situation in which the underlying random effects distribution is a mixture of normal distributions, and the variance component is small compared to the residual variance. They found that plots of the random effects predictions did not show up non-normality, despite their underlying distribution being clearly non-normal. In previous editions of this text, we proposed checking the normality of the random effects using plots of their predicted values, $\widehat{\boldsymbol{\beta}}$, against their predicted means, along with corresponding normal plots, with predicted means obtained as the means of the expected values, $\mathbf{X}\widehat{\boldsymbol{\alpha}}$, within each random effect category (e.g. calculate the means of $\mathbf{X}\widehat{\boldsymbol{\alpha}}$, within each centre). These plots are still helpful for identifying outliers (e.g. an outlying centre, see example in Section 2.5), which may affect particularly the standard errors of fixed effects but will not always show up an underlying non-normality. The influence of any outliers detected may be assessed by analysing data with the relevant random effect group removed. When there is interest in the random effects themselves, however, their predictions should be interpreted cautiously, as they are more sensitive to deviations from normality, and plots of random effect predictions will not give full assurance of normality. Recently, Verbeke and Molenberghs (2013) have proposed

an alternative graphical method to check the normality of random effects and, we understand, are currently developing a diagnostic test. These checks will provide a more robust means to check normality but may not be straightforward to implement until introduced into readily available software such as `PROC MIXED`. Verbeke and Molenberghs (2013) also provide an overview of other research, considering the implications of non-normality of the random effects.

2.4.7 Missing data

Mixed models are much more flexible than fixed effects models in the treatment of missing values. For example, in a two-period, cross-over trial, information on subjects with one value missing is completely lost when a fixed effects analysis is used. In contrast, mixed models are capable of handling the imbalance caused by missing observations, provided that they are missing at random. This is often a reasonable assumption to make. If a subject withdraws from a cross-over trial after receiving one treatment, then we may have no idea of how the subject would have responded to the other treatments, and to handle these non-observed periods as if they were missing at random seems eminently sensible. It is helpful at this stage to classify missing data into one of three widely used categories.

Missing completely at random

As the name suggests, observations are missing completely at random if the probability of an observation being missing is the same for all potential data points, and the probability of 'missingness' is unaffected by whether other observations are missing. In practice, datasets where data are missing completely at random will be uncommon. It is most recognisable if the 'missingness' is due to causes that are unpredictable and can be viewed in a broad sense as accidental. The loss by accidental spillage of a blood sample prior to its analysis would be seen as an observation missing completely at random. Similarly, patient non-attendance at a follow-up visit because of a family bereavement could readily be regarded as a cause for an observation to be missing completely at random. Most missing value situations are, however, less clear cut than these two examples. Tests of randomness have been suggested by Diggle (1989) and Ridout (1991), and different mechanisms for missing data are described by Little and Rubin (2002). If the application of mixed models required that any missing data were missing completely at random, their applicability would be severely curtailed. Fortunately, such a strict definition is not necessary, and the more relaxed requirement that observations be *missing at random* is sufficient.

Missing at random

The requirement for an observation to be missing at random is that its expected value should be unaffected by whether the observation is missing. One of the

most common reasons for missing values is patient non-compliance. If a patient decides that continued participation in a trial is just too much effort or the patient dislikes the clinical procedures and that decision is unrelated to any change in the outcome variables for the study, then the missing values will be missing at random. This example is uncontroversial, but a more contentious example occurs if the patient withdraws from the trial because of a perceived adverse reaction to treatment. If the adverse reaction is unrelated to the outcome variables, then the missing data with respect to these efficacy variables can be regarded as missing at random. In taking this approach, we are, in effect, attempting to estimate the effects of treatment on our outcome variables if adverse events do not occur. Of course, we do not ignore this important effect of treatment, and analysis of adverse events will form an important part in the evaluation of any clinical trial. By employing this philosophy in analysis, we are seeking to separate the effect of the treatments on the efficacy outcomes of interest from the assessment of the treatments' tolerability. It is probably the most usual way of dealing with patient withdrawal, but it will not always be the method of choice. For example, a pragmatic approach to a holistic evaluation of treatment would be to regard drug intolerability as a failure of treatment and predefine an appropriate value of the response variable for substitution when such a 'failure' occurs.

A decision as to whether missing values can be legitimately regarded as missing at random is rarely obvious. The advantages to the analysis if this assumption can be made are substantial, however, and so in practice, this assumption is usually made unless there are strong grounds for concluding that the missing values are not missing at random.

Missing not at random

It will sometimes be clear that the censoring mechanism is not acting at random. If a patient withdraws from the trial because of clinical deterioration, then to assume that subsequent observations were missing at random would be totally inappropriate. Under such circumstances, a variety of methods are employed in an attempt to reduce the bias due to poor responders having missing values. None of the methods is universally applicable, but the most widely applied is the 'last-value-carried-forward' approach. In this method, the last observed value of the response variable is substituted for every subsequent missing observation. If the anticipated pattern of response to the intervention is of improving measurements, then this may be an effective way to minimise bias, but in some circumstances, unthinking application of 'last-value-carried-forward' can worsen the bias from withdrawal of poor responders. This would be the case in trials of treatments for chronic bronchitis where the aim is to prevent further deterioration in lung function. The pattern of unchanging observations arising from 'last-value-carried-forward' would artificially generate a 'good' outcome from a patient who has withdrawn due to a poor response – not the type of imputation one would hope for.

Handling missing data is, however, a substantial topic in its own right, and we will not consider it further in this book. The text *Missing Data in Clinical* Studies by Molenberghs and Kenward (2007) is recommended for readers wishing to gain more knowledge in this area.

It is important to be aware that mixed models do not overcome the problems caused by missing independent variables, for example, when a fixed or random effects category is unknown or a baseline value missing. As with a fixed effects model, when this occurs the observation will be automatically deleted from the analysis unless a value is imputed to denote the missing effect.

2.4.8 Determining whether the simulated posterior distribution has converged

When a Bayesian analysis is carried out, the simulated posterior should be examined to obtain reassurance that it has converged, that is, formed a stationary distribution. Several approaches are available for assessing convergence.

Various plots of the simulated values for each parameter may reveal problems with convergence. A plot of the sampled parameter values against the iteration number (often called a 'trace plot') shows how quickly the simulation process has settled to a range of values and whether the values stay in this range. If the simulated values take a while to settle to a range of values, it would be advisable to omit the first set of samples. The omitted samples are then often described as 'burn-in' samples.

An 'autocorrelation' plot shows the correlation of simulated values depending on how widely they are separated. If widely separated values are correlated, it is an indication that a larger sample size is needed. However, sometimes, the increase will cause the sample to become unmanageably large. In this situation, it is often satisfactory to use only every 1 in N of the samples to construct the posterior distribution. This process is known as 'thinning'. However, we note that it is not essential to use thinning when the sample size is increased as long as the sample is manageable.

A histogram of the simulated values is helpful to indicate the shape of the simulated posterior. If this is not smooth, it is often a sign that more samples are required.

Summary statistics such as the correlation of parameter samples by time lag, and the estimated time lag for parameter samples to become uncorrelated, will indicate if correlation is high between the sampled parameter values and again suggest a larger sample size. It can be helpful too to estimate the 'effective sample size' after taking into account the amount of autocorrelation. It is possible to determine the proportion of the posterior variability that is due to the simulation – a small proportion is a good indication that convergence has been achieved.

Another approach to assessing convergence is by using statistical tests. These look for particular symptoms that may indicate that the iterations have

not converged to a stationary process, which will provide the true parameter distributions. Some of the tests available are:

- Geweke test – Compares mean estimates from early and late samples using a Z test.
- Gelman–Rubin – Compares results of several sets of simulations to see if they all converge to the same distribution.
- Heidelberger–Welch – Tests whether the simulated posterior has become stationary.
- Rafferty–Lewis – Evaluates the accuracy of the percentiles.

Checking convergence has some similarities to model assumption checking in that a degree of judgement is needed. It is rarely possible to be 100% certain that either convergence is achieved or model assumptions are met.

Use of these methods will be illustrated with the aid of the example in Section 2.5.

2.5 Example

In Section 1.2, we introduced a multi-centre trial of treatments for hypertension. In this trial, a new hypertensive drug (A) was compared with two standard drugs (B and C). In this section, we will consider analyses of the trial in greater detail and cover some of the practical points made in the previous section. The `SAS` code used for each model will be given at the end of the section. We will follow this pattern of supplying `SAS` code following each example throughout the book.

2.5.1 Analysis models

The main response variable in the trial (DBP at the final visit) will be analysed. The last post-treatment visit attended is used for patients who do not complete the trial, forming an 'intention-to-treat' analysis. Analyses were carried out using the models listed in below. Initial DBP (baseline) was included as a covariate in all models to reduce between-patient variation.

Model	Fixed effects	Random effects	Method
1	Baseline, treatment, centre	–	OLS[a]
2	Baseline, treatment	Centre	REML
3	Baseline, treatment	Centre, treatment·centre	REML
4	Baseline, treatment	Centre	Bayes
5	Baseline, treatment	Centre, treatment·centre	Bayes

[a]OLS = ordinary least squares. This is the fixed effects method used by `PROC GLM` (see Section 2.2.2).

Fixed effects models (Model 1)

The data were analysed first using the more conventional fixed effects approach. Initially, a full model including baseline, treatment, centre and centre·treatment effects was fitted to test whether centre·treatment effects were significant. They were found to be non-significant ($p = 0.19$) and were therefore omitted from the model to give Model 1. This is the usual action taken when a fixed effects approach is used, since inclusion of fixed centre·treatment effects will cause the overall treatment estimate to be an unweighted average of the individual centre estimates. However, note that in this instance, it would not have been possible in any case to estimate overall treatment effects when centre·treatment effects were included because all the treatments were not received at every centre.

When centre·treatment effects are included in a fixed effects analysis, this causes a separate treatment effect, specific to each centre included in the trial, to be fitted. The inferences strictly apply only to those centres that were included in the trial. If the model that omits the centre·treatment effect is used, the inferences again should still strictly apply only to the centres used. However, extrapolation of the inferences regarding the effect of treatment to other centres seems more reasonable when an assumption has been made that the treatment effect does not depend on the centre in which it has been applied.

Random effects models fitted using REML (Models 2 and 3)

In Model 2, centre effects are fitted as random and baseline and treatment effects as fixed, using REML. With this model, treatment effects are estimated using information from the residual error stratum and, additionally, from the centre error stratum, since balance does not occur across random effects (centres). By comparing the results with those from Model 1, it is possible to determine whether any additional information has been recovered on treatments from the centre error stratum. Note that only under the strong assumption that there is no centre·treatment interaction will inferences apply to the population of centres from which those in the trial can be considered a sample.

In Model 3, both centre and centre·treatment effects are fitted as random, and baseline and treatment effects as fixed, using REML. Unlike the fixed effects approach, centre·treatment effects are retained in the model, provided their variance component is positive. The logic here is that use of the model implies a belief that treatment effects may differ between centres, and we wish this to be included in our estimates, even if the interaction is non-significant. Since centre·treatment effects are taken as random in this model, treatment effects are assumed to vary randomly between the centres, and results can be related with some confidence to the wider population of centres.

Bayesian models (Models 4 and 5)

Models 4 and 5 are the same as Models 2 and 3, except that they are fitted using the Bayesian approach. They were fitted by taking 100,000 samples of the model

parameters using the Metropolis algorithm in `PROC MCMC` and a thinning factor of 5 (i.e. only one in five samples) to construct the posterior (see Section 2.3.5). Frequencies of the sampled values were used directly to obtain the marginal distribution for each parameter. Fixed effects estimates were obtained by simply calculating the means and standard deviations of the sampled parameter values. Treatment parameters were obtained directly for A – C and B – C. Samples for the treatment difference A – B were obtained by calculating the differences between the A – C and B – C samples.

The variance components are of less direct interest in this example. However, they can be estimated by taking statistics such as the mode, median or mean of the sampled parameters (see Section 2.3.3). Here we obtain an idea of their size by taking the medians of their posterior distributions. These estimates will not be exactly the same as the REML estimates because the posterior density is not exactly proportional to the likelihood function (since a flat prior is not used) and also because the REML estimates correspond to modes rather than to medians. A measure of the spread of the variance components can be obtained by calculating probability intervals for the sampled values (see Section 2.3.3).

2.5.2 Results

The results obtained from each model are shown in Table 2.1. The variance components for centres and treatment centres are much smaller than the residual in all models. This indicates that most of the variation in the data is due to differences between patients and not to differences between centres. The Bayesian probability intervals for the variance components illustrate the skewness of their distributions. These probability intervals also highlight the inadequacy of the confidence intervals produced by the REML analyses, which have very different ranges for the centre and centre·treatment variance components.

The treatment standard errors are slightly smaller in Model 2 than in Model 1, indicating that only a small amount of information on treatments has been recovered from the centre error stratum. This is mainly because there was only a small degree of imbalance in the allocation of treatments within centres. We note, however, that the treatment comparisons involving B have been modified appreciably. When the treatment·centre interaction is included as a random effect (Models 3 and 5), the treatment effects differ from Models 1, 2 and 4 because the correlation between patients on the same treatments and at the same centres is taken into account, causing a different weighting of the observations. The treatment standard errors are increased in Models 3 and 5, reflecting the positive centre·treatment variance component. The amount of this increase is related to the size of this variance component and to the number of centres used. Although the centre·treatment variance component is small compared with the residual, its influence on the standard error is noticeable because it is related inversely to the number of centres, and not to the number of observations. (The relationship

Table 2.1 Estimates of variance components and fixed effects.

Model	Fixed effects	Random effects	Method
1	Baseline, treatment, centre	–	OLS
2	Baseline, treatment	Centre	REML
3	Baseline, treatment	Centre, treatment·centre	REML
4	Baseline, treatment	Centre	Bayes
5	Baseline, treatment	Centre, treatment·centre	Bayes

		Treatment effects (SEs)		
Model	**Baseline**	**A – B**	**A – C**	**B – C**
1	0.22 (0.11)	1.20 (1.24)	2.99 (1.23)	1.79 (1.27)
2	0.22 (0.11)	1.03 (1.22)	2.98 (1.21)	1.95 (1.24)
3	0.28 (0.11)	1.29 (1.43)	2.93 (1.41)	1.64 (1.45)
4[a]	0.28 (0.11)	1.03 (1.23)	3.00 (1.22)	1.97 (1.23)
5[a]	0.26 (0.11)	1.18 (1.40)	2.93 (1.37)	1.75 (1.42)

	Variance components (95% confidence or probability intervals)		
Model	**Centre**	**Treatment·centre**	**Residual**
1	–	–	71.9 (60.9 – 86.2)
2	7.82 (3.52–29.74)	–	70.9 (60.2 – 84.7)
3	6.46 (2.41–46.05)	4.10 (0.88–1515.83)	68.4 (57.2 – 83.2)
4[b]	7.41 (1.11–16.95)	–	71.4 (60.1 – 84.6)
5[b]	5.40 (0.00–15.05)	1.80 (0.00–15.73)	69.7 (57.6 – 83.0)

[a]Strictly speaking, the SEs in these Bayesian models are parameter standard deviations.
[b]Medians are given with 95% HPD probability intervals.

between the treatment effect standard errors and the variance components in a multi-centre analysis is defined more precisely in Section 5.2.)

The fixed effects and standard errors from Model 2 fitted using REML are similar to those from Model 4 using a Bayesian analysis. However, there is a noticeable difference in the estimates of A – B and B – C between Models 3 (REML) and 5 (Bayesian analysis) where random centre·treatment effects have been included, although the corresponding standard errors are similar. The differences in the effect estimates are likely to be due to an unusual shape for the posterior density, causing the mean parameter value from the Bayesian analysis (Model 5) to differ from the value maximising the likelihood as provided by the REML analysis (Model 3). It is unclear which estimates are preferable, and more research is needed to study situations where parameters estimates differ between the approaches and to recommend the preferred estimate.

The similarity between the fixed effects standard errors from the REML and Bayesian analyses is reassuring, given there are inherent differences in the

'standard errors' obtained in the Bayesian analysis. They are based directly on the simulated distribution of $\boldsymbol{\alpha}$, which takes into account the distribution of the variance parameters. As such, they are related directly to the probability intervals. In contrast, the REML analysis calculates standard errors using the Kenward–Roger approximation (Kenward and Roger, 1997), which helps to alleviate bias. Confidence intervals are then obtained using multiples of the standard errors, based on the *t* distribution with appropriate DF (calculated using the Kenward–Roger approximation). Thus, if confidence intervals and probability intervals are in agreement, we should only anticipate similar 'standard errors' from the two approaches when the DF with the REML approach are relatively large. When the DF are small, the standard error from the REML analysis will be less than the corresponding parameter standard deviation in the Bayesian analysis. This point is considered further in the following section.

2.5.3 Discussion of points from Section 2.4

Each of the practical points covered in Section 2.4 is discussed in the following sections in relation to the example.

Negative variance component estimates (2.4.1)

Negative variance component estimates were not obtained by any model in this example.

Variance parameter accuracy (2.4.2)

Since 29 centres were used in this study, we would not expect to be concerned that insufficient DF were available for estimating the variance components. The wide confidence intervals obtained for Models 2 and 3 partly reflect the inadequacy of the methods for determining these limits. However, it is surprising that there are also wide probability intervals in the Bayesian analyses (Models 4 and 5).

Bias in fixed effects standard errors (2.4.3)

Since 29 centres were used, we would expect the weightings (derived from the variance components) used to estimate the fixed effects and their standard errors also to be fairly accurate. An indication of the amount of standard error bias present in the two mixed models analyses considered in Section 2.5.2 can be obtained by comparing the model-based standard error (with no bias correction) to the standard error corrected by the Kenward–Roger adjustment (1997). In Model 2, the standard errors are almost identical, and the bias is negligible. This is likely to be because most of the information on treatments is estimated from the residual error stratum. However, in Model 3, the adjusted standard errors are about 3% higher than the model-based values. More bias has occurred in this model because treatment effects are estimated mainly from the centre·treatment error stratum.

Method	Model 2: centre random (model based, KR)	Model 3: centre, centre·treat random (model based, KR)
A – B	1.221, 1.221	1.391, 1.430
A – C	1.211, 1.211	1.373, 1.411
B – C	1.239, 1.240	1.405, 1.445

Significance testing (2.4.4)

We demonstrate the use of significance tests for Models 3 and 5, showing their calculation from both likelihood-based and Bayesian results. In Model 3, the significance of treatment differences can be assessed with Wald F tests using the Kenward–Roger adjustment for the fixed effects variances and their corresponding denominator DF. An $F_{2,25}$ value of 2.16 was obtained for the composite test of treatment equality which is non-significant ($p = 0.14$). Test statistics for pairwise treatment comparisons gave

$$\begin{aligned} \text{A} - \text{B} \quad & t_{23.8} = 0.90, \quad p = 0.38, \\ \text{A} - \text{C} \quad & t_{25.6} = 2.07, \quad p = 0.048, \\ \text{B} - \text{C} \quad & t_{25.7} = 1.14, \quad p = 0.27. \end{aligned}$$

Thus DBP was shown to be significantly lower on treatment C than on treatment A. The denominator DF differ slightly for each comparison owing to the differences in the distribution of the treatments across centres.

In Model 5, exact p-values are obtained by calculating the proportion of sampled parameters with a value of less than (or greater than) zero and then doubling the smaller of these values to obtain a two-sided p-value. The p-values obtained in this way are as follows:

$$\begin{aligned} \text{A} - \text{B} \quad & p = 0.39 \\ \text{A} - \text{C} \quad & p = 0.03 \\ \text{B} - \text{C} \quad & p = 0.22 \end{aligned}$$

These differ a little from the REML p-values but lead to the same conclusions.

Confidence intervals (2.4.5)

The 95% confidence intervals for the differences between treatments using Model 3 are calculated by SAS using t distributions with Satterthwaite's DF. These are compared in the following table to the 95% probability intervals obtained using the equal tails method from Model 5 where a Bayesian analysis was used.

	Model 3	*Model 5*
A – B	(−1.59, 4.15)	(−1.51, 4.02)
A – C	(0.10, 5.75)	(0.23, 5.61)
B – C	(−1.25, 4.53)	(−1.11, 4.47)

The width of the intervals is slightly smaller for the Bayesian model. This may be due partly to the smaller centre·treatment variance component estimated for this model. The differences in location of the estimates (discussed in main results) between the models is also reflected in the intervals.

Model checking (2.4.6)

Checks of model assumptions are carried out for Model 3 as an illustration of the techniques that can be used. We present plots of residuals and random effects against their predicted values, as well as normal plots. In addition, the homogeneity of the variance components is checked by calculating the variances of the residuals and centre·treatment effects by treatment.

Residuals Residual and normal plots are shown in Figure 2.5(a). These indicate no general lack of normality but one possible outlier with a residual of 40 (patient 314, visit 5). In practice, some statisticians would be happy to accept this degree of variation in the residuals. To illustrate the magnitude of the effect that an outlier can have, we have examined the effect of removing this observation. The graphs shown in Figure 2.5(b) indicate the difference in effect estimates that would result from removing each observation or by manually removing the outlier and reanalysing the dataset. When the outlying patient was removed from the analysis, the parameter estimates changed noticeably from their original estimates:

	Variance components		
Model	**Centre**	**Treatment·centre**	**Residual**
With outlier	6.86	4.10	68.4
Without outlier	6.97	1.51	63.3
	Treatment effects (SEs)		
Model	**A – B**	**A – C**	**B – C**
With outlier	1.29 (1.43)	2.93 (1.41)	1.64 (1.45)
Without outlier	0.74 (1.27)	2.49 (1.26)	1.76 (1.29)

On closer examination, it was likely, but not certain, that the DBP value (of 140) was due to a recording mistake. This is an extremely high reading, which is inherently unlikely, but the patient's baseline DBP was 113 mm Hg, and he had dropped out of the trial at visit 5 due to an 'insufficient effect'. Under these circumstances, we might wish to report results from analyses both including and excluding the outlier. The large changes in parameter estimates caused by the exclusion of one patient in a reasonably large trial illustrate the potential importance of at least basic checks of model assumptions.

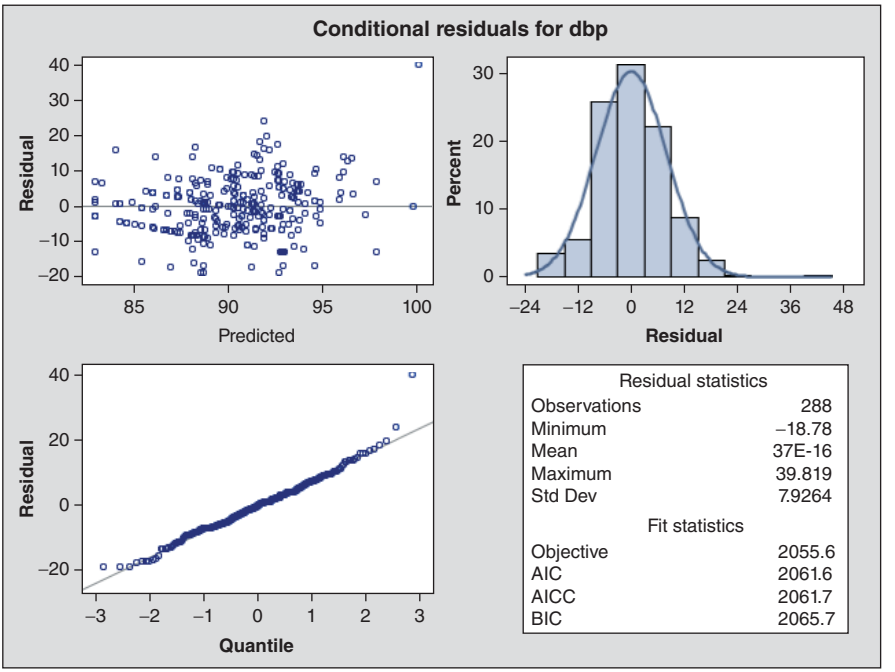


Figure 2.5 (a) Residual and normal plots.

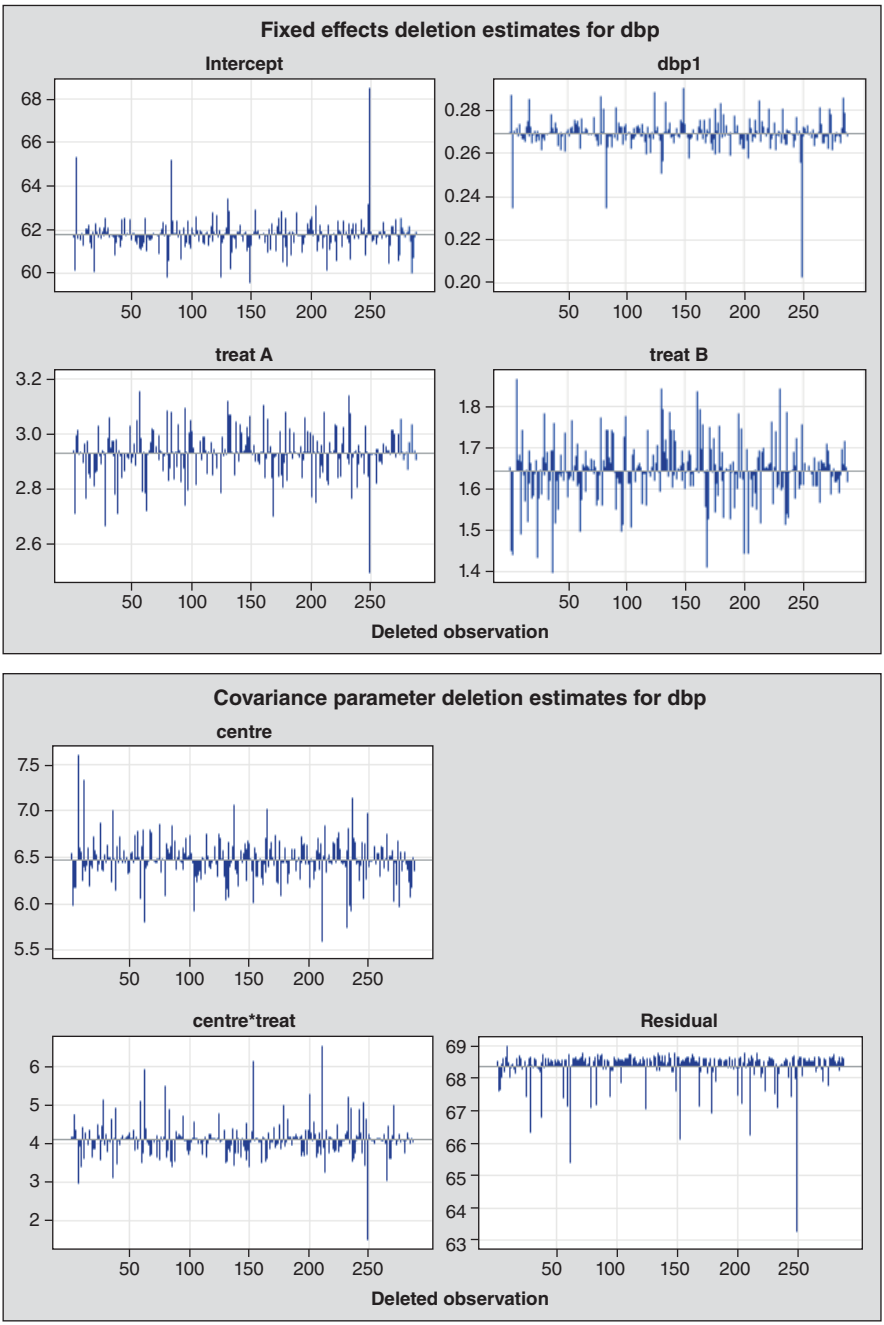


Figure 2.5 (b) Effect of removing each observation on model parameters.

Centre effects Predicted values are obtained for each centre by taking the means of the predicted observations, $\mathbf{X}\hat{\boldsymbol{\alpha}}$, for that centre. The plots (Figure 2.6(a)) indicate one possible outlying centre (centre 31). The influence of removing each centre is shown in Figure 2.6(b). However, note that the x axis in these plots does not relate directly to the centre numbers, and the corresponding SAS output listing the influence of each centre needs to be consulted. When the analysis was repeated with centre 31 removed, the following results were obtained:

	Variance components (SEs)		
Model	**Centre**	**Treatment·centre**	**Residual**
With centre 31	6.46	4.10	68.4
Without centre 31	0.81	5.37	73.1
	Treatment effects (SEs)		
Model	**A – B**	**A – C**	**B – C**
With centre 31	1.29 (1.43)	2.93 (1.41)	1.64 (1.45)
Without centre 31	1.33 (1.58)	3.03 (1.56)	1.70 (1.60)

The variance component estimates have changed fairly noticeably from their previous values. However, the exclusion of this centre did not greatly change the treatment estimates, although their standard errors are increased. Thus, the centre could be retained in the analysis with reasonable confidence. However, possible reasons for the centre being outlying should be investigated, at least to check that the protocol was followed correctly.

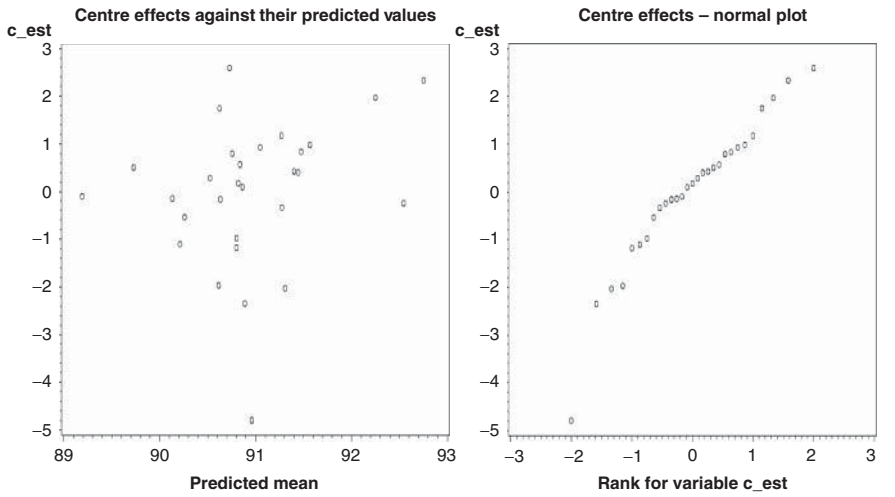


Figure 2.6 (a) Plots of centre effects.

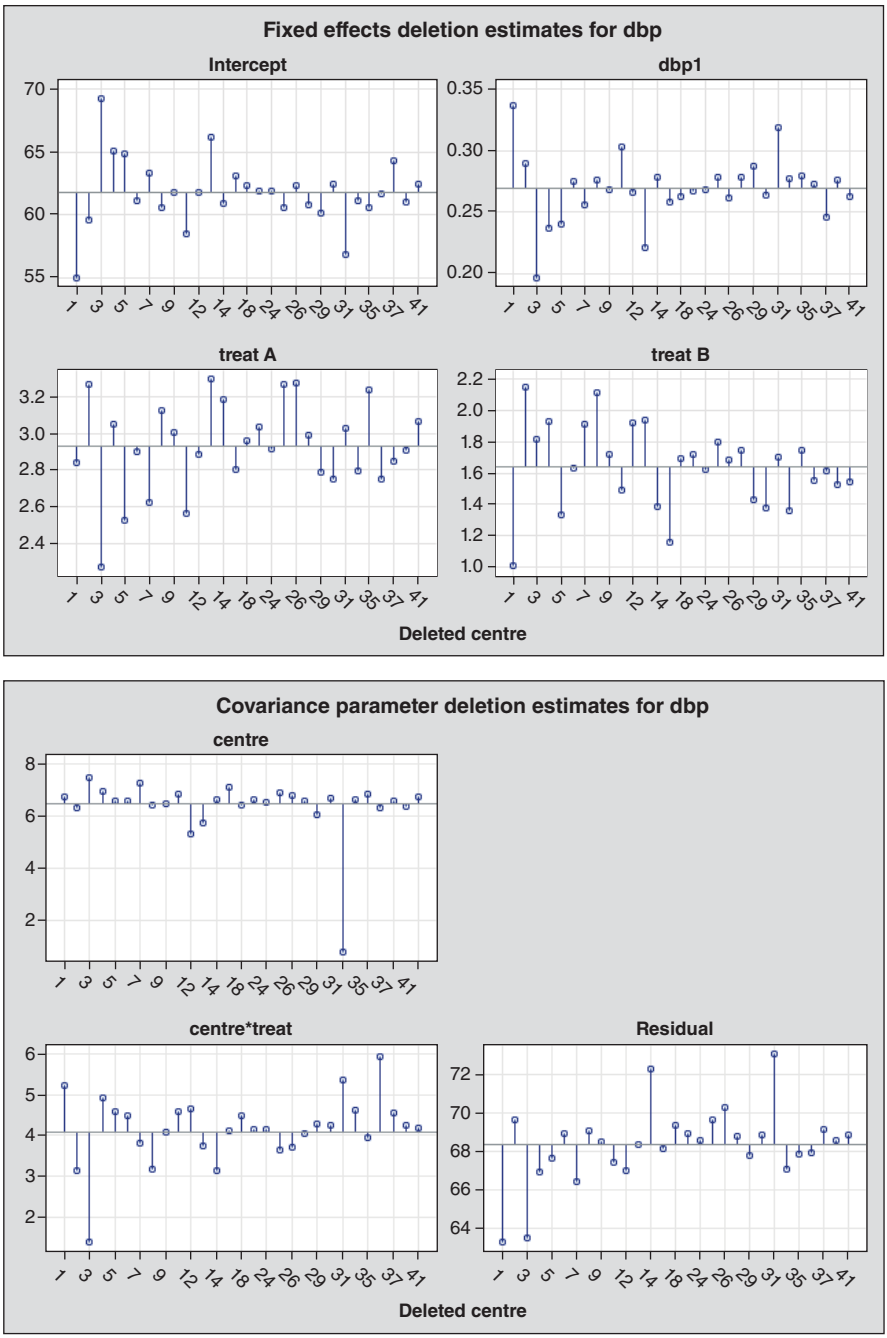


Figure 2.6 (b) Effect of removing centres on model parameters.

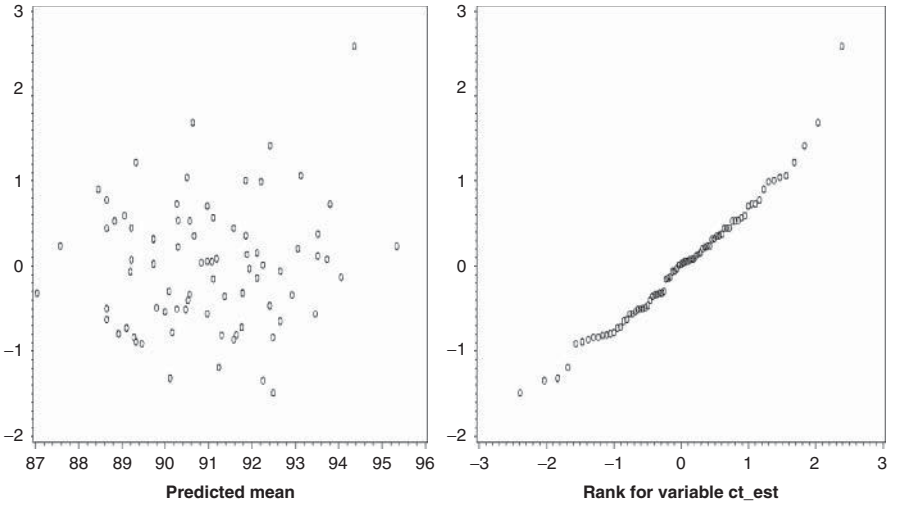


Figure 2.7 Plots of centre·treatment effects.

Centre·treatment effects The plots (Figure 2.7) of centre·treatment effects do not indicate any noticeably outlying values.

Homogeneity of treatment variances The standard deviations of the residuals and the centre·treatment effects were similar between the treatment groups, indicating no strong evidence of non-homogeneity of variance.

Treatment	Residual	Centre·treatment effects
A	9.15	0.85
B	6.53	0.70
C	7.86	0.67

Missing data (2.4.7)

Nearly all missing data in the study were caused by dropouts (usually following adverse effects of treatment). The number of patients dropping out following randomisation was A, 17; B, 12; and C, 3. The larger numbers for the first two treatments indicate that the dropout rate has been influenced by treatment. Missing values are therefore not missing completely at random. They may be considered as missing at random, however, and in Section 6.3, we present an analysis treating the observations from all visits as repeated measures data. In the present analysis, based on measurements at the final visit, we have used the 'last-value-carried-forward' method to minimise any bias.

Determining whether the simulated posterior distribution has converged (2.4.8)

Some of the approaches suggested in 2.4.8 will be used to assess the convergence of the parameters in Model 4. This model was initially fitted using 100,000 samples and a thinning factor of 5 (i.e. only one in five samples was used to construct the posterior).

Three types of diagnostic plots are readily available when using `PROC MCMC`. The first plot in each set is a 'trace plot', which plots simulated values of the parameter against the iteration number. Visual examination of this plot helps to determine if:

- The sampling process is reliable from the start or if it takes a number of iterations to settle down and sample in the right region. If it is not, then the samples taken before the process settles down should be discounted. These are often referred to as 'burn-in' samples.
- The samples are autocorrelated, that is, adjacent samples are correlated.
- Whether convergence is failing to occur, even after many samples.

The second 'autocorrelation' plot shows the correlation between samples depending on their separation. When there is autocorrelation, then more samples should be taken to ensure the full parameter range is adequately covered. The concurrent use of 'thinning' where only 1 in every 'n' samples is used to form the posterior has gained popularity, as it provides a more manageable number of samples to form the marginal parameter distributions. An n-fold increase in the sample size is often used with corresponding thinning level of n; so the eventual sample size is unchanged. However, we note that it is not essential to thin the sample. Greater increases in the sample size should be used when there is more autocorrelation.

The last plot is a histogram illustrating the marginal posterior distribution for the parameter.

The trace plots for the two treatment effect parameters (Figure 2.8) indicate the sampling appears to have settled very quickly for the fixed treatment effects, and it is not necessary to discount early samples. Had this not been the case, inclusion of an `NBI=m` option in the `PROC MCMC` statement will cause the first m samples to be omitted when obtaining the parameter estimates and summaries. The sampling has stayed in the same region (about -2 to 6) across all iterations, giving no indication that convergence may have failed. The trace plots are dense and do not indicate autocorrelation, and this is further illustrated in the autocorrelation plots. There is a slight kink in the posterior densities; however, the distributions are relatively smooth, and the tails appear adequately sampled.

Plots for the treatment difference A – B would need to be constructed and plotted manually from the sampled data. However, if plots for A – C and B – C are satisfactory, then it is very likely that plots of A – B will be too, since the A – B samples are based on the differences in A – C and B – C samples. So it is likely to be sufficient to examine convergence for A – C and B – C only.

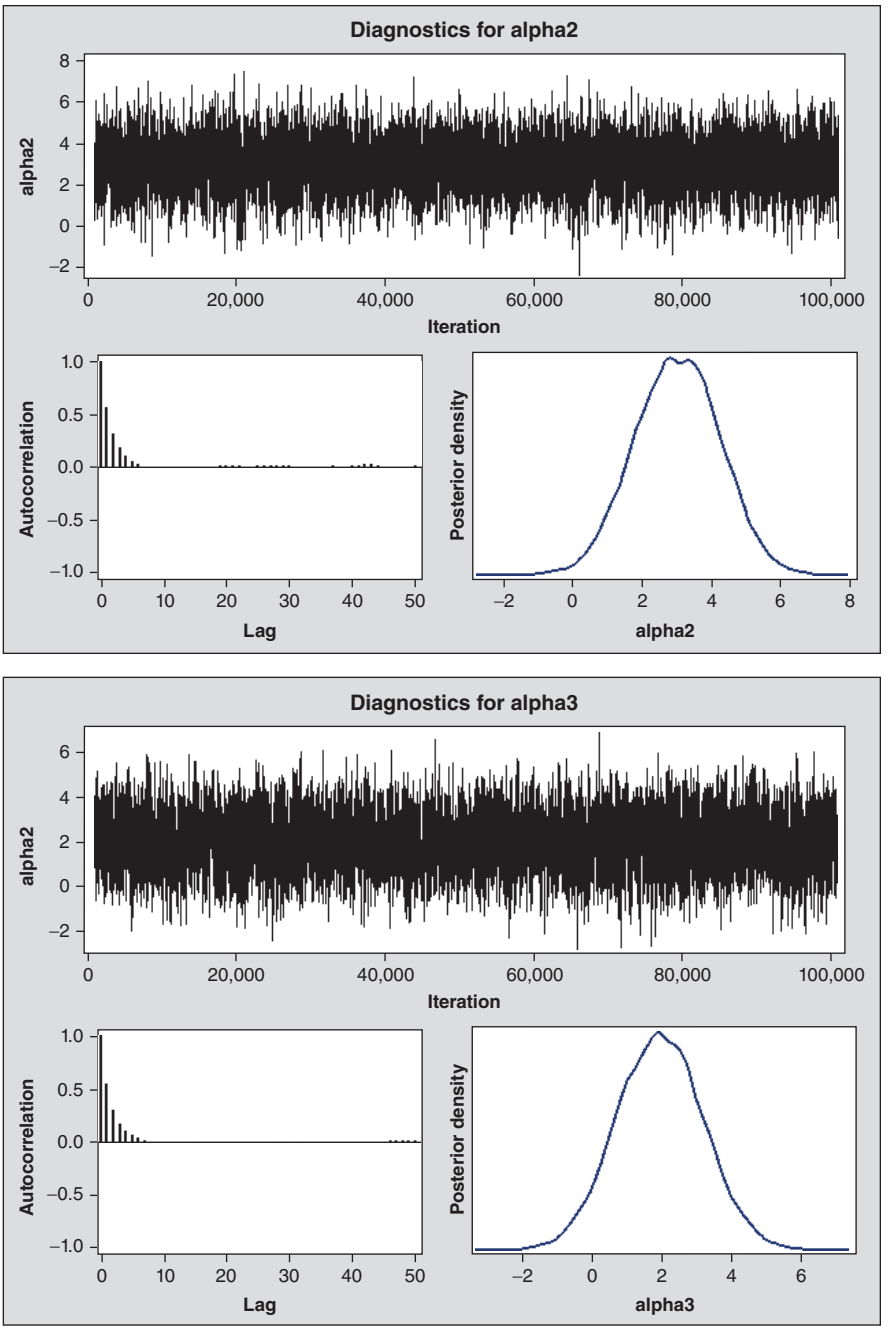


Figure 2.8 Diagnostic plots for two of the treatment effects, A − C (alpha2) and B − C (alpha3).

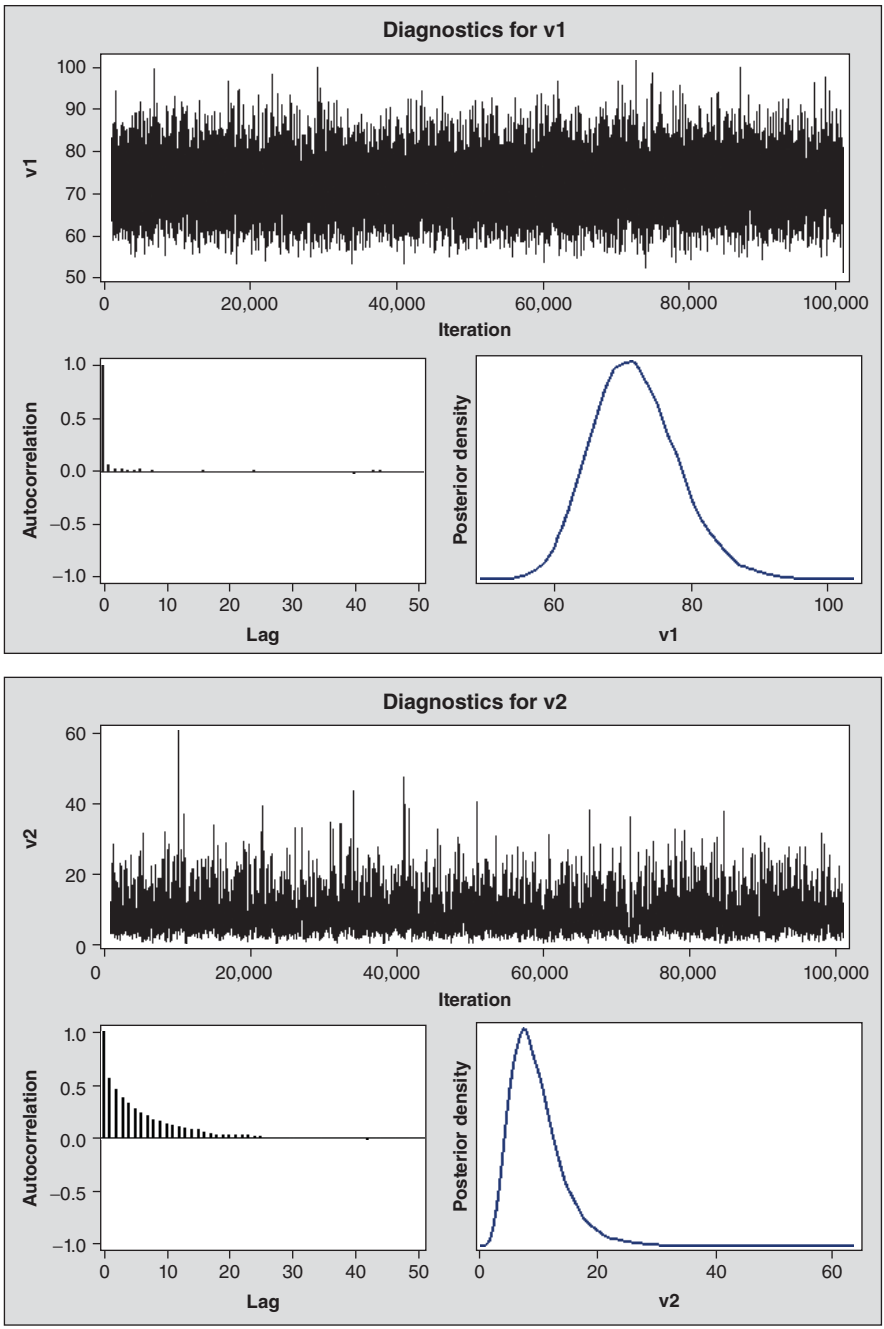


Figure 2.9 Diagnostic plots for the residual (v1) and centre variance component (v2) parameters.

All plots for the residual look acceptable (Figure 2.9). However, the centre variance component plots show a higher degree of autocorrelation, which only becomes negligible when samples are separated by about 20. We will try to alleviate problems that this may cause by using a sample of double the size and increasing the thinning factor so that only 1 observation in every 10 is used, to see if this affects the results.

Analysis with a thinning factor of 10

There are small changes to the treatment effect mean differences of the order of about 0.5–1% (Table 2.2). Surprisingly, there is little difference to the centre variance component estimates for which there was most evidence of autocorrelation. Since there have been changes to the treatment effects, it would seem preferable to use the results based on the larger sample size and higher thinning level of 10.

For those working in the pharmaceutical industry, it may be necessary to specify how decisions will be made about thinning and burn-in in advance in the study protocol.

Table 2.2 Comparing results from Model 4 with larger sample size and higher thinning factor.

Treatment effects (SE)			
Sample size	*Thinning factor*	*A – C*	*B – C*
100,000	5	3.00 (1.22)	1.97 (1.23)
200,000	10	2.98 (1.22)	1.96 (1.25)
Variance components (95% probability intervals)			
Sample size	*Thinning factor*	Centre	*Residual*
100,000	5	7.41 (1.11 – 16.95)	71.4 (60.1 – 84.6)
200,000	10	7.42 (0.90 – 17.22)	71.4 (60.6 – 85.5)

Further assessments of convergence using summary statistics and diagnostic tests

Some of the summary statistics and tests for assessing convergence that are readily available in `SAS` are now considered. The following table shows the 'Monte Carlo Standard Errors' (`MCSE`), the `SDs` of the parameters (from original output) and their ratio (`MCSE/SD`). The small ratios indicate that only a fraction of the posterior variability is due to the simulation, indicating that an adequate sample size is likely to have been achieved.

Monte Carlo Standard Errors

Parameter	MCSE	Standard Deviation	MCSE/SD
alpha0	0.1162	11.2354	0.0103
alpha1	0.00111	0.1083	0.0103
alpha2	0.0119	1.2173	0.00981
alpha3	0.0121	1.2453	0.00975
v1	0.0490	6.3472	0.00771
v2	0.0701	4.5493	0.0154

This table shows the autocorrelations among posterior samples. They are fairly small after 10 lags, and a thinning factor of 10 therefore seems appropriate.

Posterior Autocorrelations

Parameter	Lag 1	Lag 5	Lag 10	Lag 50
alpha0	0.3278	0.0131	-0.0159	0.0011
alpha1	0.3268	0.0104	-0.0165	0.0027
alpha2	0.3216	-0.0056	0.0138	0.0022
alpha3	0.3118	0.0074	0.0086	0.0078
v1	0.0297	0.0011	-0.0079	-0.0012
v2	0.4637	0.1437	0.0319	-0.0045

This table shows results of the Geweke test comparing mean estimates obtained from early and late samples. No parameter failed the test.

Geweke Diagnostics

Parameter	z	Pr > |z|
alpha0	-0.3730	0.7091
alpha1	0.2331	0.8157
alpha2	1.0436	0.2967
alpha3	0.8770	0.3805
v1	-1.1986	0.2307
v2	0.5445	0.5861

The following table reports the effective sample sizes for each parameter after taking into account the autocorrelation for each parameter, that is, the sample size expected if the samples were totally independent. The effective sample size is the lowest for centre variance component, which was found to have the highest degree of autocorrelation. However, the estimated 4200 independent samples are still likely to be sufficient. It is difficult to set general rules for how to use the statistics in this table. However, if the effective sample size was very small, it would

be sensible to increase the number of samples further. The 'time' column estimates amount of separation between samples required for them to become uncorrelated, and the 'efficiency' column is the ratio of the effective sample size to the total sample size after thinning (20,000 in this case).

Effective Sample Sizes

Parameter	ESS	Autocorrelation Time	Efficiency
alpha0	9349.7	2.1391	0.4675
alpha1	9481.4	2.1094	0.4741
alpha2	10387.4	1.9254	0.5194
alpha3	10527.1	1.8999	0.5264
v1	16806.7	1.1900	0.8403
v2	4212.5	4.7478	0.2106

SAS code and output

There now follows a description of the SAS code and the outputs, which were produced in the analysis of the previous example. This is a pattern that will be followed throughout the rest of the book. Some readers may be relative novices at using SAS and may find some of the aspects of the code or output confusing. Anyone in this position is referred forward to Sections 9.2 and 9.4 where the PROC MIXED and PROC MCMC procedures are introduced. The SAS code and datasets may be obtained electronically from web page www.wiley.com/go/brown/applied_mixed.

The variable names used for the program are as follows:

centre = centre number,
treat = treatment (A,B,C),
patient = patient number,
dbp = diastolic blood pressure at last attended visit,
dbp1 = baseline diastolic blood pressure.

SAS code only is given for Models 1 and 2 because their outputs have a similar form to that from Model 3. The full code and output are given for Model 3 to illustrate the use of PROC MIXED in fitting a mixed model using REML and also to show how model checking can be undertaken. The full code and output are given for Model 4 to demonstrate the use of SAS for performing a Bayesian analysis. The output for Model 5 is not given because it is very similar to that for Model 4.

Model 1

```
PROC MIXED CL; CLASS centre treat;
MODEL dbp = dbp1 treat centre;
LSMEANS treat/ DIFF PDIFF CL;
```

Model 2

```
PROC MIXED CL; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR;
RANDOM centre;
LSMEANS treat/ DIFF PDIFF CL;
```

Model 3

```
PROC MIXED CL; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR;
RANDOM centre centre*treat;
LSMEANS treat/ DIFF PDIFF CL;
```

```
                    Model Information
Data Set                        WORK.A
Dependent Variable              dbp
Covariance Structure            Variance Components
Estimation Method               REML
Residual Variance Method        Profile
Fixed Effects SE Method         Kenward-Roger
Degrees of Freedom Method       Kenward-Roger

                  Class Level Information
Class      Levels     Values
centre     29         1 2 3 4 5 6 7 8 9 11 12 13 14 15 18 23 24
                      25 26 27 29 30 31 32 35 36 37 40 41
treat      3          A B C

                       Dimensions
                Covariance Parameters      3
                Columns in X               5
                Columns in Z             108
                Subjects                   1
                Max Obs Per Subject      288

                   Number of Observations
              Number of Observations Read        288
              Number of Observations Used        288
              Number of Observations Not Used      0
```

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	2072.30225900	
1	3	2055.64188178	0.00000322
2	1	2055.63936685	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
centre	6.4628	0.05	2.4064	46.0513
centre*treat	4.0962	0.05	0.8793	1515.23
Residual	68.3677	0.05	57.1917	83.1916

Fit Statistics

-2 Res Log Likelihood	2055.6
AIC (smaller is better)	2061.6
AICC (smaller is better)	2061.7
BIC (smaller is better)	2065.7

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	284	6.16	0.0137
treat	2	25	2.16	0.1364

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
treat	A	92.3491	1.1233	52	82.21	<.0001	0.05	90.0951	94.6032
treat	B	91.0632	1.1695	51.4	77.86	<.0001	0.05	88.7157	93.4107
treat	C	89.4217	1.1326	58.7	78.96	<.0001	0.05	87.1552	91.6882

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
treat	A	B	1.2859	1.4300	23.8	0.90	0.3775	0.05	-1.6665	4.2384
treat	A	C	2.9274	1.4109	25.6	2.07	0.0482	0.05	0.02511	5.8297
treat	B	C	1.6415	1.4453	25.7	1.14	0.2666	0.05	-1.3314	4.6144

Model 3 with model checking

The following code can be used in any version of SAS to obtain the residual and normal plots. Following this, alternative code using ODS GRAPHICS code available in SAS Version 9 will be given. Note that only the plots of centre and centre·treatment effects resulting from this code are shown in the text (Figures 2.6(a) and 2.7). Figures 2.5 and 2.6(b) result from the ODS GRAPHICS code given in a later section.

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR OUTP=pred OUTPM=predm;
RANDOM centre centre*treat/ SOLUTION;
LSMEANS treat/ DIFF PDIFF CL;
ODS LISTING EXCLUDE SOLUTIONR;
ODS OUTPUT SOLUTIONR=solut; RUN;

DATA solut; SET solut;
centrex=centre*1; * obtain numeric centre variable;
DROP centre;

DATA c_est(KEEP=centre c_est) ct_est(KEEP=centre treat
  ct_est);
SET solut;
centre=centrex;
IF effect='centre' THEN DO;
 c_est=estimate;
 OUTPUT c_est;
END;
ELSE DO;
 ct_est=estimate;
 OUTPUT ct_est;
END;
PROC SORT DATA=ct_est; BY centre treat;

PROC SORT DATA=predm; BY centre treat;

DATA c_est; MERGE predm c_est; BY centre;
PROC MEANS NOPRINT; BY centre; ID c_est;
VAR pred; OUTPUT OUT=c_est MEAN=c pred N=freq;

DATA ct_est; MERGE predm ct_est; BY centre treat;
PROC MEANS NOPRINT; BY centre treat; ID ct_est;
VAR pred; OUTPUT OUT=ct_est MEAN=ct pred N=freq;

SYMBOL1 V=CIRCLE;

PROC GPLOT DATA=pred; PLOT resid*pred;
TITLE 'RESIDUALS AGAINST THEIR PREDICTED VALUES';
```

```
PROC RANK DATA=pred OUT=norm NORMAL=TUKEY; VAR resid;
   RANKS s_est;
PROC GPLOT DATA=norm; PLOT resid*s est;
TITLE 'RESIDUALS - NORMAL PLOT';

PROC GPLOT DATA=c_est; PLOT c_est*c pred;
TITLE 'CENTRE EFFECTS AGAINST THEIR PREDICTED VALUES';
PROC RANK OUT=norm NORMAL=TUKEY DATA=c_est; VAR c_est;
   RANKS s est;
PROC GPLOT DATA=norm; PLOT c_est*s est;
TITLE 'CENTRE EFFECTS - NORMAL PLOT';

PROC GPLOT DATA=ct_est; PLOT ct_est*ct pred;
TITLE 'CENTRE.TREAT EFFECTS AGAINST THEIR PREDICTED VALUES';
PROC RANK OUT=norm NORMAL=TUKEY DATA=ct_est; VAR ct_est;
   RANKS s est;
PROC GPLOT DATA=norm; PLOT ct_est*s est;
TITLE 'CENTRE.TREAT EFFECTS - NORMAL PLOT';

* CHECK HETEROGENEITY OF RESIDUAL AND CENTRE*TREAT VARIANCE
   BY TREATMENT;
PROC SORT DATA=pred; BY treat;
PROC MEANS DATA=pred; VAR resid; BY treat;
TITLE 'PROC MEANS TO CHECK RESIDUAL VARIANCE IS HOMOGE-
NEOUS ACROSS TREATMENTS';

PROC SORT DATA=ct_est; BY treat;
PROC MEANS DATA=ct_est; VAR ct_est; BY treat;
TITLE 'PROC MEANS TO CHECK CENTRE.TREAT EFFECT VARIANCE IS
HOMOGENEOUS ACROSS TREATMENTS';
```

This code may not at first sight be straightforward to understand. The steps used may be summarised as follows:

- Fit Model 3.
- Output the residuals and predicted values given by $\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}}$ to dataset `pred` using the `OUTP` option; output the predicted values given by $\mathbf{X}\hat{\boldsymbol{\alpha}}$ to dataset `predm` using the `OUTPM` option; output the random effects estimates to dataset `solut` using the `SOLUTION` option in the `RANDOM` statement and an `ODS OUTPUT` statement. Note that the `ODS LISTING EXCLUDE` statement causes the random effects to be excluded from the main `PROC MIXED` output.
- Create datasets `c_est` and `ct_est` containing random effects estimates for the centre and centre·treatment effects, respectively.
- Merge datasets `c_est` and `ct_est` with the predicted values and calculate the mean predicted value within each random effects category.
- Print and plot the residuals and the centre and centre·treatment effect_estimates.

- Calculate the standard deviations of the residuals and centre·treatment effects within each treatment group.

The output from PROC MIXED is identical to that given in Chapter 4, and the residual plots produced by this code were given earlier in this section.

SAS diagnostic graphics using the Results Viewer in Version 9.3

Using the default options in SAS Version 9.3, the following code can be used as a quick and simple way to obtain a range of diagnostic plots. Either influence plots, including those shown in Figure 2.5(b), or residual plots, including Figure 2.5(a), can be obtained depending on the options chosen in the MODEL statement.

```
PROC MIXED PLOTS=ALL CL; CLASS centre treat;
MODEL dbp=dbp1 treat / DDFM=KR INFLUENCE(ITER=5);
RANDOM centre centre*treat;
```

Replacing INFLUENCE(ITER=5) by RESIDUAL in the MODEL statement will produce a variety of residual plots, instead of influence plots. It is not possible to produce residual and normal plots for the centre or centre·treatment effects without using the extensive code described earlier. Influence plots can be produced for these effects, however, by replacing the MODEL statement with the following:

```
MODEL dbp=dbp1 treat / DDFM=KR INFLUENCE (EFFECT=centre ITER=5);
```

Part of the output this produces appears as Figure 2.6(b), and similar code can be used to generate influence plots for centre·treatment.

Model 4

```
PROC MCMC OUTPOST=post4 NMC=100000 THIN=5 SEED=7893;
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
   POSTINTERVALS;
PARMS alpha0 alpha1 alpha2 alpha3 v1 v2;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_centre ~ NORMAL(0, VAR=v2) SUBJECT=centre;
mu = alpha0 + alpha1*dbp1 + alpha2*treata + alpha3*treatb
   + b_centre;
MODEL dbp ~ NORMAL(mu, VAR = v1);
```

Parameters

Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	v1	Conjugate	0.00990	igamma(0.01, scale = 0.01)
2	v2	Conjugate	0.00990	igamma(0.01, scale = 0.01)
3	alpha0	N-Metropolis	0	normal(0, var = 10000)
	alpha1		0	normal(0, var = 10000)
	alpha2		0	normal(0, var = 10000)
	alpha3		0	normal(0, var = 10000)

This table is a reminder of the prior distributions specified in the code and gives the initial sample values. In this case, these were not specified within the `PARMS` statement, and it would appear modes of the prior distributions have been used.

Random Effects Parameters

Parameter	Subject	Levels	Prior Distribution
b_centre	centre	29	normal(0, var = v2)

This table reminds us that a normal distribution with zero mean and variance equal to the centre variance component (`v2`) has been assumed for centre effects.

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
alpha0	20000	59.7667	11.1272	52.2638	59.8739	67.2426
alpha1	20000	0.2869	0.1073	0.2148	0.2864	0.3602
alpha2	20000	3.0034	1.2197	2.1748	3.0109	3.8387
alpha3	20000	1.9731	1.2302	1.1229	1.9689	2.7984
v1	20000	71.7779	6.3109	67.3362	71.3917	75.8079
v2	20000	8.2553	4.5165	5.0898	7.4128	10.4457

This table summarises the location of the model parameters. `alpha0-alpha3` denote parameters for the fixed effects – intercept, baseline DBP, treatment A and treatment B. In this case, the mean estimates for the fixed effects parameters

are very close to the medians (50% percentile), indicating that their distribution is symmetrical. The standard deviations of the parameters are analogous to standard errors of the mean. `v1` and `v2` are the variance components for the residual and centre effects. Usually, the mean is not the best summary of location for variance components, as their distributions are usually skewed. The medians values (50% percentiles) are preferable.

Posterior Intervals

Parameter	Alpha	Equal-Tail Interval		HPD Interval	
alpha0	0.050	37.5928	81.5463	38.8495	82.6188
alpha1	0.050	0.0781	0.5019	0.0879	0.5079
alpha2	0.050	0.6256	5.3571	0.6571	5.3703
alpha3	0.050	-0.4242	4.3941	-0.4535	4.3582
v1	0.050	60.5610	85.2193	60.1364	84.6258
v2	0.050	2.0651	19.3398	1.1116	16.9583

This table provides probability intervals for the parameters, which are analogous to confidence intervals. The HPD interval is obtained such that within which all points have a higher probability density than all points outside it (see Section 2.3.3). The equal tail interval ensures that each tail is of equal size. Compared to the HPD interval, this interval will be shifted towards the smallest tail of the HPD. It is not clear which interval is preferable, although the equal-tail interval is easier to calculate and may be more acceptable to Regulatory bodies. There can be noticeable differences between the intervals; so it would seem important for those working in an experimental setting to specify which interval will be used in advance.

Only results for treatment differences A – C (alpha2) and B – C (alpha3) are given by `PROC MCMC`. There is no `ESTIMATE` statement available as in `PROC MIXED` to provide results for the treatment difference A – B. This may be obtained from the differences between the A – C and B – C samples.

Obtaining an estimate of A – B, probability intervals and *p*-values

To obtain an estimate of A – B, it is necessary to summarise the simulated values in the posterior distribution as the model is parameterised in terms of A – C and B – C, and `PROC MCMC` only summarises these samples. The following code obtains samples for the treatment difference A – B, based on the differences in the sampled values of A – C and B – C, which were output to dataset 'post4' from Model 4. An alternative approach would have been to relabel the treatments so that A – B was explicitly modelled and reanalysed the data.

p-values are not directly available in the procedure output, and it is necessary to obtain them from probability intervals that have zero on the boundaries.

p-values of the three treatment comparisons are then obtained by calculating the probabilities that the parameters take a value below or above zero. To be comparable to a classical 'two-sided' test, the *p*-value is calculated as twice the proportion of samples falling in the smallest tail with zero on the boundary (equivalent to the value of α for an equal-tail interval with zero on the boundary).

```
DATA p1; SET post4;
dbp1=alpha1;
a_c=alpha2;
b_c=alpha3;
a_b=alpha2-alpha3; * obtain samples of A-B;
IF dp1<0 THEN DBP0=1; ELSE dbp0=0;
IF a_b<0 THEN a_b0=1; ELSE a_b0=0;
IF a_c<0 THEN a_c0=1; ELSE a_c0=0;
IF b_c<0 THEN b_c0=1; ELSE b_c0=0;

PROC MEANS NOPRINT DATA=p1; VAR a_b dbp0 a_b0 a_c0 b_c0;
OUTPUT OUT=p2 SUM=dum dbp0_n a_b0_n a_c0_n b_c0_n
   N=samples mean=a_b_mean std=a_b_std;

DATA p3; SET p2;
* macro to calculate p-values for fixed effects parameters;
%MACRO p_calc(var);
&var.0_p=&var.0_n/samples;
IF &var.0<0.5 THEN &var._p=&var.0_p*2;
   ELSE &var._p=(1-&var.0_p)*2;
%MEND;

%p_calc(dbp); %p_calc(a_b); %p_calc(a_c); %p_calc(b_c);

PROC UNIVARIATE DATA=p1;
VAR a_b a_c b_c; OUTPUT OUT=ci PCTLPTS=2.5 97.5
   PCTLPRE=a_b a_c b_c PCTLNAME=lower upper;

PROC PRINT NOOBS DATA=p3; VAR a_b_mean a_b_std;
TITLE 'Mean and SE for A-B';

PROC PRINT NOOBS DATA=p3; VAR a_b_p a_c_p b_c_p;
TITLE 'p-values for pairwise treatment comparisons';

PROC PRINT NOOBS DATA=ci;
TITLE 'Equal tail 95% CIs';
```

Mean and SE for A - B

a_b_mean	a_b_std
1.03027	1.22645

```
p-values for pairwise treatment comparisons

a_b_p     a_c_p     b_c_p

0.3928    0.0139    0.109

Equal tail 95% CIs

a_blower    a_bupper    a_clower    a_cupper    b_clower    b_cupper

-1.36294     3.45211     0.62563     5.35712    -0.42416     4.39410
```

This output shows the mean for A – B and its standard deviation and *p*-values corresponding to each of the treatment differences. Only the difference between treatments A and C is significant at the 5% level. Note the equal tail confidence intervals for A – C and B – C are the same as those already given in the default SAS output. We now also have the confidence interval for A – B.

Determining whether the simulated posterior distribution has converged

Adding an ODS graphics statement to SAS MCMC code will cause three diagnostic plots to appear for each model parameter that can be used to help check convergence: a trace plot, a plot of the correlation between samples by their distance apart in the chain, and a plot of the marginal density. The code is given for Model 4.

```
ODS GRAPHICS ON;
PROC MCMC OUTPOST=post4 NMC=100000 THIN=5 SEED=7893;
PARMS alpha0 0 alpha1 0 alpha2 0 alpha3 0 v1 1 v2 1;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_centre ~ NORMAL(0, VAR=v2) SUBJECT=centre;
mu = alpha0 + alpha1*dbp1 + alpha2*treata + alpha3*treatb
   + b_centre;
MODEL dbp ~ NORMAL(mu, VAR = v1);
ODS GRAPHICS OFF;
```

An `ODS SELECT` statement has not been used in this code, and parameter summaries are produced. Surprisingly, once an `ODS SELECT` statement is included (as in the original `SAS` code), the default model diagnostics will not appear unless specifically requested by the `TADPANEL` option within the `ODS` statement. Thus, if both parameter summaries and model diagnostics are

required, then the TADPANEL option should be added to the ODS statement to ensure the three diagnostic plots appear.

```
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
   POSTINTERVALS TADPANEL;
```

Output

The plots created for the fixed effects and variance components are shown earlier in the main example text, along with the summary statistics and the Gewerke test.

Changing the sample size thinning factor

This may be easily achieved by altering the value used for NMC option to change the sample size and the THIN option to change the thinning factor. The following code requests a sample size of 200,000 and a thinning factor of 10, causing only one in 10 of the values sampled to be used to form the posterior distribution. Hence, the posterior is formed from 20,000 samples.

```
PROC MCMC OUTPOST=post4 NMC=200000 THIN=10 SEED=7893;
```

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
alpha0	20000	59.7624	11.2354	52.1993	59.8480	67.3192
alpha1	20000	0.2870	0.1083	0.2136	0.2864	0.3609
alpha2	20000	2.9799	1.2173	2.1669	2.9779	3.7976
alpha3	20000	1.9620	1.2453	1.1073	1.9551	2.8069
v1	20000	71.8359	6.3472	67.3414	71.4243	75.8141
v2	20000	8.2548	4.5493	5.0690	7.4218	10.4203

Posterior Intervals

Parameter	Alpha	Equal-Tail Interval		HPD Interval	
alpha0	0.050	37.6091	81.9172	36.5540	80.5030
alpha1	0.050	0.0765	0.5006	0.0874	0.5095
alpha2	0.050	0.5938	5.3785	0.5539	5.3198
alpha3	0.050	-0.4722	4.3940	-0.4854	4.3726
v1	0.050	60.6394	85.4908	59.8878	84.5619
v2	0.050	2.0057	19.4847	0.8987	17.2243

Model 5

```
PROC MCMC OUTPOST=post5 NMC=100000 THIN=5 SEED=7893;
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
   POSTINTERVALS;
PARMS alpha0 alpha1 alpha2 alpha3 v1 v2 v3;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_centre ~ NORMAL(0, VAR=v2) SUBJECT=centre;
RANDOM b_ct ~ NORMAL(0, VAR=v3) SUBJECT=centre_treat;
mu = alpha0 + alpha1*dbp1 + alpha2*treata + alpha3*treatb
   + b_centre + b_ct;
MODEL dbp ~ NORMAL(mu, VAR = v1);
```

3

Generalised linear mixed models

Up until now, we have considered models with normally distributed errors. However, there are many situations where data are not of this type, for example, where the presence/absence of an adverse event is recorded and the normality assumption cannot be made. A class of models known as generalised linear models (GLMs) is available for fitting fixed effects models to such non-normal data. These models can be further extended to fit mixed models and are then referred to as generalised linear mixed models (GLMMs). Random effects, random coefficients or covariance patterns can be included in a GLMM in much the same way as in normal mixed models, and again either balanced or unbalanced data can be analysed. Although GLMMs can be used to analyse data from any distribution from the exponential family, binary data and Poisson data are most frequently encountered, and for this reason, this book will primarily concentrate on the use of GLMMs for these data types.

In introducing these topics, we will, of necessity, be less than comprehensive. In their excellent book *Generalised Linear Models*, McCullagh and Nelder (1989) take 500 pages to cover the subject and start by an assumption of 'a knowledge of matrix theory, including generalised inverses, together with basic ideas of probability theory, including orders of magnitude in probability'. At one end of the readership spectrum, therefore, those with no experience of GLMs may wish to skip all but the introductory paragraphs of each section because some sections inevitably draw on the assumption of prior knowledge. This will enable such readers to identify where such methods might prove useful. In case the reader with little background knowledge of GLMs identifies a need to apply GLMs and GLMMs and is horrified at the prospect of having to master textbooks on the subject, we would emphasise that fitting such models can often be achieved without an encyclopaedic knowledge of the topic. The final section of this chapter and sections of subsequent chapters will illustrate the application of these models and present the SAS code needed to implement them.

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

We will start by describing the GLM in Section 3.1 and then show how it is extended to the GLMM in Section 3.2. GLMMs are more complex than normal mixed models, and there is therefore more potential for problems such as biased estimates and a failure to converge. These are considered in Section 3.3, which also gives some practical information on fitting GLMMs. A worked example is given in Section 3.4.

3.1 Generalised linear models

3.1.1 Introduction

GLMs can be used to fit fixed effects models to certain types of non-normal data: those with a distribution from the exponential family. Consider the following example. We wish to conduct a clinical trial to investigate the effect of a new treatment for epilepsy. A suitable variable for assessing efficacy is the number of seizures that occur during a predetermined period. Thus, the response variable is a count. Such variables are often found to follow a Poisson distribution. This is a member of the exponential family, and GLMs or GLMMs can be considered, depending on the details of how the trial is designed. Such an example is considered in Section 6.4. As a second example, consider the analysis of a particular adverse event in a clinical trial. In some situations, a simple contingency-table-based analysis will be sufficient. If, however, there are baseline effects or if the trial design is more complicated, a GLM may be preferred. In the multi-centre trial, which we are regularly revisiting in this book, the occurrence of cold feet was such an adverse event and could be reported at any of the follow-up visits. As a binary outcome, this is also from the exponential family, and in Section 3.4, we will show how GLMMs can be applied to these data.

As with the models we have met for normally distributed data, the models use a linear combination of variables to 'predict' the response. In the case of normally distributed data, the fixed effects model is $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e}$. That is, the response is determined by the linear component, $\mathbf{X}\boldsymbol{\alpha}$, which gives the expected response, which we will denote by $\boldsymbol{\mu}$ and by a randomly determined error term. In a somewhat convoluted way, we could write the model as

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e},$$

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\alpha}.$$

The GLM can easily be specified from this artificial-looking model by allowing $\boldsymbol{\mu}$ and $\mathbf{X}\boldsymbol{\alpha}$ to be related by a 'link function', g, so that

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha}.$$

Thus, normal models are a special case of GLMs in which the link function is the identity function. In general, the link function is not the identity function but takes a form suitable for the distribution of the data.

An alternative, less mathematical way of familiarisation with the concept of the GLM is to think of the link function as a method of mapping the response data from their scale of observation to the real scale $(-\infty, +\infty)$. For example, binomial probabilities have a range 0–1, and the logit link function, $\log(\mu/(1-\mu))$, will translate this range to the real scale. This is necessary because fitting a linear model directly to the binomial parameter could lead to estimates of probabilities that were negative or greater than one. Use of the link function allows the model parameters to be included in the model linearly, just as in the models we have described for normal data. This often gives the GLM an advantage over contingency table methods, which are sometimes used to analyse binary data (e.g. chi-squared tests) because these methods cannot incorporate several fixed effects simultaneously.

In this chapter, we will give only a brief introduction to GLMs. However, more detail can be found in McCullagh and Nelder (1989). Before defining the GLM, basic details of the binomial and Poisson distributions will be given for those who are not completely familiar with these distributions, and the general form for distributions from the exponential family will be specified. This general distributional form will be needed for setting a particular form of link function known as the 'canonical' link.

3.1.2 Distributions

We now define the Bernoulli, binomial and Poisson distributions. These can all be described as 'one-parameter' distributions, that is using a single parameter completely describes the distribution.

The Bernoulli distribution

This distribution is used to model binary data where observations have one of two possible outcomes, which can be thought of as 'success' or 'failure'. If one is used to denote success and zero failure, the density function is

$$f(y) = \mu^y(1-\mu)^{(1-y)}, \quad y = 0, 1.$$

Thus, μ corresponds to the probability of success, and the mean and variance are given by

$$\text{mean}(y) = \mu,$$

$$\text{var}(y) = \mu(1-\mu).$$

The binomial distribution

This distribution is also suitable for binary data. However, observations are now recorded as the number of successes out of a number of 'tries'. The parameter of interest is the proportion of successes. If z and n are the observed numbers of successes and tries, respectively, then the proportion $y = z/n$ has a density function

$$f(y, n) = \frac{n!(\mu)^{ny}(1-\mu)^{n-ny}}{(ny)!(n-ny)!}$$

and

$$\text{mean}(y) = \mu,$$
$$\text{var}(y) = \mu(1-\mu)/n.$$

Note that when $n = 1$, the Bernoulli density function is obtained. Thus, the Bernoulli distribution is a special case of the binomial distribution.

The Poisson distribution

This distribution can be used to model 'count' data. The number of episodes of dizziness over a fixed period and the number of abnormal heart beats on an ECG tape of a prescribed length are examples of count data. Its density function is given as

$$f(y) = \mu^y \mathrm{e}^{-\mu}/y!, \quad y = 0, 1, 2, \ldots$$

and

$$\text{mean}(y) = \mu,$$
$$\text{var}(y) = \mu.$$

The Poisson distribution with offset

Sometimes, the underlying scale for count data varies with each observation. For example, observations may be made over varying periods (e.g. number of epileptic seizures measured over different numbers of days for each patient). Alternatively, the underlying scale may relate to some other factor such as the size of a geographical region over which counts of subjects with a specific disease are taken. To take account of such a varying scale, the scale for each observation needs to be utilised in forming the distribution density. The scale variable is often referred to as the *offset*. The parameter of interest is then the number of counts per unit scale of the offset variable. If we denote the offset variable by t (even though it is not always time) and the observed number of counts by z, the distribution of $y = z/t$ has a density function

$$f(y, t) = (\mu t)^{yt}\mathrm{e}^{-\mu t}/(yt)!$$

and

$$\text{mean}(y) = \mu,$$

$$\text{var}(y) = \mu/t.$$

Note that when $t = 1$, the density function for the Poisson distribution without an offset variable as shown previously is obtained, confirming it as a special case of this distribution.

3.1.3 The general form for exponential distributions

To show how the GLM can be used for data with any exponential family distribution, we first need to define a general form in which all exponential family density functions can be expressed. This can be written as

$$f(y; \theta, \phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y, \phi)\},$$

where

θ = a location parameter (not necessarily the mean),
ϕ = a dispersion parameter (only appears in distributions that have two parameters such as the normal distribution).

The form of the functions a, b and c will be different for each distribution. Distributions that are not from the exponential family cannot be expressed in this way.

The one-parameter distributions considered in this book can be defined solely in terms of the location parameter, θ, and the general form then simplifies to

$$f(y; \theta) = \exp\{[y\theta - b(\theta)]/a + c(y)\},$$

where a is now a constant. Expressions for a, $b(\theta)$ and $c(y)$ are listed in the following table for the Bernoulli, binomial and Poisson distributions:

Distribution	a	$b(\theta)$	$c(y)$
Bernoulli	1	$\log(1 + \exp(\theta))$	1
Binomial	$1/n$	$\log(1 + \exp(\theta))$	$\log[n!/((ny)!\,(n - ny)!)]$
Poisson	1	$\exp(\theta)$	$-\log(y!)$
Poisson with offset	$1/t$	$\exp(\theta)$	$yt\log(yt) - \log(yt!)$

where n and t are the denominator and offset terms, respectively. Details of exactly how these expressions are obtained from their density functions will be given later in Section 3.1.6.

It can be shown that the mean and variance of a distribution can then be written in terms of the functions a and b as

$$\text{mean}(y) = \mu = b'(\theta),$$
$$\text{var}(y) = ab''(\theta).$$

Hence, $\theta = b'^{-1}(\mu)$, and we can alternatively write the variance in terms of μ as

$$\text{var}(y) = ab''(b'^{-1}(\mu)).$$

Using these expressions, the means and variances for the distributions can then be written in terms of θ or μ as follows:

Distribution	**Mean $\mu = b'(\theta)$**	**Variance in terms of θ, $ab''(\theta)$**	**Variance in terms of μ**
Bernoulli	$(1+\exp(-\theta))^{-1}$	$\exp(\theta)/(1+\exp(\theta))^2$	$\mu(1-\mu)$
Binomial	$(1+\exp(-\theta))^{-1}$	$\exp(\theta)/(1+\exp(\theta))^2/n$	$\mu(1-\mu)/n$
Poisson	$\exp(\theta)$	$\exp(\theta)$	μ
Poisson with offset	$\exp(\theta)$	$\exp(\theta)/t$	μ/t

3.1.4 The GLM definition

In Section 3.1.1, we defined the GLM using the matrix notation used for normal models by

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}$$

and related $\boldsymbol{\mu}$ to a linear sum of the fixed effects, $\mathbf{X}\boldsymbol{\alpha}$ (the linear component), by a 'link function', g, so that

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha}.$$

To relate the GLM directly to the general exponential density function introduced in the previous section, we label the linear component $\boldsymbol{\theta}$ so that $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\alpha}$. We will next consider how link functions can be constructed.

Canonical link functions

A type of link function known as the *canonical link function* is given by

$$g = b'^{-1},$$

where b is obtained from the general form for the density function for exponential distributions given previously. For the distributions we have considered, the canonical link functions are given by

Distribution	$g(\mu) = b'^{-1}(\mu)$	**Name**
Bernoulli	$\log(\mu/(1-\mu))$	Logit
Binomial	$\log(\mu/(1-\mu))$	Logit
Poisson	$\log(\mu)$	Log
Poisson with offset	$\log(\mu)$	Log

In most situations, use of the canonical link function will lead to a satisfactory analysis model. However, we should also mention that there is not a strict requirement for canonical link functions to be used in the GLM, and non-canonical link functions are also available. These are not derived from the density function but still map the data from their original scale onto the real scale. For example, a link function known as the probit function, $g(\mu) = \Phi^{-1}(\mu)$ (where Φ is the cumulative normal density function), is sometimes used for binary data recorded in toxicology experiments, since values of μ corresponding to specific probabilities can easily be obtained using the normal density function. Despite not being canonical, this link function does still map the original range of the data (0 to 1) to $-\infty$ to ∞ as required for the GLM. In this book, we will not be considering non-canonical link functions. However, further information can be found in McCullagh and Nelder (1989).

We earlier specified a general formula for the variance in the GLM as $\text{var}(y) = ab''(\boldsymbol{\theta})$. Using the relationship $g = b'^{-1}$, for canonical link functions, we can now equivalently write the variance in terms of μ and the function g as

$$\text{var}(y) = ag'^{-1}(\mu).$$

The variance matrix, V

The variance matrix for the GLM may be written as

$$\text{var}(\mathbf{y}) = \text{var}(\mathbf{e}) = \mathbf{V}.$$

Since the GLM is a fixed effects model, the observations are assumed to be uncorrelated, and the variance matrix, $\mathbf{V}$, is therefore diagonal. The diagonal terms of this matrix are equal to the variances of each observation given the underlying distribution. So, for example, in an analysis of six Bernoulli observations, where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_6)'$, $\mathbf{V}$ can be written as

$$\mathbf{V} = \begin{pmatrix} \mu_1(1-\mu_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu_2(1-\mu_2) & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu_3(1-\mu_3) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_4(1-\mu_4) & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_5(1-\mu_5) & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_6(1-\mu_6) \end{pmatrix}.$$

Note that unlike the fixed effects model for normal data, the variances are different for each observation.

In the previous subsection, we showed that the variance could be written in the general form

$$\text{var}(\mathbf{y}) = ab''(\boldsymbol{\theta}) = ag'^{-1}(\boldsymbol{\mu}).$$

We can use a and $b''(\boldsymbol{\theta})$ to express $\mathbf{V}$ in matrix form as a product of two diagonal matrices:

$$\mathbf{V} = \mathbf{AB},$$

where

$\mathbf{A} = \text{diag}\{a_i\}$,
$\mathbf{B} = \text{diag}\{b''(\theta_i)\} = \text{diag}\{g'^{-1}(\mu_i)\}$.

For the binomial distribution, $\mathbf{A}$ is a diagonal matrix of inverses of the denominator terms (number of 'tries'). For example, for a dataset with six observations, $\mathbf{A}$ would be

$$\mathbf{A} = \begin{pmatrix} 1/n_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/n_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/n_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/n_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/n_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/n_6 \end{pmatrix}.$$

For a Poisson distribution with offset, the n_i are replaced by the offset variable values, t_i.

$\mathbf{B}$ is a diagonal matrix of variance terms. For the Bernoulli and binomial distributions, it is

$$\mathbf{B} = \begin{pmatrix} \mu_1(1-\mu_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu_2(1-\mu_2) & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu_3(1-\mu_3) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_4(1-\mu_4) & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_5(1-\mu_5) & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_6(1-\mu_6) \end{pmatrix}.$$

Note that for the Bernoulli and Poisson distributions, $a_i = 1$ and thus $\mathbf{A} = \mathbf{I}$ (the identity matrix) and $\mathbf{V} = \mathbf{B}$.

The dispersion parameter

Variance in the model can be increased (or decreased) from the observation variances specified by the underlying distribution (i.e. $a_i b''(\theta_i)$), by multiplying the variance matrix by a dispersion parameter, ϕ:

$$\mathbf{V} = \phi\mathbf{AB}.$$

An alternative dispersion parameter is suggested by Williams (1982) for binary data. In this case, ϕ_i is calculated using a formula that varies depending on the denominator of each observation and so adjusts for their differing variances:

$$\mathbf{V} = \text{diag}\{\phi_i\}\mathbf{AB}.$$

It is referred to as the 'Williams modification'. Before GLMMs were developed, dispersion parameters were frequently used as a limited facility to model variance at the residual level in one-parameter distributions. We will consider the implications of using a dispersion parameter further in Section 3.3.7.

3.1.5 Fitting the GLM

For readers who wish to understand the 'mechanics' of how the GLM is fitted, we now look at the numerical procedures involved.

The GLM is fitted by maximising the log likelihood function. Using the general form for a one-parameter exponential distribution given in Section 3.1.3:

$$f(\mathbf{y};\boldsymbol{\theta}) = \exp\{[\mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta})]/a + c(\mathbf{y})\}.$$

The log likelihood for a set of observations can be written as

$$\log(L) = \sum_i (y_i\theta_i - b(\theta_i))/a_i + K,$$

or, in matrix/vector notation, as

$$\log(L) = \mathbf{y}'\mathbf{A}^{-1}\boldsymbol{\theta} - b(\boldsymbol{\theta})^{1/2'}\mathbf{A}^{-1}b(\boldsymbol{\theta})^{1/2} + K, \qquad \text{(A)}$$

where

$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)'$,
$b(\boldsymbol{\theta}) = (b(\theta_1), b(\theta_2), \ldots, b(\theta_n))'$,
$K =$ constant.

In fixed effects models for normal data, we saw that a solution for $\boldsymbol{\alpha}$ was easily obtained by differentiating the log likelihood with respect to $\boldsymbol{\alpha}$ and setting the resulting expression to zero (Section 2.2.2). However, in GLMs, the differentiated log likelihood, $\text{d}\log(L)/\text{d}\boldsymbol{\alpha}$, is non-linear in $\boldsymbol{\alpha}$, and an expression giving a direct solution for $\boldsymbol{\alpha}$ cannot be formed. To obtain $\text{d}\log(L)/\text{d}\boldsymbol{\alpha}$, we differentiate log(L) (A) after substituting $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\alpha}$:

$$\begin{aligned}
\text{d}\log(L)/\text{d}\boldsymbol{\alpha} &= \text{d}(\mathbf{y}'\mathbf{A}^{-1}\mathbf{X}\boldsymbol{\alpha})/\text{d}\boldsymbol{\alpha} - \text{d}(b(\mathbf{X}\boldsymbol{\alpha})^{1/2'}\mathbf{A}^{-1}b(\mathbf{X}\boldsymbol{\alpha})^{1/2})/\text{d}\boldsymbol{\alpha} \\
&= \mathbf{X}'\mathbf{A}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{A}^{-1}b'(\mathbf{X}\boldsymbol{\alpha}) \\
&= \mathbf{X}'\mathbf{A}^{-1}(\mathbf{y} - b'(\mathbf{X}\boldsymbol{\alpha})).
\end{aligned}$$

Setting this differential to zero leads to equations that can, in principle, be solved for $\boldsymbol{\alpha}$. However, because they are non-linear in $\boldsymbol{\alpha}$, one of the approaches below is usually used instead to obtain estimates of $\boldsymbol{\alpha}$.

The likelihood function can be maximised directly for $\boldsymbol{\alpha}$ using an iterative procedure such as Newton–Raphson (see Section 2.2.4). The variance of the resulting $\widehat{\boldsymbol{\alpha}}$ can be calculated at the final iteration by (from McCullagh and Nelder, 1989, Chapter 9)

$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{B}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{B})^{-1},$$

where

$$\mathbf{B} = \text{diag}\{g'^{-1}(\boldsymbol{\mu})\} = \text{diag}\{b''(\boldsymbol{\theta})\}.$$

We see that $\mathbf{B}$ is a diagonal matrix containing the variances of the individual observations.

Alternatively, the likelihood can be maximised using an *iterative weighted least squares* method. This approach (defined by McCullagh and Nelder, 1989, Section 2.5) can be based on analysing the following pseudo variable, $\mathbf{z}$, which can be thought of as a linearised observation vector:

$$\begin{aligned}\mathbf{z} &= g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})g'(\boldsymbol{\mu}) \\ &= g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1}.\end{aligned}$$

$\mathbf{z}$ can, in fact, be defined as a first-order Taylor series expansion for $g(\mathbf{y})$ about $\boldsymbol{\mu}$.

Recalling that $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha}$, we see that $\mathbf{z}$ has variance

$$\begin{aligned}\mathbf{V}_{\mathbf{z}} &= \text{var}(\mathbf{X}\boldsymbol{\alpha}) + \mathbf{B}^{-1}\text{var}(\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1} \\ &= \mathbf{0} + \mathbf{B}^{-1}\mathbf{A}\mathbf{B}\mathbf{B}^{-1} \\ &= \mathbf{A}\mathbf{B}^{-1}.\end{aligned}$$

A normal model can then be expressed in terms of the linearised pseudo variable, $\mathbf{z}$:

$$\mathbf{z} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e}.$$

$\mathbf{z}$ is then analysed iteratively using weighted least squares (see Section 2.2.1). Weights are taken as the inverse of the variance matrix $\mathbf{V}_{\mathbf{z}}$. Iteration is required because $\mathbf{z}$ and $\mathbf{V}_{\mathbf{z}}$ are dependent on $\boldsymbol{\alpha}$. The raw data, $\mathbf{y}$, can be taken as initial values for $\mathbf{z}$. Alternatively, $g(\mathbf{y})$ can be used for the initial values, although an adjustment may be necessary to prevent infinite values. The identity matrix can be used initially for $\mathbf{V}_{\mathbf{z}}$. The initial fixed effects solution is calculated using these values of $\mathbf{z}$ and $\mathbf{V}_{\mathbf{z}}$ as

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{z},$$

as in Section 2.2.2. This forms the first iteration. $\widehat{\boldsymbol{\alpha}}$ is then used to calculate new values for $\mathbf{z}$ and $\mathbf{V}_{\mathbf{z}}$. From these new values, $\widehat{\boldsymbol{\alpha}}$ is recalculated to form the second

iteration. The process is continued until $\widehat{\boldsymbol{\alpha}}$ converges. The asymptotic variance of $\widehat{\boldsymbol{\alpha}}$ can be calculated at the final iteration by

$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}.$$

The GLM can be fitted using `PROC GENMOD` or using `PROC GLIMMIX` without the inclusion of random effects terms (see Section 9.3).

3.1.6 Expressing individual distributions in the general exponential form

In Section 3.1.3, we introduced the idea of expressing distributions in a general form for exponential distributions:

$$f(y;\theta,\phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y,\phi)\}.$$

Forms for a, b and c were given for the Bernoulli, binomial and Poisson distributions. We now show how these forms are obtained from the distribution densities.

The Bernoulli distribution

The density function

$$f(y) = \mu^y(1-\mu)^{(1-y)}$$

is, by logging and then exponentiating the right-hand side of this equation, rearranged in the general exponential form as

$$f(y) = \exp[y\log(\mu/(1-\mu)) + \log(1-\mu)].$$

Thus, $\theta = \log(\mu/(1-\mu))$, and we obtain the logit as the canonical link function. The mean, μ, can then be expressed as the inverse of the logit function, $\mu = \exp(\theta)/(1+\exp(\theta)) = (1+\exp(-\theta))^{-1}$. Writing the distribution in terms of θ, we obtain

$$\begin{aligned} f(y) &= \exp\{y\theta + \log[1 - \exp(\theta)/(1+\exp(\theta))]\} \\ &= \exp\{y\theta - \log[1+\exp(\theta)]\}. \end{aligned}$$

Therefore, $b(\theta) = \log(1+\exp(\theta))$, $a = 1$ and $c(y) = 1$.

The binomial distribution

The density function

$$f(y,n) = \frac{n!(\mu)^{ny}(1-\mu)^{n-ny}}{(ny)!(n-ny)!}$$

is rearranged in the general exponential form using the same trick as was used for the Bernoulli distribution, as

$$f(y,n) = \exp\{[y\log(\mu/(1-\mu)) + n\log(1-\mu)]n + \log[n!/((ny)!(n-ny)!)]\}.$$

This again gives $\theta = \log(\mu/(1-\mu))$ and the logit as the canonical link function. Therefore, μ can again be expressed as $\exp(\theta)/(1+\exp(\theta)) = (1+\exp(-\theta))^{-1}$, and we can write

$$\begin{aligned} f(y,n) &= \exp\{[y\theta + \log\{1-\exp(\theta)/(1+\exp(\theta))\}]n + \log[n!/((ny)!(n-ny)!)]\} \\ &= \exp\{[y\theta - \log(1+\exp(\theta))]n + \log[n!/((ny)!(n-ny)!)]\}. \end{aligned}$$

Therefore, $b(\theta) = \log(1+\exp(\theta))$, $a = 1/n$, and $c(y) = \log[n!/((ny)!\,(n-ny)!)]$.

The Poisson distribution

The density function

$$f(y) = \mu^y \mathrm{e}^{-\mu}/y!$$

is rearranged in the general exponential form as

$$f(y) = \exp[y\log(\mu) - \mu - \log(y!)].$$

Therefore, $\theta = \log(\mu) = g(\mu)$, and we obtain the log as the canonical link function. Substituting for $\mu = \exp(\theta)$ gives

$$f(y) = \exp[y\theta - \exp(\theta) - \log(y!)].$$

Thus, $b(\theta) = \exp(\theta)$, $a = 1$ and $c(y) = -\log(y!)$.

The Poisson distribution with offset

The density function

$$f(y,t) = (\mu t)^{yt}\mathrm{e}^{-\mu t}/(yt)!$$

is rearranged in the general exponential form as

$$f(y,t) = \exp\{[y\log(\mu) - \mu]t + yt\log(t) - \log((yt)!)\}.$$

Thus, $\theta = \log(\mu)$ and again the log is the canonical link function. Substituting $\mu = \exp(\theta)$, we may write

$$f(y) = \exp\{[y\theta - \exp(\theta)]t + yt\log(t) - \log(yt!)\},$$

giving $b(\theta) = \exp(\theta)$, $a = 1/t$ and $c(y) = yt\log(yt) - \log(yt!)$.

The normal distribution

The normal distribution has both a location and a dispersion parameter. Although it is well known that a GLM is not necessary for analysing normal data, it is helpful to see this by showing that the canonical link function for the normal distribution is the identity function. The density function is

$$f(y) = \exp(-(y-\mu)^2/2\sigma^2)/\sqrt{(2\pi\sigma^2)},$$

which can be rearranged as

$$f(y) = \exp\left[\left(y\mu - \mu^2/2\right)/\sigma^2 - y^2/2\sigma^2 - \frac{1}{2}\log(2\pi\sigma^2)\right].$$

This is now in the general exponential form for two-parameter distributions. Thus, the canonical link is the identity function $\theta = g(\mu) = \mu$, and $\phi = \sigma^2$, $a(\phi) = \phi$, $b(\theta) = \theta^2/2$ and $c(y, \phi) = -[y^2/\phi + \log(2\pi\phi)]/2$.

3.1.7 Conditional logistic regression

This model does not, strictly speaking, form a GLM, but it is mentioned because it can be very useful for modelling binary data in datasets where there are only a few observations in each category of a fixed effect (e.g. there are only two observations per patient in a two-period, cross-over trial). We shall see later in Section 3.3.2 that fitting problems can arise in GLMs with fixed effects containing only a few observations per category. Using a conditional logistic regression analysis is one way in which these problems can be avoided. It works by omitting the 'problem' effect (e.g. patients in a cross-over trial) as a fixed effect in the model but instead the likelihood is 'conditioned' on this effect. Other effects are fitted as fixed just as in ordinary GLMs, and results can be interpreted as if the 'problem' effect had been fitted as fixed. A more complete description of the method can be found in Clayton and Hills (1993, Chapter 29) and Collett (1991, Section 7.7.1). The model can be fitted using `PROC LOGISTIC` with the 'problem' effect defined within a `STRATA` statement. An example of its use will be given in Section 8.5.4.

3.2 Generalised linear mixed models

GLMMs are based on extending the fixed effects GLM to include random effects, random coefficients and covariance patterns. In Section 3.2.1, we will specify a general form for the GLMM that encompasses all types of mixed model. In Section 3.2.2, we define the likelihood function for random effects and random coefficients GLMMs and introduce a similar function known as the

quasi-likelihood, which is required for fitting covariance pattern models. Following this, in Section 3.2.3, fitting methods for GLMMs will be outlined. These last two sections can be omitted by readers who do not desire a detailed understanding of the more theoretical aspects of fitting GLMMs.

3.2.1 The GLMM definition

The GLMM can be defined by

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}.$$

As in the GLM, $\boldsymbol{\mu}$ is the vector of expected means of the observations and is linked to the model parameters by a link function, g:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}.$$

$\mathbf{X}$ and $\mathbf{Z}$ are the fixed and random effects design matrices, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the vectors of fixed and random effects parameters as in the normal mixed model. The random effects, $\boldsymbol{\beta}$, can again be assumed to follow a normal distribution:

$$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{G})$$

and $\mathbf{G}$ is defined just as in Section 2.2. The variance matrix can be written as

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathrm{var}(\boldsymbol{\mu}) + \mathbf{R},$$

where $\mathbf{R}$ is the residual variance matrix, $\mathrm{var}(\mathbf{e})$. However, $\mathbf{V}$ is not as easily specified as it was for normal data where $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$. This is because $\boldsymbol{\mu}$ is not now a linear function of $\boldsymbol{\beta}$. A first-order approximation used by some fitting methods is

$$\mathbf{V} \approx \mathbf{BZGZ}'\mathbf{B} + \mathbf{R},$$

where $\mathbf{B}$ is a diagonal matrix of variances determined by the underlying distribution as described in Section 3.1 (e.g. $\mathbf{B} = \mathrm{diag}\{\mu_i(1-\mu_i)\}$ for binary data). In random effects and random coefficients models, the residual matrix, $\mathbf{R}$, is diagonal, since the residuals are assumed uncorrelated. The diagonal variance terms are equal to the expected variances given the underlying distribution, and thus $\mathbf{R} = \mathbf{AB}$ as in the GLMs. For random effects and random coefficients models, $\mathbf{V}$ can then be written as

$$\mathbf{V} \approx \mathbf{BZGZ}'\mathbf{B} + \mathbf{AB}.$$

In covariance pattern models, correlated residuals are allowed, and $\mathbf{R}$ can be expressed as a product of a correlation matrix defined on the linear scale, $\mathbf{P}$, and $\mathbf{AB}$:

$$\mathbf{R} = \mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{P}\mathbf{B}^{1/2}\mathbf{A}^{1/2}.$$

The reason for defining $\mathbf{P}$ on a linear scale is because it can then be parameterised to have a covariance pattern in the same way as normal data. We will be looking

more closely at how to define covariance patterns in Section 6.2. The approximation to $\mathbf{V}$ then becomes

$$\mathbf{V} \approx \mathbf{BZGZ'B} + \mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{PB}^{1/2}\mathbf{A}^{1/2}.$$

This formula can be considered a general form for $\mathbf{V}$, since by taking $\mathbf{P} = \mathbf{I}$ for random effects and random coefficients models, we obtain $\mathbf{V} \approx \mathbf{BZGZ'B} + \mathbf{AB}$ as shown previously.

The dispersion parameter

As in the GLMs, variance at the residual level can be increased (or decreased) by using a dispersion parameter. The residual variance is multiplied by the dispersion parameter, ϕ, so that

$$\mathbf{R} = \phi\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{PB}^{1/2}\mathbf{A}^{1/2}.$$

If the observed residual variance were exactly equal to that predicted ($\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{PB}^{1/2}\mathbf{A}^{1/2}$), then the dispersion parameter would equal one. However, often, this is not the case. The value of the dispersion parameter is influenced by several factors, and these will be considered in more detail in Section 3.3.7.

3.2.2 The likelihood and quasi-likelihood functions

As in normal mixed models, a popular way of fitting the GLMM is based on maximising the likelihood function for the model parameters. However, a difficulty with this is that true likelihood functions can only be defined for random effects and random coefficients models. A true likelihood function is not available for covariance pattern models, since a general multivariate distributional form does not exist for non-normal data (for normal data the multivariate normal distribution was used). However, we will show how it is possible to get around this difficulty by defining an alternative function known as the *quasi-likelihood* function, which has very similar properties to the likelihood function. In this section, we will specify the likelihood function for random effects and random coefficients models, define a quasi-likelihood function for covariance pattern models and then give a general form of the quasi-likelihood function that is appropriate for all types of mixed model.

The likelihood function for random effects and random coefficients models

For these models, we can obtain a true likelihood function from the product of the likelihoods based on $\mathbf{y}|\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ (by $\mathbf{y}|\boldsymbol{\beta}$ we mean $\mathbf{y}$ conditional on $\boldsymbol{\beta}$ so that $\boldsymbol{\beta}$ is treated as constant when defining the variance of $\mathbf{y}|\boldsymbol{\beta}$). A true likelihood function is possible because the distributions of $\mathbf{y}|\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ are known, and hence likelihood

functions can be formed from them. The likelihood for the fixed effects, $\boldsymbol{\alpha}$, and the variance parameters in the $\mathbf{G}$ matrix, $\boldsymbol{\gamma}_{\mathbf{G}}$, can be written as

$$L(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{G}}; \mathbf{y}) = L(\boldsymbol{\alpha}; \mathbf{y} \,|\, \boldsymbol{\beta}) L(\boldsymbol{\gamma}_{\mathbf{G}}; \boldsymbol{\beta}). \tag{B}$$

Now, $\boldsymbol{\beta}$ is assumed to have a multivariate normal distribution, $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{G})$, so substituting the multivariate normal density for $L(\boldsymbol{\gamma}_{\mathbf{G}}; \boldsymbol{\beta})$ we have

$$L(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{G}}; \mathbf{y}) \propto L(\boldsymbol{\alpha}; \mathbf{y} \,|\, \boldsymbol{\beta}) \,|\, \mathbf{G} \,|^{-1/2} \exp(-1/2 \boldsymbol{\beta}' \mathbf{G}^{-1} \boldsymbol{\beta}).$$

The $\mathbf{y}|\boldsymbol{\beta}$ are independent because we have assumed uncorrelated residuals ($\mathbf{R}$ is diagonal), and therefore $L(\boldsymbol{\alpha}; \mathbf{y}|\boldsymbol{\beta})$ is simply defined using the assumed distribution of $\mathbf{y}|\boldsymbol{\beta}$ (e.g. binomial, Poisson). This can be expressed using the same form obtained in Section 3.1.6 for the GLMs:

$$\begin{aligned} L(\alpha, \mathbf{y}|\boldsymbol{\beta}) &= \exp[\mathbf{y}' \mathbf{A}^{-1} \boldsymbol{\theta} - b(\boldsymbol{\theta})^{1/2'} \mathbf{A}^{-1} b(\boldsymbol{\theta})^{1/2} + K] \\ &\propto \exp[\mathbf{y}' \mathbf{A}^{-1} \boldsymbol{\theta} - b(\boldsymbol{\theta})^{1/2'} \mathbf{A}^{-1} b(\boldsymbol{\theta})^{1/2}], \end{aligned}$$

where

$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}$,
$\mathbf{A} = \mathrm{diag}\{a_i\}$, where a_i are constant terms (see Section 3.1.4),
$b(\boldsymbol{\theta}) = (b(\theta_1), b(\theta_2), \ldots, b(\theta_n))'$, where b is the function used in the general distributional form (see Section 3.1.3),
$K =$ constant.

The overall likelihood for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}_{\mathbf{G}}$ can then be expressed as

$$L(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{G}}; \mathbf{y}) \propto \exp[\mathbf{y}' \mathbf{A}^{-1} \boldsymbol{\theta} - b(\boldsymbol{\theta})^{1/2'} \mathbf{A}^{-1} b(\boldsymbol{\theta})^{1/2}] |\mathbf{G}|^{-1/2} \exp(-1/2 \boldsymbol{\beta}' \mathbf{G}^{-1} \boldsymbol{\beta}),$$

and the log likelihood as

$$\begin{aligned} \log\{\mathbf{L}(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{G}}; \mathbf{y})\} = \mathbf{y}' \mathbf{A}^{-1} \boldsymbol{\theta} - \mathbf{b}(\boldsymbol{\theta})^{1/2'} \mathbf{A}^{-1} \mathbf{b}(\boldsymbol{\theta})^{1/2} - 1/2 \log |\, \mathbf{G} \,| \\ - 1/2 \boldsymbol{\beta}' \mathbf{G}^{-1} \boldsymbol{\beta} + K. \end{aligned} \tag{C}$$

The quasi-likelihood function for covariance pattern models

In these models, the observations are correlated, and the model is parameterised by the fixed effects, $\boldsymbol{\alpha}$, and the variance parameters used in the $\mathbf{R}$ matrix, $\boldsymbol{\gamma}_{\mathbf{R}}$. However, since a general multivariate distributional form is not available for non-normal data, we cannot define a true likelihood function. This difficulty is overcome by instead specifying a quasi-likelihood function, $QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{R}}; \mathbf{y})$, which has similar properties to a true likelihood. It is defined so that the differential of its log with respect to $\boldsymbol{\alpha}$ has the same form as that of a true log likelihood with respect to $\boldsymbol{\alpha}$. The differential of the true log likelihood for random effects models (C) with respect to $\boldsymbol{\alpha}$ can be shown to be

$$\delta \log\{L(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{G}}; \mathbf{y})\} / \delta \boldsymbol{\alpha} = \mathbf{X}' \mathbf{A}^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

(it is obtained in a similar way to $\mathrm{d}\log(L)/\mathrm{d}\boldsymbol{\alpha}$ in Section 3.1.6).

Now since $\mathbf{R} = \mathbf{AB}$, this can be equivalently written by substituting $\mathbf{A}^{-1} = \mathbf{BR}^{-1}$ as

$$\delta \log\{L(\boldsymbol{\alpha}; \mathbf{y} \mid \boldsymbol{\beta})\}/\delta\boldsymbol{\alpha} = \mathbf{X'BR}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

where $\mathbf{R}$ is the residual covariance matrix. This form can then be used to define the differentiated log quasi-likelihood function for covariance pattern models, so we may write

$$\delta \log\{QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_\mathbf{R}; \mathbf{y})\}/\delta\boldsymbol{\alpha} = \mathbf{X'BR}^{-1}(\mathbf{y} - \boldsymbol{\mu}).$$

We note that some authors define the quasi-likelihood function as the log of the quasi-likelihood specified here. However, it would seem to make more sense to define it as we have so that it corresponds to the likelihood function.

A general quasi-likelihood function for all GLMMs

It is helpful to define a quasi-likelihood form that is appropriate for all types of GLMMs, which may contain random effects, coefficients and covariance patterns. This is obtained by replacing $L(\boldsymbol{\alpha}; \mathbf{y}|\boldsymbol{\beta})$ in (B) by $QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_\mathbf{R}; \mathbf{y}|\boldsymbol{\beta})$. Doing this, we obtain

$$QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y}) = QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_\mathbf{R}; \mathbf{y} \mid \boldsymbol{\beta})L(\boldsymbol{\gamma}_\mathbf{G}; \boldsymbol{\beta}),$$

and

$$\log\{QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y})\} = \log\{QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_\mathbf{R}; \mathbf{y} \mid \boldsymbol{\beta})\} - 1/2 \log|\mathbf{G}| - 1/2\boldsymbol{\beta'}\mathbf{G}^{-1}\boldsymbol{\beta} + K, \quad \text{(D)}$$

where

$$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_\mathbf{G}, \boldsymbol{\gamma}_\mathbf{R}).$$

This function will correspond to a true likelihood function whenever the residuals are uncorrelated (i.e. no $\boldsymbol{\gamma}_\mathbf{R}$ parameters are included). From now on, the term 'quasi-likelihood' will be used to infer either a true likelihood (for models with uncorrelated residuals) or a quasi-likelihood (for models with correlated residuals).

3.2.3 Fitting the GLMM

The quasi-likelihood is less straightforward to maximise than the likelihood function defined for normal mixed models. This is mainly because it is not now a linear function of $\boldsymbol{\alpha}$, and a solution for $\boldsymbol{\alpha}$ cannot be expressed directly in terms of the variance parameters. Several methods are available for maximising the quasi-likelihood. We are unable to recommend a single 'best' approach for fitting all types of GLMM; however, we will introduce two methods that are used within SAS procedures to maximise the quasi-likelihood: generalised estimating equations (GEEs) and pseudo-likelihood. We also describe how a Bayesian approach can be used. For those who wish for a greater understanding of the area, Breslow and Clayton (1993) give an in-depth coverage of approaches to fitting GLMMs.

Generalised estimating equations (GEEs)

This method was first suggested by Liang and Zeger (1986) and was initially developed for analysing covariance pattern models. This method is based on alternating between solutions for the fixed effects and for the variance parameters. The iterative procedure can be defined as follows:

1. A solution for the fixed effects is obtained while holding the variance parameters constant.
2. A solution for the variance parameters is obtained while holding the fixed effects constant.
3. The variance parameters are fixed at the values obtained in step 2, and a second solution for the fixed effects is obtained.
4. Steps 2 and 3 are repeated until the parameter estimates converge.

We now describe how solutions for the fixed effects and variance parameters are obtained at each step.

The fixed effects solution (steps 1 and 3) The solution for the fixed effects can be obtained by differentiating the log quasi-likelihood with respect to $\boldsymbol{\alpha}$ and setting the resulting expression to zero. Differentiating the log quasi-likelihood given by (D) previously and setting to zero gives

$$\mathbf{X}'\mathbf{B}\mathbf{R}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = 0.$$

This expression gives rise to n equations which are often referred to as *score equations* or *GEEs*. A solution for $\boldsymbol{\alpha}$ cannot be obtained by rearranging this expression, as was the case in normal mixed models, because the equations are non-linear in $\boldsymbol{\alpha}$ (recall that $\boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\alpha})$), and for this reason, the solution needs to be found iteratively. There are various methods available for obtaining the solution. Often, a linearised pseudo variable of the form that will be described for the pseudo-likelihood is used. However, we will not go into more detail on these methods in this book.

The variance parameters solution (step 2) Having obtained estimates of the fixed effects, the variance parameters are estimated in a similar way to that used for iterative generalised least squares (see Section 2.2). The matrix of products of the full residuals ($\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}$) is set equal to the variance matrix, $\mathbf{V}$, specified in terms of the variance parameters. This gives

$$\mathbf{V} = (\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})',$$

which leads to a set of $n \times n$ simultaneous equations (one for each element in the $n \times n$ matrices) that can be solved for the variance parameters (n = number of observations). The equations will have a similar form to those used in Section 2.2.4.

The equations do not take account of the fact that $\boldsymbol{\alpha}$ will be estimated and not known. However, as in normal mixed models, it is possible to adapt them to provide the unbiased REML estimators (see Section 2.2.1).

GEEs are used to fit covariance pattern models when the `REPEATED` statement is used in `PROC GENMOD`.

Pseudo-likelihood

This method can be used to fit both random effects models and covariance pattern models. It was proposed by Wolfinger and O'Connell (1993) and is one of the algorithms available within `PROC GLIMMIX` (see Section 9.3). Pseudo-likelihood maximises the quasi-likelihood by iteratively analysing a linearised pseudo variable (i.e. a transformation of $\mathbf{y}$ onto the linear scale) using weighted normal mixed models. The method is referred to as 'pseudo-likelihood' because the likelihood function maximised at each iteration is that of the pseudo variable and not that of the original data. The 'pseudo variable' introduced in Section 3.1.5 based on a first-order Taylor series expansion is again used:

$$\begin{aligned}\mathbf{z} &= g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1} \\ &= \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + (\mathbf{y} - \boldsymbol{\mu}))\mathbf{B}^{-1}.\end{aligned}$$

$\mathbf{z}$ has variance

$$\begin{aligned}\mathbf{V}_{\mathbf{z}} &= \text{var}(\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}) + \mathbf{B}^{-1}\text{var}(\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1} \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{B}^{-1}\mathbf{R}\mathbf{B}^{-1}.\end{aligned}$$

By rewriting the residual matrix $\mathbf{R}$ as a product of a correlation matrix on the linear scale, $\mathbf{P}$, and $\mathbf{A}\mathbf{B}^{-1}$, $\mathbf{V}_{\mathbf{z}}$ can be re-expressed as

$$\mathbf{V}_{\mathbf{z}} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{A}^{1/2}\mathbf{B}^{-1/2}\mathbf{P}\mathbf{B}^{-1/2}\mathbf{A}^{1/2}.$$

In random effects and random coefficients models, the residuals are uncorrelated and $\mathbf{P} = \mathbf{I}$, so $\mathbf{V}_{\mathbf{z}}$ then simplifies to

$$\mathbf{V}_{\mathbf{z}} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{A}\mathbf{B}^{-1}.$$

Conditioning the $\mathbf{z}$ on $\boldsymbol{\beta}$ allows $\mathbf{Z}\mathbf{G}\mathbf{Z}'$ to be omitted from the variance matrix formula. $\mathbf{z} \mid \boldsymbol{\beta}$ has the following multivariate normal distribution:

$$\mathbf{z}|\boldsymbol{\beta} \sim \text{N}(\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}, \mathbf{A}^{1/2}\mathbf{B}^{-1/2}\mathbf{P}\mathbf{B}^{-1/2}\mathbf{A}^{1/2}).$$

$\mathbf{z}$ can now be analysed as a weighted normal mixed model with residual matrix $\mathbf{P}$ (defined on the linear scale) and diagonal weight matrix $\mathbf{A}^{-1}\mathbf{B}$ (inverse product of pre- and post-multipliers of $\mathbf{P}$, $\mathbf{A}^{1/2}\mathbf{B}^{-1/2}\mathbf{B}^{-1/2}\mathbf{A}^{1/2}$). Because $\mathbf{z}$ and $\mathbf{B}$ are dependent on the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the normal mixed model needs to be fitted iteratively. Any of the methods described in Chapter 2 can be used to do this. However, again, variance parameters will be biased downwards to some extent if maximum likelihood (ML) or IGLS are used, but this problem is not expected with REML and RIGLS. We will now define the iterative procedure more explicitly.

The iterative procedure The raw data, $\mathbf{y}$, can be taken as initial values for $\mathbf{z}$, and the identity matrix, $\mathbf{I}$, can be used for the initial weight matrix. Alternatively, $g(\mathbf{y})$ can be taken as initial values for $\mathbf{z}$, with an adjustment if necessary to prevent infinite values. A weighted normal mixed model is then fitted using a method such as REML. This completes the first iteration and provides initial estimates for $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$. New values of $\mathbf{z}$ and $\mathbf{B}$ are calculated using $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ and from these a second weight matrix, $\mathbf{A}^{-1}\mathbf{B}$. A second weighted mixed model is then fitted, and the process is continued until the parameter estimates converge (i.e. until parameter values change very little between successive iterations). This method is computationally costly because normal mixed models, themselves requiring iterations, are fitted iteratively.

Fixed and random effects variance Once the model has converged, the fixed and random effects and their variances can be calculated using the formula specified for normal mixed models in Sections 2.2.3 and 2.2.4:

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{z},$$
$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1},$$

and, for random effects

$$\widehat{\boldsymbol{\beta}} = \mathbf{G}\mathbf{Z}'\mathbf{V}_{\mathbf{z}}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\alpha}),$$
$$\text{var}(\widehat{\boldsymbol{\beta}}) = \mathbf{G}\mathbf{Z}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{Z}\mathbf{G} - \mathbf{G}\mathbf{Z}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{Z}\mathbf{G}.$$

Just as in normal mixed models, because $\mathbf{V}_{\mathbf{z}}$ is estimated and not known, there can be a small amount of downward bias in $\text{var}(\widehat{\boldsymbol{\alpha}})$ and $\text{var}(\widehat{\boldsymbol{\beta}})$ (see Sections 2.4.3 and 3.3.6).

The results obtained from GEEs and pseudo-likelihood are expected to be similar; however, there are often slight differences due to the different computational approaches.

Bayesian methods

Bayesian methods can be used to fit GLMMs in as much the same way as normal mixed models. Again, they have been developed most fully for fitting random effects and random coefficients models where the true likelihood function is available. The posterior distribution of the parameters can be determined in the same way as defined in Section 2.3.2 except that a non-normal distribution is assumed for the data. The resulting posterior density surface can then be used to provide estimates, standard deviations and probability intervals for the model parameters. As in normal mixed models, some of the bias problems are overcome when a Bayesian approach is used, although we will see later that the potential problems associated with random effects shrinkage are not avoided. A GLMM may be fitted in SAS using a Bayesian approach using `PROC MCMC`.

3.3 Practical application and interpretation

In this section, some points relating to the practical application and interpretation of GLMs and GLMMs will be covered. Experience with these models is still limited, and therefore some of the issues are still far from resolved.

3.3.1 Specifying binary data

Binary data can be specified either as a series of zeros and ones (*Bernoulli form*) or as frequencies of 'success' out of a number of 'tries' (*binomial form*). If data are recorded in Bernoulli form (as is often the case in clinical trials), then it is usually most convenient to analyse them as such. This also has the advantage that other measurements made at the observation level (e.g. baseline effects) can be included in the model as covariates. If there are no baseline effects, then data can alternatively be aggregated to give frequencies (e.g. at the centre·treatment level) and analysed in binomial form. In this situation where data are in binomial form, we would suggest that the dispersion parameter is fixed at one, and a random effect is fitted at the observation level (e.g. an effect for each centre·treatment frequency). This approach might be appropriate for an analysis of centre·treatment frequencies of an adverse event from a multi-centre trial. Fitting centre·treatment effects as random with the dispersion parameter fixed at one will allow variation at the observation level to be modelled by the centre·treatment variance component. The meta-analysis example in Section 5.7 shows how binomial trial·treatment frequencies can be analysed in this way.

Analysing binomial frequencies is less intensive computationally than using Bernoulli data. However, results will differ to some extent between the two analyses because variance at the residual level is modelled separately from that at the random effects level in a Bernoulli model. The difference will be more noticeable in datasets where there are uniform random effects categories. In this situation, we would prefer an analysis of the data in Bernoulli form because the dispersion parameter can then help overcome any biases caused by random effects shrinkage.

3.3.2 Uniform effects categories

We define a fixed or random effects category as 'uniform' if an infinite value is obtained when the link function is applied to the mean of observations in the category. It occurs when all observations within the category are zero in binary and Poisson data or when all observations are n_i/n_i in binary data. For example, centre effects are fitted in some of the analysis models we will use in Section 3.4 to analyse an adverse event 'cold feet' in a multi-centre trial. In some centres, no subjects were recorded as having cold feet, and this caused these centre effects to be uniform.

For *uniform fixed effects*, a corresponding effect estimate on the linear scale cannot then be estimated and will to tend towards plus or minus infinity. For example, if in a simple between-patient trial the frequency of success for one of the treatments is 100%, then a model using a logit link function would attempt to estimate the treatment mean as log(1.00/0.00).

However, sensible estimates of *uniform random effects* (and their corresponding variance components) can still be obtained, provided not all categories within a particular random effect are uniform (although in some circumstances the estimates of the variance components may be biased, see Section 3.3.5). This is possible because the random effects estimates are shrunken, and thus information from all observations, not just those for the particular effect, is used in forming the estimates. However, when all categories of a random effect are uniform, the model will not converge. In other situations, whether convergence is achieved is less predictable and depends on the number of uniform categories, the number of fixed effects fitted and on how the fixed effects relate to random effects.

Uniform categories are most likely when the probability of success is very small or large in binary data or when the event rate is very small in Poisson data. They are also likely when there are small numbers of observations within some of the fixed effects categories.

Uniform fixed effects categories are easily identified in the results by estimates and standard errors that are extremely large. When this occurs, the results given for other effects are in fact equivalent to a reanalysis of the data, with observations corresponding to the uniform category removed, and thus they can still be used. If it is important to test the overall significance of a set of fixed effects containing a uniform category (e.g. treatments), this can be done by comparing the likelihoods (or quasi-likelihoods) between models that include and exclude the effects using a likelihood ratio test (see Section 2.4.4).

Alternatively, when a fixed effects model is being fitted to binary data, an exact logistic regression where all possible combinations of data values are considered can be used. This can be carried out using the `EXACT` statement in `PROC LOGISTIC`.

Uniform random effects categories are often indicated by a low dispersion parameter or a lack of convergence.

3.3.3 Negative variance components

Negative estimates of variance components can occur just as in the normal mixed model, and the points made in Section 2.4.1 apply. An additional influence in the GLMM making a negative variance component estimate more likely is the possible bias due to the effects of random effects shrinkage (see section 3.3.1). When a Bayesian method is used, usually, a non-informative prior distribution will be set for variance components, and therefore it is not possible for the posterior distribution of the variance components to include negative values.

3.3.4 Presentation of fixed and random effects estimates

In this section, we look at how results for fixed and random effects estimates can be presented and interpreted.

The logit link function

The logit link function, $\log(\mu/(1-\mu))$, is the canonical link function for Bernoulli and binomial distributions (see Section 3.1.4). Fixed effects GLMs using this link function are often referred to as *logistic regression* analyses. Results for both GLMs and GLMMs are obtained on the logit scale and can be expressed in terms of *odds ratios (ORs)* when exponentiated. To see this, we consider a simple, single-centre, parallel group trial to compare two treatments, with a binary outcome. The model could be written as

$$\log(\mu/(1-\mu)) = a + bx,$$

where x is an indicator variable denoting treatment (say, one if the treatment is A and zero if the treatment is B).

Thus, using p_{A} and p_{B} to denote the probabilities of success,

$$\log(p_{\mathrm{A}}/(1-p_{\mathrm{A}})) = a + b,$$

$$\log(p_{\mathrm{B}}/(1-p_{\mathrm{B}})) = a,$$

and, by subtraction,

$$\log(p_{\mathrm{A}}/(1-p_{\mathrm{A}})) - \log(p_{\mathrm{B}}/(1-p_{\mathrm{B}})) = b.$$

Hence,

$$\log \frac{(p_{\mathrm{A}}/(1-p_{\mathrm{A}}))}{(p_{\mathrm{B}}/(1-p_{\mathrm{B}}))} = b,$$

and, on exponentiating,

$$\frac{p_{\mathrm{A}}/(1-p_{\mathrm{A}})}{p_{\mathrm{B}}/(1-p_{\mathrm{B}})} = \mathrm{e}^{b}.$$

The numerator of this expression gives us the odds of success on treatment A (i.e. the probability of success divided by the probability of failure). Similarly, the denominator is the odds of success on treatment B, leading to e^{b} being the estimate of the OR.

Note that software packages vary in their parameterisations of the fixed effects, and therefore care is needed to ensure the direction of effects is interpreted correctly. Calculation of ORs will be illustrated in the analysis of cold feet in the hypertension trial in Section 3.4.

The log link function

The log function is the canonical link function for Poisson distributions. Models using this link function can be described as *log-linear* models. Results are obtained

on the log scale, which, as for logistic regression, is not easy to interpret directly. In a similar way, though, exponentiating the coefficients allows them to be expressed this time in terms of *relative rates (RRs)*. The RR is simply the ratio of two event rates; for example, the rate on treatment A divided by the rate on treatment B can be used to compare the two treatments:

$$\text{RR} = \frac{\text{Rate of event on treatment A}}{\text{Rate of event on treatment B}}.$$

The calculation of an RR will be illustrated in the analysis of epileptic seizure frequencies in Section 6.4.

If the log link function is used to analyse data that are binary, the rates then become risks (or probabilities), and RRs are then sometimes instead referred to as *relative risks*.

3.3.5 Accuracy of variance parameters and potential bias

It is important that variance parameters are estimated with a reasonable accuracy because of their effect on the calculation of fixed effects and their standard errors. As in normal mixed models, the accuracy of variance parameters is dependent on the number of DF used to estimate them. Although there are no hard and fast rules, it would seem inadvisable to fit an effect as random if less than about five DF were available (e.g. a multi-centre trial with five or less centres).

An additional cause of bias in GLMMs is due to the shrinkage of random effects estimates, particularly when there are uniform random categories. The random effects GLMM assumes that $\mathbf{y}|\boldsymbol{\beta}$ has a distribution where the variance (in one-parameter distributions) is defined as a function of the expected means (e.g. $\boldsymbol{\mu}$ ($\mathbf{1} - \boldsymbol{\mu}$) and $\boldsymbol{\mu}$ for binomial and Poisson data, respectively). However, this relationship between the mean and variance does not hold exactly because random effects estimates are shrunken compared with their raw means.

For binary data, this causes the predicted residual variance, $\mathbf{AB}$, to be greater than that observed (i.e. $\text{var}(\mathbf{y} - g^{-1}(\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}))$) whenever shrinkage occurs towards 0.5 and less than that observed otherwise. This can be seen for binomial data by considering observations within a particular random effect category. If their raw mean is μ and their shrunken mean is μ_s, their observed variance can be shown to be $\text{var}(y) = \Sigma_i(y_i - \mu_s)^2/n = \mu_s^2 - 2\mu\mu_s + \mu$. This is greater than the predicted variance, $\mu_s(1 - \mu_s)$, whenever μ_s is in the range $(\mu, 0.5)$ for $\mu < 0.5$ or $(0.5, \mu)$ for $\mu > 0.5$. For uniform categories, shrinkage is always towards 0.5, and the predicted variance is therefore always greater than that observed. In datasets with several uniform random effect categories, this discrepancy can cause appreciable bias in the variance component estimates, particularly when methods using a linear approximation are used, such as pseudo-likelihood. Bayesian methods do not use approximations of the linear part of the model, and their results are expected to be less prone to bias in the presence of uniform random categories. When no random effect categories are uniform, the variance parameter bias is likely to be small.

For Poisson data, it can be shown that $\text{var}(y) = \mu_s^2 - 2\mu\mu_s + \mu^2 + \mu$, where μ is the raw mean. This is greater than the predicted variance, μ_s, except when μ_s is in the range $(\mu, \mu + 1)$. For uniform categories (where $\mu = 0$), the shrunken estimate may or may not be in the range $(0, 1)$, and so the predicted variance can be either less than or greater than that predicted. For this reason, the problem of biased variance components is less likely in analyses of Poisson data.

3.3.6 Bias in fixed and random effects standard errors

Standard errors of fixed and random effects are calculated from variance parameters, and therefore any bias in these parameters will also cause bias in the fixed effects standard errors. In addition, a downward bias can occur for the same reasons given for normal mixed models whenever an effect is estimated using information from several error strata (see Chapter 2). However, this second source of bias is usually small and can be adjusted for using the Kenward–Roger option available in the `GLIMMIX` procedure. In `SAS/STAT` 12.1 and later versions, there is an alternative option (`DDFM=KR2`) that obtains Kenward and Roger's improved adjustment (Kenward and Roger, 2009) and further reduces the bias in certain types of covariance pattern model. Among the examples considered in this text, this new option would provide a different standard error to the `DDFM=KR` option for the first-order autoregressive structure and the heterogeneous structures ((vi)–(viii)) described in Section 6.2.1.

The 'empirical' variance estimator has been suggested as a more robust estimator of the fixed effects variance for covariance pattern models. This estimator takes into account the observed covariance in the data and may help alleviate some of the bias in variance parameter estimates and any misspecification of the covariance pattern. However, it is not appropriate if there are any subjects with only one observation. There are two ways the empirical variance can be calculated. If the analysis is based on a linearised approximation (e.g. pseudo-likelihood), it can be based directly on the linearised data and their variance matrix at the last iteration. For example, in pseudo-likelihood the approximation $\mathbf{z} = g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1}$ is used and the empirical estimator of $\text{var}(\widehat{\boldsymbol{\alpha}})$ can be calculated as

$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\,\text{cov}(\mathbf{z})\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{V}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}.$$

Alternatively, it can be based on the raw data and its variance matrix by substituting a 'linked' design matrix, $\mathbf{BX}$, for the $\mathbf{X}$ matrix in the formula given in Chapter 2 (i.e. the usual design matrix $\mathbf{X}$ is pre-multiplied by the diagonal matrix of expected observation variances, $\mathbf{B}$) to give

$$\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{B}\mathbf{V}^{-1}\,\mathbf{B}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}\mathbf{V}^{-1}\;\;\text{cov}(\mathbf{y})\,\mathbf{V}^{-1}\mathbf{B}\mathbf{X}(\mathbf{X}'\mathbf{B}\mathbf{V}^{-1}\mathbf{B}\mathbf{X})^{-1}.$$

However, the empirical variance is known to be biased (see `SAS PROC GLIMMIX` documentation) and studies of normal data have found it to be less reliable than

the model-based variance, particularly in small datasets. It is possible that careful modelling of the covariance pattern will provide more reliable estimates of fixed effects variances than the empirical estimator. Alternatively a variety of bias adjustments to the empirical estimator are available in `PROC GLIMMIX`, but we have had no experience in their application.

The different influences on biases both in variance components and effect standard errors are summarised in Table 3.1, and suggested actions are made for the different situations. When a random effects model appears unsuitable, an alternative may be to reparametrise the model in terms of a covariance pattern model. For example, instead of fitting patient effects as random in a crossover trial, a covariance pattern model allowing correlations between the repeated observations on the same patients could instead be fitted.

Table 3.1 Influences on variance component and fixed effects standard error biases.

Problem	Influences	Comments and suggestions
Variance component bias and inaccuracy	Too few random effects	If less than about five categories, inadvisable to fit effect as random
	Uniform random effect categories	If too many (perhaps greater than about 30%) categories are uniform, it may be inadvisable to fit effect as random Influence could be assessed by fitting a dispersion parameter (see Section 3.3.7) and assessing if it is very different from one Uniform categories most likely in situations with few observations per category (e.g. two-way crossover with two observations per patient)
	Approximations used in likelihood-based fitting method	Likely to accentuate any bias due to too few random effects and uniform random effect categories This influence can be eliminated by using a Bayesian approach If enough random effect categories and few uniform categories, a likelihood-based approach may also be satisfactory
Downward bias in fixed effects SEs	Formula assuming variance components known	Overcome using Bayesian approach or Kenward–Roger adjustment More bias when: fewer random effects, fewer observations per random effect, more imbalance (as in normal mixed models)
	Variance component bias	As indicated above

3.3.7 The dispersion parameter

The dispersion parameter allows the residual variances to increase or decrease from their predicted values; the residual variance matrix is taken to be $\mathbf{R} = \phi\mathbf{AB}$ in GLMs and $\mathbf{R} = \phi\mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{P}\mathbf{A}^{1/2}\mathbf{B}^{1/2}$ in GLMMs.

The GLM

In the GLM, it is only advisable to fit a dispersion parameter when binomial or Poisson frequencies are being modeled. It will then model any *over-* or *under-dispersion* and will have a role similar to a variance component at the residual level. Observed residual variation may be greater or less than that predicted by the underlying distribution of the data (i.e. $\mathbf{R} = \mathbf{AB}$), and the dispersion parameter will take account of this to some extent. For example, if trial·treatment frequencies from a meta-analysis were in binomial form and trial and treatment effects were fitted, the dispersion parameter would reflect the trial·treatment variation. A value greater than one indicates more variation than expected by chance. This is equivalent to obtaining a positive trial·treatment variance component and can be referred to as 'over-dispersion'. Conversely, when the dispersion parameter is less than one, it is equivalent to obtaining a negative trial·treatment variance component estimate that can be referred to as 'under-dispersion'. In this situation, a decision needs to be made as to whether this lower than expected variance is genuine. If the lower variance does not seem plausible, then the dispersion parameter should be omitted from the model. This is equivalent to fixing a negative variance component at zero.

When data are in Bernoulli form, the observed variation is usually almost exactly equal to the predicted variances given by the diagonal terms in $\mathbf{AB}$, and the dispersion parameter will be close to one. If, however, there are uniform effect categories, its value will be less than one, but in that situation, a GLM is not advisable. Thus, there is no reason to include the parameter when data are in Bernoulli form.

The GLMM

In the GLMM, the dispersion parameter can be influenced by both over- or under-dispersion of the data and by the effects of random effects shrinkage. Random effects shrinkage can cause the predicted residual variance to be greater than that observed (particularly when uniform random effect categories are present in binary data), and this is likely to cause a smaller value of the dispersion parameter. Thus, interpretation of the dispersion parameter in the GLMM can be difficult, since it is not always clear which factors have affected it.

When binomial data (i.e. observations expressed as frequencies over denominators) with no uniform random effect categories are analysed, the dispersion value can be interpreted largely in terms of over- or under-dispersion. However, it is usually preferable to analyse binomial data by fitting a random effect at the observation level (i.e. with each observation taken as a separate category) rather

than fitting a dispersion parameter. This more adequately allows for differences in the sizes of the denominators and the influence in determining variance at the observation level.

In binomial datasets where there are uniform random effect categories, it is more difficult to differentiate between the influences of random effects shrinkage and of over- or under-dispersion (for this reason, we would usually recommend that such datasets are analysed in Bernoulli form.) However, a dispersion parameter of less than one is most likely to be due to random effects shrinkage, whereas a dispersion parameter greater than one is likely to indicate over-dispersion, since random effects shrinkage is expected to cause the dispersion parameter to decrease.

In Bernoulli data, values of the dispersion parameter that are noticeably different to one are likely to indicate the model is not fitting well and that estimates of the variance components are likely to be biased. This is most likely to occur when there are uniform random categories. If the value is considerably less than one, then a simulation approach could be tried in place of a likelihood-based method. However, when there is a high proportion of uniform categories, it may be more appropriate to respecify the model, either not fitting the offending random effect or in terms of a covariance pattern model with correlated observations rather than random effects.

In Poisson datasets, the dispersion value can usually be interpreted largely in terms of over- or under-dispersion. This is because uniform random categories do not usually affect its value greatly.

Should a dispersion parameter be fitted? We believe that it is usually helpful to include a dispersion parameter unless a random effect is being fitted to Bernoulli data at the residual level. Its role will be either to model genuine over- or under-dispersion in binomial or Poisson data or to help overcome discrepancies between the observed and predicted residual variation caused by random effects shrinkage. It can also be a useful diagnostic aid to help determine when the GLMM may be misspecified or inappropriate.

Covariance pattern models

The dispersion parameter in covariance pattern models fitting no random effects has the same role as in GLMs and the points made previously for GLMs apply.

3.3.8 Significance testing

GLMs

Fixed effects can be tested using chi-squared tests. These are based on asymptotic theory; that is, as sample sizes become larger, the distribution of fixed effects estimates conform more closely to normal distributions, the standard deviations

of which are increasingly well estimated by fixed effects standard errors. If a dispersion parameter has been fitted, it may be more appropriate to use instead Wald F tests to take account of the fact that a residual variance parameter is estimated. The Wald F statistic is calculated as described in Section 2.4.4. However, note that the F test still relies on the asymptotic normality of the fixed effects just as the chi-squared test does. Theory to underpin the analysis of datasets with small sample sizes is, however, remarkably sparse.

GLMMs

We first consider significance tests when a likelihood-based approach is used.

Fixed and random effects These effects can be tested using Wald F and t tests calculated on the linear scale as described in Section 2.4.4. These tests are preferable to chi-squared tests, as they take into account the uncertainty in the variance parameters (including the dispersion parameter). However, both the chi-squared and F tests rely on the asymptotic normality of the fixed effects estimates. The Satterthwaite DF can be calculated for the F (and t) tests in just the same way as for normal mixed models. However, there is a situation where these DF would not be appropriate, although it is a situation that we have already recommended should be avoided. If neither a dispersion parameter, nor a variance component is modelled at the residual level, a conservative estimate could be taken as the lowest DF of the error strata used for estimating the fixed effect. If a Bayesian method is used, then tests can be performed by calculating exact 'Bayesian' p-values from the marginal posterior distribution for each effect (see Section 2.3.3).

Variance parameters A variance parameter can be tested by comparing likelihoods (or quasi-likelihoods) between models fitting and not fitting the parameter using a likelihood ratio test as described in Section 2.4.4. The theoretical basis for this approach has only been proved for true likelihoods, although we believe that quasi-likelihood-based tests will still give good approximations. In `SAS`, `PROC GLIMMIX` outputs pseudo-likelihoods and the `COVTEST` statement provide statistical inferences about the variance parameters for a limited range of situations (see the on-line `SAS` help for further details). Note that Bayesian models are usually set up to sample only positive variance components, and in that situation, variance components cannot be tested for significance.

3.3.9 Confidence intervals

When a likelihood-based approach such as pseudo-likelihood is used, confidence intervals for an effect may be obtained using the mean and standard error estimates given on the linear (linked) scale. These confidence intervals can

often be converted to a more interpretable scale, usually by exponentiation. For example, ORs can be obtained when the logit link function is used, as well as relative risks for the log link function.

Percentage points from the t distribution with the DF calculated as described in Section 2.4.4 can be used to take into account the fact that the variance and dispersion parameters are estimated. If the software used does not provide t statistics, the normal distribution value of 1.96 can be substituted to provide a more approximate (and unduly narrow) confidence interval.

$$\text{Lower } 95\% \text{ confidence limit} = \text{mean effect} - t_{\text{DF},0.975} \times \text{SE},$$

$$\text{Upper } 95\% \text{ confidence limit} = \text{mean effect} + t_{\text{DF},0.975} \times \text{SE}.$$

If a Bayesian approach is used, exact probability intervals may be obtained from the marginal posterior distribution for each effect (see Section 2.3). This is demonstrated in Sections 2.5 and 3.4.

3.3.10 Model checking

In the GLMM, it is assumed that the random effects are normally distributed and uncorrelated. As for normal mixed models (see Section 2.4.6), fixed effects and variance component estimates are not usually sensitive to a misspecification of the random effects distribution. However, random effect predictions may be affected when their distribution is misspecified and, if of interest, should be interpreted cautiously. Plots of the random effects are helpful for identifying any outlying effect categories but may not always detect a lack of normality.

The residuals $\mathbf{y} - \hat{\boldsymbol{\mu}}$ do not need to be checked for Bernoulli data. However, if the data are binomial (i.e. expressed as frequencies with denominators) or Poisson, then residual plots can be used to identify outlying observations. It should be borne in mind that normal plots will not necessarily produce straight lines. The residuals, $\mathbf{y} - \hat{\boldsymbol{\mu}}$, do not have equal variances, and therefore they should first be linearised by dividing by their predicted standard errors, $\mathrm{A}^{1/2}\mathrm{B}^{1/2}$ (e.g. $(\mu_i(1-\mu_i))^{1/2}$ for binary data) to give the 'Pearson' residuals. Note that when data are analysed using a linearised pseudo variable (as in pseudo-likelihood), the 'pseudo' residuals will already be on a linear scale and can be checked directly.

3.3.11 Determining whether the simulated posterior distribution has converged

When a Bayesian analysis is carried out, the simulated posterior should be examined to obtain reassurance that it has converged, that is formed a stationary distribution. The approaches described for normal data in Section 2.4.8 may be used.

3.4 Example

3.4.1 Introduction and models fitted

The multi-centre trial of treatments for lowering blood pressure introduced in Section 1.3 is used again in this section. An adverse event, 'cold feet', is analysed as a binary variable, and observations at the final or last attended visit are used. Cold feet was, in fact, recorded on a scale of 1–5: 1 = none, 2 = occasionally, 3 = on most days, 4 = most of the time and 5 = all of the time. A binary 'cold feet' variable was created by taking categories 1 and 2 as negative and categories 3–5 as positive. The frequencies of cold feet by treatment and centre are shown in Table 3.2. In this trial, 'cold feet' was recorded at baseline so, in order to include a baseline covariate in the model (and so reduce between-patient variation), we will analyse the data in Bernoulli form.

Table 3.2 indicates that there are several zero frequencies of cold feet, and these will lead to uniform centre and centre·treatment categories. This in turn may cause variance component bias (see Section 3.3.5), and it is not clear whether a random effects model will be satisfactory. In this section, we will fit a variety of models (see Table 3.3) and discuss their strengths and weaknesses. In practice, only Model 1 is likely to be considered as a fixed effects model, since the large number of uniform categories will cause problems in estimating satisfactory treatment effects in Models 2 and 3 (as discussed in Section 3.4.2). In Model 4, centre effects are fitted as random, and in Model 5, both centre and centre·treatment effects are fitted as random. Model 5 takes into account the random variation in the treatment effect between centres, and results can be related with more confidence to the 'population' of potential centres. Models 4 and 5 are both fitted using pseudo-likelihood.

Models 6 and 7 are the same as Models 4 and 5, except that they are fitted using a Bayesian model with non-informative priors to obtain a joint (posterior) distribution of the model parameters. The Bayesian models are set up in a similar way to the normal example described in Section 2.5, except that Bernoulli distributions are now assumed for the observations. Again, normal distributions with zero means and very large variances (of 10,000) were used as non-informative priors for the fixed effects (baseline and treatment), and inverse gamma distributions with very small parameters (of 0.01) were used as non-informative prior distributions for the centre and centre·treatment variance components. Note that this prior specification for the variance components ensures that negative variance component samples cannot be obtained. Five hundred thousand samples were taken using MCMC (see Section 2.3.5) and a thinning factor of 10 to provide the posterior distribution of the model parameters. The values sampled were then used directly to obtain parameters estimates, standard deviations, probability intervals and 'Bayesian' p-values in exactly the same way as described for the example given in Section 2.5.

Table 3.2 Frequencies of cold feet by treatment and centre.

	Treatment			
Centre	**A**	**B**	**C**	**Total**
1	3/13	5/14	1/12	9/39
2	2/3	0/4	0/3	2/10
3	0/3	0/3	0/2	0/8
4	1/4	1/4	0/4	2/12
5	1/4	3/5	0/2	4/11
6	0/2	1/1	1/2	2/5
7	0/6	1/6	0/6	1/18
8	1/2	0/1	1/2	2/5
9	–	–	0/1	0/1
11	0/4	1/4	0/4	1/12
12	0/3	1/3	0/4	1/10
13	1/1	0/1	0/2	1/4
14	0/8	2/8	1/8	3/24
15	1/4	0/4	0/3	1/11
18	0/2	0/2	0/2	0/6
23	1/1	–	0/2	1/3
24	–	–	0/1	0/1
25	0/3	0/2	0/2	0/7
26	0/3	1/4	0/3	1/10
27	–	1/1	0/1	1/2
29	1/1	–	0/1	1/2
30	0/1	0/2	0/2	0/5
31	0/12	0/12	0/12	0/36
32	1/2	0/1	0/1	1/4
35	0/2	0/1	–	0/3
36	0/9	5/6	0/8	5/23
37	0/2	0/1	1/2	1/5
40	0/1	1/1	–	1/2
41	0/2	0/1	0/1	0/4
Total	13/98	23/92	5/93	41/283

3.4.2 Results

Estimates of the variance components and fixed effects for each model are shown in Table 3.4.

Fixed effects models (1–3)

Several uniform effects categories occur in Models 2 and 3. These categories are easily identified in the results by estimates and standard errors that are

Table 3.3 Models used to analyse 'cold feet' in a multi-centre trial.

Model	Fixed effects	Random effects	Method
1	Baseline, treatment	–	GLM
2	Baseline, treatment, centre	–	GLM
3	Baseline, treatment, centre·treatment	–	GLM
4	Baseline, treatment	Centre	P-L[a]
5	Baseline, treatment	Centre, centre·treatment	P-L[a]
6	Baseline, treatment	Centre	Bayes
7	Baseline, treatment	Centre, centre·treatment	Bayes

[a]P - L = pseudo - likelihood.

Table 3.4 Estimates of variance components and fixed effects (on the logit scale).

	Variance components			
Model	**Centre**	**Treatment·centre**	**Dispersion parameter**	**− 2log(*L*)**
1	–	–	1.00[a] (1.01)	178.17
2	–	–	1.00[a] (0.79)	147.78
3	–	–	1.00[a] (0.53)	100.05
4	0.09	–	0.94	–
5	0.00	1.88	0.54	–
6	0.08[b]	–	1.00[a]	–
7	0.09[b]	0.53[b]	1.00[a]	–

	Treatment effects (SEs)			
Model	**Baseline**	**A − B**	**A − C**	**B − C**
1	2.97 (0.49)	−0.77 (0.44)	0.94 (0.60)	1.70 (0.57)
2	2.67 (0.55)	−0.98 (0.49)	1.05 (0.65)	2.03 (0.63)
3	3.09 (0.80)	–	–	–
4	2.91 (0.48)	−0.77 (0.43)	0.93 (0.58)	1.70 (0.56)
5	3.06 (0.46)	−0.59 (0.57)	1.18 (0.66)	1.78 (0.66)
6	3.06 (0.52)	−0.81 (0.47)	1.01 (0.63)	1.82 (0.60)
7	3.27 (0.61)	−0.78 (0.57)	1.18 (0.74)	1.96 (0.72)

[a]Dispersion parameter is fixed at one (value in brackets is its estimate).
[b]Estimates are median values from the marginal posterior distributions.

extremely large. The fixed effects estimates resulting from Model 2 are listed in Table 3.5. The uniform centre categories can be identified as centres 3, 9, 18, 24, 25, 30, 31 and 35. (Note that estimates for these centres become very large when the intercept term of − 26.6 is added, while estimates for other centres are then more reasonable.) These are the centres where no patients had cold feet.

Table 3.5 Fixed effects estimates resulting from Model 2.

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr>ChiSq
Intercept		1	−26.5793	2.1509	−30.7950	−22.3635	152.70	<.0001
cf1		1	2.6711	0.5548	1.5838	3.7584	23.18	<.0001
treat	A	1	1.0470	0.6477	−0.2224	2.3164	2.61	0.1060
treat	B	1	2.0302	0.6294	0.7967	3.2637	10.41	0.0013
treat	C	0	0.0000	0.0000	0.0000	0.0000	.00	.0001
centre	1	1	23.4900	2.1070	19.3604	27.6197	124.29	<.0001
centre	2	1	23.2142	2.2258	18.8516	27.5768	108.77	<.0001
centre	3	1	−0.1467	110060.4	−215715	215714.3	0.00	1.0000
centre	4	1	23.2940	2.2582	18.8680	27.7201	106.40	<.0001
centre	5	1	24.1606	2.1989	19.8508	28.4704	120.72	<.0001
centre	6	1	25.3009	2.2938	20.8051	29.7966	121.66	<.0001
centre	7	1	21.8392	2.3609	17.2120	26.4665	85.57	<.0001
centre	8	1	24.7352	2.2914	20.2442	29.2262	116.53	<.0001
centre	9	1	1.2139	322114.2	−631331	631333.5	0.00	1.0000
centre	11	1	22.6362	2.3220	18.0852	27.1872	95.04	<.0001
centre	12	1	22.1269	2.4024	17.4183	26.8355	84.83	<.0001
centre	13	1	24.5423	2.4163	19.8066	29.2781	103.17	<.0001
centre	14	1	23.1055	2.1751	18.8423	27.3687	112.84	<.0001
centre	15	1	22.9148	2.3279	18.3521	27.4775	96.89	<.0001
centre	18	1	−0.0486	126467.5	−247872	247871.8	0.00	1.0000
centre	23	1	24.3186	2.7168	18.9938	29.6434	80.12	<.0001
centre	24	1	1.2139	322114.2	−631331	631333.5	0.00	1.0000
centre	25	1	−0.0211	117643.3	−230577	230576.5	0.00	1.0000
centre	26	1	22.9963	2.3328	18.4242	27.5685	97.18	<.0001
centre	27	1	25.5642	2.6333	20.4029	30.7254	94.24	<.0001
centre	29	1	24.7202	2.9282	18.9811	30.4594	71.27	<.0001
centre	30	1	−0.0857	137672.4	−269833	269832.9	0.00	1.0000
centre	31	1	−0.0486	51630.15	−101193	101193.2	0.00	1.0000
centre	32	1	23.3668	2.4567	18.5518	28.1817	90.47	<.0001
centre	35	1	−0.2451	183098.8	−358867	358866.7	0.00	1.0000
centre	36	1	23.2206	2.1447	19.0170	27.4241	117.22	<.0001
centre	37	1	23.5944	2.3797	18.9302	28.2585	98.30	<.0001
centre	40	0	23.7051	0.0000	23.7051	23.7051	.30	.0001
centre	41	0	0.0000	0.0000	0.0000	0.0000	.30	.0001

Centre 40 has no DF. This has occurred because centre 41 (the usual reference category) is uniform, and so centre 40 is used for reference. Standard errors for the other centre effects are based on comparisons with centre 40. However, their estimates in this output are still based on comparisons to centre 41. Thus, SAS does not produce useful centre estimates and standard errors when the reference category is uniform. Usually, centre estimates will not be of interest. However, if required, they can be obtained by renumbering the centres so that the last one is non-uniform. Although we are clearly getting estimates and standard errors that are unstable, the likelihood still converges, since the uniform categories have little effect on it.

Although Models 2 and 3 are not recommended for estimating treatment effects, they can still be used to test the overall significance of the fixed centre and centre·treatment effects by using likelihood ratio tests. For example, to test centre effects in Model 2, we calculate twice the difference in the log likelihoods between Models 1 and 2, $178.17 - 147.78 = 30.39 \sim \chi^2_{28}$. This indicates that centre effects are non-significant. To test centre·treatment effects, twice the difference in the log likelihoods between Models 2 and 3 is taken, $147.78 - 100.05 = 47.73 \sim \chi^2_{47}$. This is also non-significant. However, it should be borne in mind that these tests have low power for detecting small centre or centre·treatment effects.

A preferable approach to fitting Model 2 may be to use an exact conditional logistic regression stratified by centre. This would avoid the loss of information on uniform centres. Or alternatively, the smaller centres could be combined for the purposes of analysis as 'other centres'. However, the results from Models 2 and 3 indicate that centre effects are not important, and therefore Model 1 is likely to be the most satisfactory fixed effects model.

Note that we can also examine whether the treatment effects are consistent in those with and without cold feet at baseline. Adding a baseline by treatment interaction to the models showed no statistically significant effect in any of the models presented.

Random effects models fitted using pseudo-likelihood (4 and 5)

In Model 4, there is a small positive centre variance component, indicating that some recovery of treatment information from between the centres has occurred. However, it is not possible to assess the extent of this, since there is no satisfactory equivalent fixed effects model (Model 2 has uniform centre effects). The treatment effect standard errors are very similar to those obtained in Model 1, indicating that little appears to have been gained by fitting centre effects as random in this example.

In Model 5, the centre·treatment variance component is positive, indicating that the treatment effects vary randomly between centres. This is reflected in their standard errors that are increased over those for Model 4 to allow for the additional variation occurring between centres. Since treatment effects are assumed to vary randomly between the centres, results can be related with more confidence to the population of centres. The centre variance component is zero, and so there is no overall variability in the incidence of cold feet between centres.

Bayesian models (6 and 7)

The variance component estimates in Models 6 and 7, fitted using Bayesian methods, are on the whole smaller than those obtained by pseudo-likelihood (Models 4 and 5). However, they are taken as medians of their posterior distributions and are also based on using a prior that is constrained to be positive (pseudo-likelihood is equivalent to using flat priors). For these reasons,

they cannot be compared directly. The treatment standard errors in Model 7 are noticeably increased over those in Model 6, despite the fact that the centre·treatment variance component median is small. It is difficult to decide whether the pseudo-likelihood or Bayesian models are preferable for this example. On the whole, however, the results are very similar.

3.4.3 Discussion of points from Section 3.3

We will now discuss the points from Section 3.3 that have not already been covered.

Negative variance components (Section 3.3.3)

A negative variance component estimate would have been obtained for centre effects in Model 5 had it not been constrained at zero. This would almost certainly have been an underestimate of a zero or small positive variance component. Identical fixed effects estimates and standard errors would have been obtained if the data had been reanalysed with centre effects omitted; however, the denominator DF used for F tests will be smaller when centre effects are retained. The problem does not arise in Models 6 and 7, since negative variance component samples cannot be obtained when an inverse gamma distribution is assumed for their prior.

Accuracy of variance parameters and random effects shrinkage (Section 3.3.5)

In this trial, there were uniform centre categories, and this may lead to some bias in the variance parameter estimates. However, fitting a dispersion parameter will, to some extent, help to alleviate any bias.

Bias in fixed and random effects standard errors (Section 3.3.6)

In this example (Models 4 and 5), there is a possibility that some bias in variance components (and hence fixed effects standard errors) has occurred, although to some extent, this may have been overcome by fitting a dispersion parameter. In addition, some downward bias may occur because information is combined across several error strata. However, since there are 29 centres, any bias occurring for this reason is likely to be small and has been corrected for by using the Kenward–Roger adjustment in `PROC GLIMMIX`.

The dispersion parameter (Section 3.3.7)

The dispersion parameter was fixed at one in Models 1–3. However, its estimated value is also given in brackets in Table 3.4. In Model 1, it is very close to one, and

fixing it at one has made very little difference to the fixed effects standard errors. In Models 2 and 3, the dispersion parameter is well below one. This is largely due to the influence of the uniform categories, and thus the dispersion parameter estimate can be considered as another indicator of their presence.

The dispersion parameters in Model 4 and 5 are both notably below one due to the presence of uniform centre effect categories. These parameters help to overcome the discrepancy in the mean/variance relationship caused by the random effects estimates being shrunken compared with their raw means. If dispersion parameters were omitted, it is likely that a downward bias in the centre and centre·treatment variance components would have occurred. However, it is difficult to tell how adequately the dispersion parameter has overcome this potential problem. Our own view is that the results are likely to be satisfactory. We can draw some comfort from the fact that the results from Model 4 are similar to those from Model 1, which does not suffer from the problems associated with uniform effects categories.

Significance testing (Section 3.3.8)

Significance tests are illustrated for Model 5. Tests of treatment effects were carried out using Wald F tests with Satterthwaite's approximation to the denominator DF (by using the `DDFM=KR` option). An $F_{2,67}$ value of 3.68 was obtained for the composite test of treatment equality. This gave a significant p-value of 0.03. Wald t tests were used to perform pairwise treatment comparisons:

$$\begin{aligned} \text{A} - \text{B} \quad & t_{55} = -1.04, \quad p = 0.30, \\ \text{A} - \text{C} \quad & t_{82} = 1.99, \quad p = 0.08, \\ \text{B} - \text{C} \quad & t_{75} = 2.71, \quad p = 0.008. \end{aligned}$$

Thus, cold feet are significantly less likely on treatment C than on treatment B. Treatment A is intermediate in its effect. The t statistic for baseline cold feet was 6.57 with 272 DF. This was highly significant ($p < 0.0001$), indicating that fitting baseline has greatly increased the sensitivity of the analysis.

Confidence intervals (Section 3.3.9)

The 95% confidence intervals for treatment effects were calculated from the mean treatment differences and standard errors. The confidence interval on a linear scale for A − B is

$$95\% \text{ CI} = -0.595 \pm t_{55,0.975} \times 0.571, \quad t_{55,0.975} = 2.01$$

so we obtain

$$95\% \text{ CI} = -0.595 \pm 2.01 \times 0.571 = (-1.738, 0.549).$$

A comparison of treatments A and B in terms of an OR is obtained by exponentiating the effect estimate:

$$\text{OR} = \frac{P(\text{cold feet on A})/(1 - P(\text{cold feet on A}))}{P(\text{cold feet on B})/(1 - P(\text{cold feet on B}))} = \exp(-0.595) = 0.55.$$

Confidence intervals for the OR are calculated by exponentiating the confidence intervals calculated on the linear scale:

$$\exp(-1.738, 0.549) = (0.18, 1.72).$$

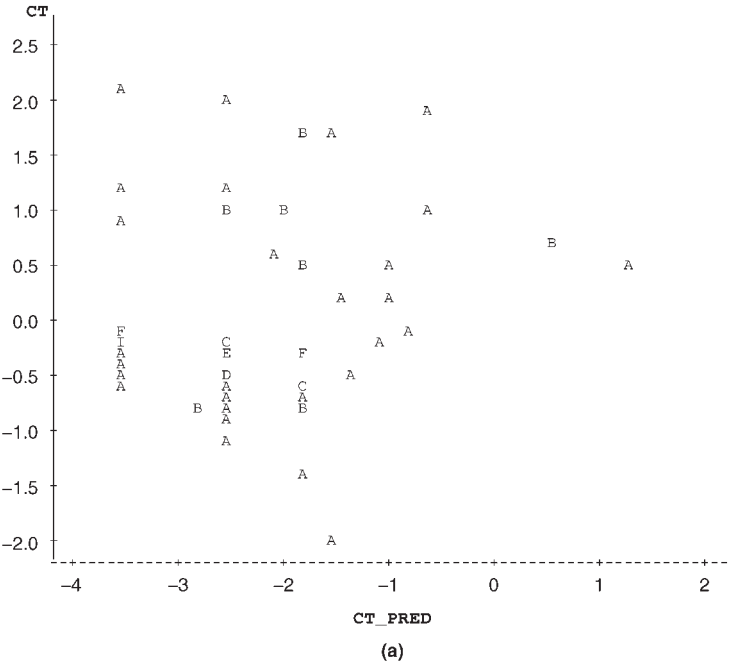
Confidence intervals were calculated in the same way for the other treatment effects:

Effect	Linear scale	Odds ratio
A – B	−0.595 (−1.732, 0.543)	0.55 (0.18, 1.73)
A – C	1.183 (−0.135, 2.501)	3.26 (0.87, 12.19)
B – C	1.777 (−0.472, 3.082)	5.91 (1.60, 21.82)

In `PROC GLIMMIX`, the `LSMEANS` statement with the `OR` option or, alternatively, the `OR` option in the `PROC GLIMMIX` statement can be used to obtain the ORs and their 95% confidence intervals directly. The use of the `ILINK` option in the `LSMEANS` statement will also give the estimated binomial probabilities for each treatment. By default, these will be evaluated at the overall average baseline value for cold feet, which, as a non-integer value is meaningless. Additional use of the `AT` option allows this to be estimated for those with and without cold feet at baseline (see `SAS` code at the end of this section).

Checking model assumptions (Section 3.3.10)

Plots of the centre·treatment effects against their predicted values and normal plots are used to check Model 5 (Figure 3.1). Note that the centre effects do not need to be checked, since the centre variance component estimate was zero. As the data are in Bernoulli form, plots of the residuals are not useful for identifying outliers. The centre·treatment effect plots show no strong evidence of outlying treatment effects at a centre. Although the normal plot indicates some deviation from normality, this is unlikely to influence the estimates of fixed effects and their standard errors (see Section 2.4.6).



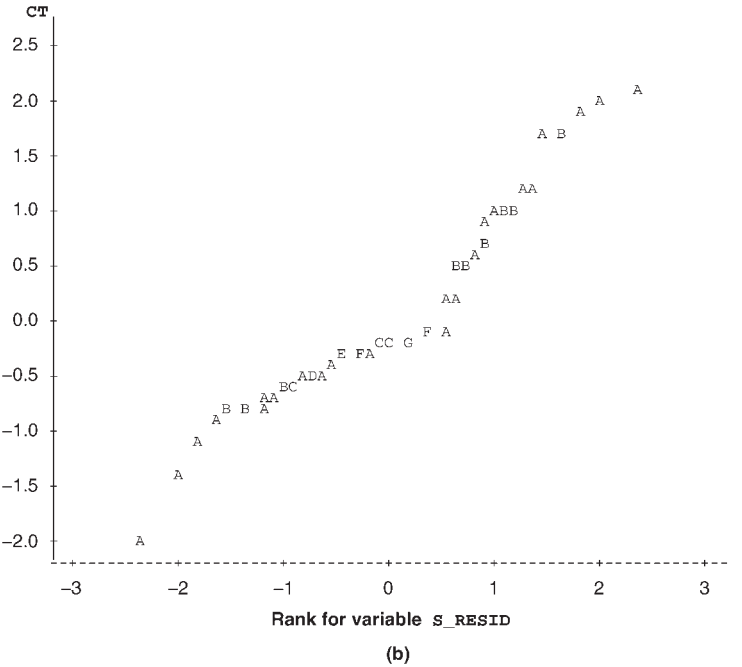


Figure 3.1 Plots of centre·treatment effects. (a) Centre·treatment effects against their predicted values. (b) Centre·treatment effects normal plot. A = 1 obs., B = 2 obs., etc.

Determining whether the simulated posterior distribution has converged (3.3.11)

The following three tables are produced by `PROC MCMC` and provide statistics to help assess whether the model has converged. The Gewerke test shows a significant result for the baseline cold feet parameter and the centre·treatment variance component. This would warrant a further analysis, either with a large sample size or by comparing results from models using different seeds for the random sampling process.

Monte Carlo Standard Errors

Parameter	MCSE	Standard Deviation	MCSE/SD
alpha0	0.0202	0.6899	0.0293
alpha1	0.0122	0.6075	0.0201
alpha2	0.00887	0.7422	0.0120
alpha3	0.00773	0.7181	0.0108
v2	0.00716	0.3220	0.0222
v3	0.0517	0.9129	0.0567

Geweke Diagnostics

Parameter	z	Pr > |z|
alpha0	-2.5359	0.0112
alpha1	3.4395	0.0006
alpha2	1.4716	0.1411
alpha3	1.7568	0.0789
v2	-0.2956	0.7675
v3	3.1427	0.0017

Effective Sample Sizes

Parameter	ESS	Autocorrelation Time	Efficiency
alpha0	1168.7	42.7829	0.0234
alpha1	2479.7	20.1637	0.0496
alpha2	6999.5	7.1434	0.1400
alpha3	8636.3	5.7895	0.1727
v2	2020.5	24.7460	0.0404
v3	311.4	160.6	0.0062

Diagnostic plots (Figure 3.2) are shown only for the centre·treatment variance component (`v3`) for which there was more doubt over convergence. Plots for the other parameters were satisfactory. The first two plots show a high degree of

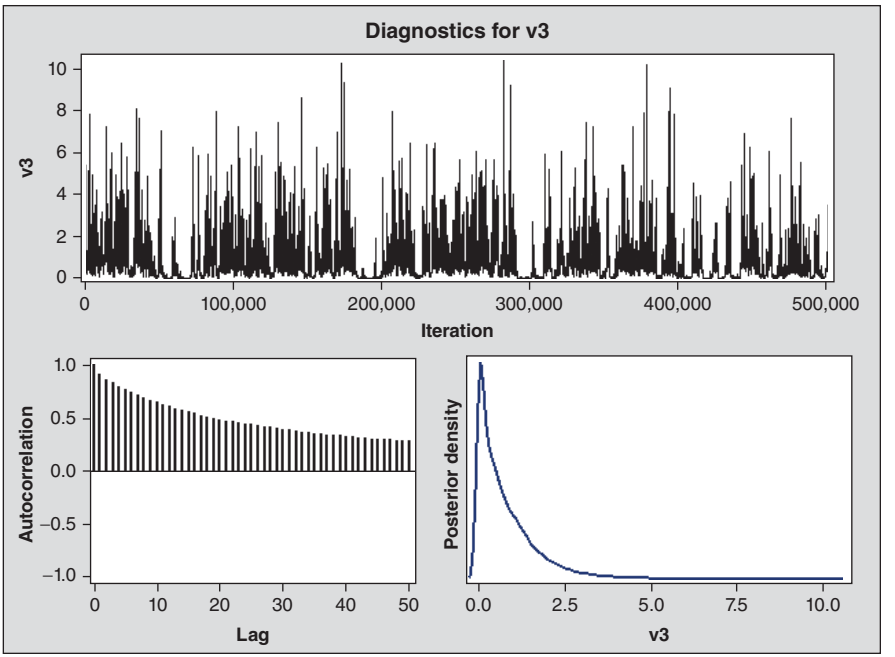


Figure 3.2 Diagnostic plots for the centre·treatment variance component (v3) parameter.

autocorrelation between the samples for `v3` and confirm that fitting a model with more samples would be advisable.

As the diagnostic statistics and graphs indicated the posterior distribution for the centre·treatment variance component may not have converged, some alternative models with more samples and different thinning levels will be considered. Firstly, a model was fitted with more samples (5,000,000) and with the thinning factor kept at 10 (Model 7b). To help consider whether the thinning level is important, we also consider the same model with thinning factors of 100 (Model 7c) and 5 (Models 7d) and no thinning (Model 7e). Lastly, to assess whether a sample size increase is required, a model with 500,000 samples and no thinning is considered (Model 7f).

The following tables compare the estimates and diagnostic statistics between the alternative models for the first treatment effect, A − C (Table 3.6) and for the centre·treatment variance component that gave particular cause for concern in Model 7a (Table 3.7). SAS reported space problems when fitting Model 7e, indicating that this size of sample could not be stored for this model.

Estimates of the treatment difference, A − C, are almost identical between the models, indicating that any of the models were likely to be adequate for estimating treatment effects and their standard errors. The diagnostic statistics do not give any cause for concern. It is interesting that the diagnostic statistics, in general,

Table 3.6 Comparison of estimates and diagnostic statistics for treatment difference A – C.

Model	Thinning level	Sample size	A – C, mean (SE)	% Variation due to sampling (MCSE/SD)	Geweke test *p*-value	Effective sample size
7a	10	500,000	1.18 (0.74)	0.012	0.14	6999
7b	10	5,000,000	1.18 (0.74)	0.005	0.95	41278
7c	100	5,000,000	1.17 (0.74)	0.006	0.86	32957
7d	5	5,000,000	1.18 (0.74)	0.005	0.89	44403
7f	1	500,000	1.18 (0.74)	0.011	0.09	8772

Table 3.7 Comparison of estimates and diagnostic statistics for the centre·treatment variance component.

Model	Thinning level	Sample size	Centre·treatment variance component, median	% Variation due to sampling (MCSE/SD)	Geweke test *p*-value	Effective sample size
7a	10	500,000	0.53	0.057	0.002	311
7b	10	5,000,000	0.49	0.019	0.97	2892
7c	100	5,000,000	0.49	0.020	0.98	2429
7d	5	5,000,000	0.49	0.016	0.96	3895
7f	1	500,000	0.53	0.031	0.002	1034

improve when no thinning is used. Thus, for this fixed effect, thinning has not provided any advantage except to provide a more manageable sample size.

The centre·treatment variance component estimate (Table 3.7) is identical between Model 7b and 7d but differs for Models 7a and 7f where only 500,000 samples were taken. Also, the Gewerke test is significant for these models. Thus 500,000 samples has not been sufficient to adequately obtain the posterior distribution to provide the centre·treatment variance component estimate. However, it is interesting that the inadequate sampling of this variance component did not prevent adequate estimates of treatment effects being obtained from Models 7a and 7f. Thus, if the main interest is in obtaining estimates of treatment effects, it would appear sufficient to check only diagnostics statistics and plots for these effects. The usual skewed distribution of the variance components makes the diagnostic statistics and plots a little less easy to assess, since they are better suited to checking normally distributed parameters.

SAS code and output

Variables
centre = centre
treat = treatment
cf = post-treatment cold feet (1 = yes, 0 = no),
cf1 = pre-treatment cold feet (1 = yes, 0 = no),
one = 1, all observations.

SAS code is provided for Models 1 – 5. However, detailed output is only given for Models 1 and 5, which should be sufficient to illustrate the use of PROC GENMOD and PROC GLIMMIX.

Model 1

```
PROC GENMOD; CLASS centre treat;
MODEL cf/one=cf1 treat/ DIST=B COVB TYPE3 WALD;
ESTIMATE 'A-B' treat 1 -1 0/ E ALPHA=0.05 EXP;
ESTIMATE 'A-C' treat 1 0 -1/ E ALPHA=0.05 EXP;
ESTIMATE 'B-C' treat 0 1 -1/ E ALPHA=0.05 EXP;
```

```
                    Model Information
            Data Set                              WORK.B
            Distribution                          Binomial
            Link Function                         Logit
            Response Variable (Events)            cf
            Response Variable (Trials)            one
            Number of Observations Read           283
            Number of Observations Used           283
            Number of Events                      41
            Number of Trials                      283
```

```
                  Class Level Information
Class        Levels      Values
centre           29      1 2 3 4 5 6 7 8 9 11 12 13 14 15 18 23 24 25
                         26 27 29 30 31 32 35 36 37 40 41
treat             3      A B C
```

Parameter Information

Parameter	Effect	treat
Prm1	Intercept	
Prm2	cf1	
Prm3	treat	A
Prm4	treat	B
Prm5	treat	C

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	279	178.1744	0.6386
Scaled Deviance	279	178.1744	0.6386
Pearson Chi-Square	279	282.4176	1.0122
Scaled Pearson X2	279	282.4176	1.0122
Log Likelihood		−89.0872	

Algorithm converged.

The deviance and Pearson chi-square are measures of model fit and have similar roles to the residual sum of squares in normal data models. The Pearson chi-square is the sum of squared Pearson residuals and the deviance is calculated as $2\log(L_y/L_m)$. L_y is the ML achievable if all available DF were used (usually this is obtained when $u_i = y_i$) and L_m is the likelihood for the model fitted.

Estimated Covariance Matrix

	Prm1	Prm2	Prm3	Prm4
Prm1	0.26581	−0.06973	−0.24923	−0.25407
Prm2	−0.06973	0.23598	0.01363	0.03002
Prm3	−0.24923	0.01363	0.35992	0.24694
Prm4	−0.25407	0.03002	0.24694	0.32686

The COVB option has caused the covariance matrix for the fixed effects (intercept, cf1, treat A and treat B) to be printed.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence	Limits	Chi-Square	Pr > ChiSq
Intercept		1	−3.3532	0.5156	−4.3637	−2.3427	42.30	<.0001
cf1		1	2.9697	0.4858	2.0176	3.9218	37.37	<.0001
treat	A	1	0.9361	0.5999	−0.2397	2.1120	2.43	0.1187
treat	B	1	1.7043	0.5717	0.5837	2.8248	8.89	0.0029
treat	C	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Wald Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
cf1	1	37.37	<.0001
treat	2	9.60	0.0082

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence	Limits	Chi-Square	Pr > ChiSq
A-B	−0.7681	0.4392	0.05	−1.6290	0.0927	3.06	0.0803
Exp (A-B)	0.4639	0.2037	0.05	0.1961	1.0971		
A-C	0.9361	0.5999	0.05	−0.2397	2.1120	2.43	0.1187
Exp (A-C)	2.5501	1.5299	0.05	0.7868	8.2645		
B-C	1.7043	0.5717	0.05	0.5837	2.8248	8.89	0.0029
Exp (B-C)	5.4973	3.1429	0.05	1.7927	16.8576		

Asymptotic chi-squared tests are performed for each fixed effects parameter. These tests should be interpreted cautiously in small datasets.

Model 2

```
PROC GENMOD; CLASS centre treat;
MODEL cf/one=cf1 treat centre/ DIST=B;
ESTIMATE 'A-B' treat 1 -1 0/ E ALPHA=0.05 EXP;
ESTIMATE 'A-C' treat 1 0 -1/ E ALPHA=0.05 EXP;
ESTIMATE 'B-C' treat 0 1 -1/ E ALPHA=0.05 EXP;
```

Although the denominator term 'one' is not needed in the `MODEL` statement, its inclusion causes `SAS` to use 0 rather than 1 as the reference category. The early parts of the output have a similar form to that given for Model 1. The following output indicates how uniform fixed effects categories can be identified.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	251	147.7812	0.5888
Scaled Deviance	251	147.7812	0.5888
Pearson Chi-Square	251	199.3547	0.7942
Scaled Pearson X2	251	199.3547	0.7942
Log Likelihood		−73.8906	

```
WARNING: Negative of Hessian not positive definite.
```

A warning of a non-positive Hessian matrix often occurs when there are uniform fixed effects categories, even though the model has fitted successfully for data

corresponding to the other (non-uniform) categories. This causes the ESTIMATE statements to fail, even though the treatment effects are in fact estimable.

Model 3

```
PROC GENMOD; CLASS centre treat;
MODEL cf/one=cf1 treat centre centre*treat/ DIST=B;
ESTIMATE 'A-B' treat 1 -1 0/ E ALPHA=0.05 EXP;
ESTIMATE 'A-C' treat 1 0 -1/ E ALPHA=0.05 EXP;
ESTIMATE 'B-C' treat 0 1 -1/ E ALPHA=0.05 EXP;
```

Output for Model 3 has a similar form to Model 2, except that the treatment effects are non-estimable, as each treatment is not received at every centre.

Model 4

```
PROC GLIMMIX OR; CLASS centre treat;
MODEL cf=cf1 treat/ DIST=B DDFM=KR;
RANDOM _RESIDUAL_;
RANDOM centre;
LSMEANS treat/ DIFF PDIFF CL;
```

Note that a denominator term is not now included, and 0 is appropriately used as the reference category for comparing fixed effects. This contrasts with `PROC GENMOD` where the highest category is taken as the reference category when the denominator term is omitted. However, note that if `DIST=BINARY` had been used in place of `DIST=B` (or equivalently `DIST=BINOMIAL`) in `PROC GLIMMIX`, SAS would then have taken the highest category, 1, to be the reference category.

Model 5

```
PROC GLIMMIX OR; CLASS centre treat;
MODEL cf=cf1 treat/ DIST=B DDFM=KR;
RANDOM _RESIDUAL_;
RANDOM centre centre*treat;
LSMEANS treat/ DIFF PDIFF CL;
LSMEANS treat/ AT cf1=0 ILINK PLOTS=MEANPLOT(ILINK)CL;
LSMEANS treat/ AT cf1=1 ILINK PLOTS=MEANPLOT(ILINK)CL;
```

Note that these final two `LSMEANS` statements produce estimated binomial probabilities for each treatment group in those with and without cold feet at baseline.

The GLIMMIX Procedure

Model Information

Data Set	WORK.B
Response Variable	cf
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Kenward-Roger
Fixed Effects SE Adjustment	Kenward-Roger

Class Level Information

Class	Levels	Values
centre	29	1 2 3 4 5 6 7 8 9 11 12 13 14 15 18 23 24 25 26 27 29 30 31 32 35 36 37 40 41
treat	3	A B C

Number of Observations Read	288
Number of Observations Used	283

Dimensions

G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	5
Columns in Z	108
Subjects (Blocks in V)	1
Max Obs per Subject	283

Optimization Information

Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	7	1286.313291	0.89690379	2.097988
1	0	5	1398.3080594	0.49367227	0.737813
2	0	5	1457.1805377	0.23227199	0.400017
3	0	5	1478.5781437	0.10235922	0.314668
4	0	3	1485.5322867	0.04087928	0.288165
5	0	2	1488.0039207	0.01553835	0.278724
6	0	2	1488.9096469	0.00580600	0.275252
7	0	1	1489.2434824	0.00215911	0.273976
8	0	1	1489.3669865	0.00079938	0.273499
9	0	1	1489.4126248	0.00029606	0.273323
10	0	1	1489.4295153	0.00010966	0.273258
11	0	1	1489.4357697	0.00004062	0.273234
12	0	1	1489.4380862	0.00001504	0.273225
13	0	1	1489.4389441	0.00001099	0.273239
14	0	0	1489.439571	0.00000000	0.273224

Convergence criterion (PCONV=1.11022E-8) satisfied.

Estimated G matrix is not positive definite.

Fit Statistics

-2 Res Log Pseudo-Likelihood	1489.44
Generalized Chi-Square	151.80
Gener. Chi-Square / DF	0.54

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
centre	0	.
centre*treat	1.8777	0.6834
Residual (VC)	0.5441	0.05113

Odds Ratio Estimates

treat	cf1	_treat	_cf1	Estimate	DF	95% Confidence	Limits
	1.0947		0.0947	21.293	272.4	8.519	53.218
A	0.0947	C	0.0947	3.264	81.55	0.873	12.194
B	0.0947	C	0.0947	5.914	74.57	1.603	21.822

Effects of continuous variables are assessed as one unit offsets from the mean. The `AT` sub-option modifies the reference value and the `UNIT` sub-option modifies the offsets.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
cf1	1	272.4	43.20	<.0001
treat	2	67.41	3.68	0.0303

treat Least Squares Means

treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
A	-2.1145	0.4117	63.25	-5.14	<.0001	0.05	-2.9371	-1.2918
B	-1.5199	0.3967	49.21	-3.83	0.0004	0.05	-2.3171	-0.7227
C	-3.2973	0.5241	100.1	-6.29	<.0001	0.05	-4.3370	-2.2576

Differences of treat Least Squares Means

treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
A	B	-0.5946	0.5707	55.34	-1.04	0.3020	0.05
A	C	1.1828	0.6626	81.55	1.79	0.0780	0.05
B	C	1.7774	0.6553	74.57	2.71	0.0083	0.05

Differences of treat Least Squares Means

treat	_treat	Lower	Upper	Odds Ratio	Lower Odds Ratio	Upper Odds Ratio
A	B	-1.7381	0.5489	0.552	0.176	1.731
A	C	-0.1354	2.5010	3.264	0.873	12.194
B	C	0.4719	3.0829	5.914	1.603	21.822

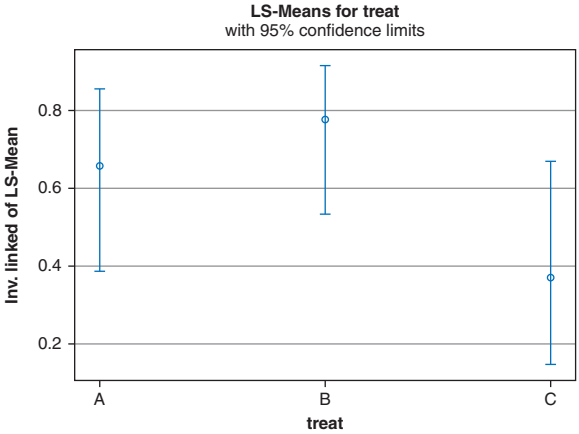
The following section shows the consequence of the use of the `AT` options in the `LSMEANS` statements.

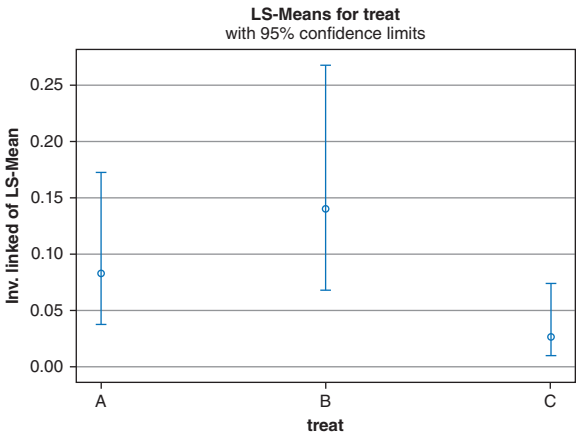
treat Least Squares Means

treat	cf1	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Mean	Standard Error Mean	Lower Mean	Upper Mean
A	1.00	0.6521	0.5637	127.5	1.16	0.2495	0.05	-0.4633	1.7675	0.6575	0.1269	0.3862	0.8542
B	1.00	1.2467	0.5654	114.6	2.20	0.0295	0.05	0.1267	2.3667	0.7767	0.09806	0.5316	0.9143
C	1.00	-0.5307	0.6233	132.6	-0.85	0.3961	0.05	-1.7637	0.7023	0.3704	0.1454	0.1463	0.6687

treat Least Squares Means

treat	cf1	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Mean	Standard Error Mean	Lower Mean	Upper Mean
A	0.00	-2.4063	0.4177	66.2	-5.76	<.0001	0.05	-3.2403	-1.5723	0.08270	0.03169	0.03768	0.1719
B	0.00	-1.8117	0.4012	51.35	-4.52	<.0001	0.05	-2.6169	-1.0064	0.1404	0.04843	0.06806	0.2677
C	0.00	-3.5891	0.5323	104.5	-6.74	<.0001	0.05	-4.6445	-2.5337	0.02688	0.01392	0.009523	0.07353





Model checking for Model 5

```
PROC GLIMMIX; CLASS centre treat;
MODEL cf/one=cf1 treat/ DIST=B DDFM=KR;
RANDOM _RESIDUAL_;
RANDOM centre centre*treat/ SOLUTION;
LSMEANS treat/ DIFF PDIFF OR CL;
OUTPUT OUT=pred pred=pred;
ODS OUTPUT SOLUTIONR=random;

DATA pred2; SET pred;
KEEP centre treat pred;
PROC SORT; BY centre treat;
```

```
PROC MEANS NOPRINT; BY centre treat; VAR pred; OUTPUT
OUT=pred3 MEAN=re_pred N=freq;

DATA random2; SET random;
IF effect='centre*treat';
re=estimate;
KEEP centre treat re;

DATA check; MERGE pred3 random2; BY centre treat;
PROC PLOT; PLOT re*re_pred;
TITLE' CENTRE.TREATMENT RANDOM EFFECTS AGAINST THEIR
   PREDICTED VALUES';
PROC RANK OUT=norm NORMAL=tukey; VAR re; RANKS rank;
PROC PLOT DATA=norm; PLOT re*rank;
TITLE 'CENTRE.TREATMENT RESIDUALS - NORMAL PLOT';
```

The PROC GLIMMIX code fits Model 5 as before with an OUTPUT statement added to output the predictions (based on $\mathbf{X}\hat{\boldsymbol{\alpha}}$, i.e. omitted the random effects element) to dataset 'pred' and a SOLUTION option in the RANDOM statement along with an ODS statement to output the random effect estimates to dataset 'random'. The mean predictions are then calculated for each centre·treatment category, and the centre·treatment random effects are plotted against their predicted means. Normal plots of the random effects are also produced as a further check.

Plots are given in the main text. Note that the plotting options available with PROC GLIMMIX do not allow these checks to be made directly. It offers a variety of residual plots, which are not helpful with binary data, plots of LSMEANS, as illustrated earlier and differences of LSMEANS.

Model 6

```
ODS GRAPHICS ON;
PROC MCMC OUTPOST=post6 NMC=100000 THIN=10 SEED=7899;
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
   POSTINTERVALS MCSE GEWEKE ESS TADPANEL;
PARMS alpha0 alpha1 alpha2 alpha3 v2;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_centre ~ NORMAL(0, VAR = v2) SUBJECT=centre;
mu = alpha0 + alpha1*cf1 + alpha2*treata + alpha3*treatb
     + b_centre;
expected = LOGISTIC(mu);
MODEL CF ~ BINARY(expected);
ODS GRAPHICS OFF;
```

The code is very similar to that used for Model 7, and the descriptions of each statement that we will give for Model 7 can also be applied to Model 6.

Model 7

```
ODS GRAPHICS ON;
PROC MCMC OUTPOST=post7 NMC=500000 THIN=10 SEED=7899;
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
   POSTINTERVALS MCSE GEWEKE ESS TADPANEL;
PARMS alpha0 alpha1 alpha2 alpha3 v2 v3;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_centre ~ NORMAL(0, VAR = v2) SUBJECT=centre;
RANDOM b_ct ~ NORMAL(0, VAR = v3) SUBJECT=centre_treat;
mu = alpha0 + alpha1*cf1 + alpha2*treata + alpha3*treatb
     + b_centre + b_ct;
expected = LOGISTIC(mu);
MODEL cf ~ BINARY(expected);
ODS GRAPHICS OFF;
```

In this analysis, 500,000 samples were taken, and a thinning factor of 10 was used. The model parameters are defined first using a `PARMS` statement. `alpha0` represents the intercept, `alpha1` the baseline value of cold feet, `alpha2` and `alpha3` the two treatment parameters and `v2` the centre variance component.

The first `PRIOR` statement specifies 'non-informative' prior normal distributions with zero means and very large variances of 10,000 for the fixed effect parameters (`alpha0`–`alpha3`). The second `PRIOR` statement specifies 'non-informative' inverse gamma distributions with very small parameters for prior distribution of the centre and centre·treatment variance components (`v2` and `v3`).

The `RANDOM` statements specify a normal distribution for the random effects (b_centre and b_ct) with zero mean and variance equal to the corresponding variance components. The `SUBJECT` options state which level the random effect will vary across.

The next three statements define the model. The first statement (starting 'mu = ') specifies the linear part of the model and calculates it as 'mu'. The second uses the logistic link function to obtain the expected means ('expected') for the binary distribution from the linear component of the model ('mu'). The third specifies that the data has a binary distribution with mean given by the variable 'expected'.

The `ODS SELECT` statement requests the statistics to be output. Summaries of the parameters are requested by the `PARAMETERS`, `REPARAMETERS`, `POSTSUMMARIES` and `POSTINTERVALS` options, and diagnostic tests and graphs are requested by the `MCSE`, `GEWEKE`, `ESS` and `TADPANEL` options. Use of the `ODS GRAPHICS` statement along with the `TADPANEL` option will cause diagnostic plots to be created.

Output

Parameters

Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	v2	Conjugate	0.00990	igamma(0.01, scale = 0.01)
2	v3	Conjugate	0.00990	igamma(0.01, scale = 0.01)
3	alpha0	N-Metropolis	0	normal(0, var = 10000)
	alpha1		0	normal(0, var = 10000)
	alpha2		0	normal(0, var = 10000)
	alpha3		0	normal(0, var = 10000)

Random Effects Parameters

Parameter	Subject	Levels	Prior Distribution
b_centre	centre	29	normal(0, var = v2)
b_ct	centre_ treat	78	normal(0, var = v3)

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
alpha0	50000	-3.8102	0.6899	-4.2253	-3.7456	-3.3261
alpha1	50000	3.2761	0.6075	2.8582	3.2455	3.6587
alpha2	50000	1.1796	0.7422	0.6733	1.1443	1.6477
alpha3	50000	1.9561	0.7181	1.4663	1.9130	2.3987
v2	50000	0.2072	0.3220	0.0316	0.0884	0.2464
v3	50000	0.8122	0.9129	0.1532	0.5272	1.1506

Posterior Intervals

Parameter	Alpha	Equal-Tail Interval		HPD Interval	
alpha0	0.050	-5.3515	-2.6440	-5.2047	-2.5396
alpha1	0.050	2.1710	4.5591	2.1286	4.5122
alpha2	0.050	-0.1749	2.7473	-0.2255	2.6796
alpha3	0.050	0.6574	3.4904	0.5818	3.3928

```
v2          0.050     0.00709      1.0889      0.00190      0.7971
v3          0.050      0.0156      3.2454      0.00277      2.5700
```

Output corresponding to the diagnostic statistics to assess convergence is shown in the main text.

SAS code to obtain estimates of the treatment difference A − B, *p*-values and odds ratios for all treatment comparisons, and equal-tailed probability intervals

```
%MACRO mcmc_stats(dat); * macro to print MCMC statistics;
DATA p1; SET &dat;
a_c=alpha2;
b_c=alpha3;
a_b=alpha2-alpha3;  * calculate samples for treatment
   difference A-B;
* define indicator variables for whether the sampled
   differences are greater than or less than zero;
IF a_b<0 THEN a_b0=1; ELSE a_b0=0;
IF a_c<0 THEN a_c0=1; ELSE a_c0=0;
IF b_c<0 THEN b_c0=1; ELSE b_c0=0;

PROC MEANS NOPRINT DATA=p1; VAR a_b a_c b_c a_b0 a_c0 b_c0;
OUTPUT OUT=p2 SUM=dum dum dum a_b0_n a_c0_n b_c0_n N=samples
   mean=a_b_mean a_c_mean b_c_mean std=a_b_std;

DATA p3; SET p2;
%p_calc(a_b); %p_calc(a_c); %p_calc(b_c);
a_c_or=exp(a_c_mean); b_c_or=exp(b_c_mean);
   a_b_or=exp(a_b_mean);

proc univariate data=p1;
var a_b a_c b_c; output out=ci pctlpts=2.5 97.5
   pctlpre=a_b a_c b_c pctlname=lower upper;

PROC PRINT NOOBS DATA=p3; VAR a_b_mean a_b_std; title
   'Mean and SE for A-B (linear scale)';
PROC PRINT NOOBS DATA=p3; VAR a_b_p a_c_p b_c_p; title
   'p-values for pairwise treatment comparisons';
PROC PRINT NOOBS DATA=ci; title 'Equal tail 95%
   CIs (linear scale)';
PROC PRINT NOOBS DATA=p3; VAR a_b_or a_c_or b_c_or; title
   'Odds ratios for treatment comparisons';

data ci2; set ci;
%exp(a_blower); %exp(a_bupper); %exp(a_clower);
%exp(a_cupper); %exp(b_clower); %exp(b_cupper);
```

```
PROC PRINT NOOBS DATA=ci2; title 'Equal tail 95%
   CIs as odds ratios';
RUN;
%MEND;

%MACRO p_calc(var); * macro to obtain p-values;
  &var.0_p=&var.0_n/samples;
  IF &var.0_p<0.5 THEN &var._p=&var.0_p*2; ELSE
     &var._p=(1-&var.0_p)*2;
%MEND;

%MACRO exp(var);
&var=exp(&var);
%MEND;

%mcmc_stats(post7);
```

```
                    Mean and SE for A-B
                       a_b_mean    a_b_std
                       -0.77650    0.56891

            p-values for pairwise treatment comparisons
                        a_b_p      a_c_p     b_c_p
                      0.16208    0.0912    .00288

                           Equal tail 95% CIs
a_blower    a_bupper    a_clower    a_cupper    b_clower    b_cupper
-1.91540     0.34591    -0.17490     2.74733     0.65740     3.49042

                Odds ratios for treatment comparisons
                      a_b_or     a_c_or     b_c_or
                     0.46001    3.25317    7.07192

                  Equal tail 95% CIs as odds ratios
a_blower    a_bupper    a_clower    a_cupper    b_clower    b_cupper
 0.14728     1.41327     0.83954     15.6009     1.92978     32.7998
```

SAS code for models with different sample sizes and thinning factors

The PROC MCMC statements used for each model are shown. The other statements used to fit the model are unchanged from Model 7a.

```
Model 7b: PROC MCMC OUTPOST=post7b NMC=5000000 THIN=10  SEED=7899;
Model 7c: PROC MCMC OUTPOST=post7c NMC=5000000 THIN=100 SEED=7899;
Model 7d: PROC MCMC OUTPOST=post7d NMC=5000000 THIN=5   SEED=7899;
Model 7e: PROC MCMC OUTPOST=post7e NMC=5000000 THIN=1   SEED=7899;
Model 7f: PROC MCMC OUTPOST=post7f NMC=500000  THIN=1   SEED=7899;
```

4

Mixed models for categorical data

Categorical data often occur in clinical trials. For example, adverse events may be classified on an ordinal scale as mild, moderate or severe. In this chapter we will primarily consider the analysis of measurements made on ordered categorical scales; however, we also describe how unordered categorical data can be analysed. A fixed effects method for analysing ordinal data known as 'ordinal logistic regression' was first suggested by McCullagh (1980) and has been widely applied. The mixed categorical model is far less well established. The model that is defined is based on extending ordinal logistic regression to include random effects and covariance patterns. As we suggested in Chapter 3 for GLMMs, some readers with a less statistical background may wish to read only the introductory paragraphs of each section which will enable them to identify where these methods might prove useful. The final section of this chapter and sections of subsequent chapters will illustrate the application of mixed categorical models.

Ordinal logistic regression will be described in Section 4.1. It is extended to a mixed ordinal logistic regression model in Section 4.2. In Section 4.3 we describe how the model can be adapted to analyse unordered categorical data. In Section 4.4 some practical issues related to model fitting and interpretation are considered, and a worked example is given in Section 4.5.

4.1 Ordinal logistic regression (fixed effects model)

Ordinal logistic regression is a fixed effects method for analysing ordinal data. It is often preferable to contingency table methods such as the Chi-squared 'test for trend' because several fixed effects can be included in the model. The method works by:

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

- **Assuming observations have a multinomial distribution which can be expressed:**

$$p(y_i) = \prod_{j=1}^{c} \mu_{ij}^{z_{ij}},$$

where

i = observation number,
j = category number,
c = number of categories,
$z_{ij} = 1$ if $y_i = j$
$= 0$ otherwise,
$\mu_{ij} = p(y_i = j)$.

- **Taking the ordered nature of the data into account by defining a model for the cumulative categorical probabilities.** The cumulative probabilities, $\mu_{ij}^{[c]} = \sum_{k=1}^{j} \mu_{ik}$, correspond to the probability that observation i is in a category less than or equal to j. They can be thought of as probabilities arising from partitioning the categories in every possible place. For example, if the response variable had categories labelled 1, 2, 3 and 4, three partitions would be possible: 1/2–4, 1–2/3–4 and 1–3/4. The cumulative probabilities are linked to the model parameters using the logit link function (see Section 3.1.4):

$$\log(\mu_{ij}^{[c]}/(1 - \mu_{ij}^{[c]})) = I_j + \mathbf{x}_i\boldsymbol{\alpha}, \qquad j = 1, 2, \ldots, c - 1,$$

where

I_j = intercept parameter for each partition j,
$\mu_{ij}^{[c]} = p(y_i \leq j) = \sum_{k=1}^{j} \mu_{ik}$,
$\mathbf{x}_i = i$th row of fixed effects design matrix $\mathbf{X}$,
$\boldsymbol{\alpha}$ = vector of fixed effects parameters.

Note that there is a separate equation for each partition, j.

- **Maximising the likelihood function for the model parameters (the I_j and $\boldsymbol{\alpha}$) based on the multinomial distribution.** Methods for maximising the likelihood function coincide with those for the mixed ordinal logistic regression model and are described in Section 4.2. They are based on using the matrix notation below to represent the multinomial distribution.

Expressing the model in matrix notation

The ordinal logistic model can alternatively be expressed using matrix notation. However, the multinomial distribution is not a member of the exponential family and cannot be linked to the model parameters using a single link function. This hurdle can be overcome by re-expressing the data in binary form. To do this we

allow each observation to become a vector of $c-1$ correlated binary observations (c = number of categories). For example, if there are four categories, then we could let: $y=1$ become (1, 0, 0), $y=2$ become (0, 1, 0), $y=3$ become (0, 0, 1) and $y=4$ become (0, 0, 0). Thus, the three terms correspond to the presence or absence of the first three categories, while the presence of the fourth category is implied by the absence of the first three. The vector, $\mathbf{y}$, containing the n extended observations can then be defined as

$$\mathbf{y} = (y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}, \ \cdots\ , y_{n1}, y_{n2}, y_{n3})'.$$

To illustrate this redefinition consider the following data constituting the first five observations from a repeated measures trial in which $\mathbf{y}$ has range 1–4.

Patient	Visit	Treatment	y
1	1	A	2
1	2	A	1
1	3	A	4
2	1	B	3
2	2	B	1

When $\mathbf{y}$ is expressed in its extended binary form it becomes

$$\mathbf{y} = (0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0)'.$$

The ordinal logistic model can now be specified in a form of a GLMM using the cumulative probabilities that result from partitioning the categories in each possible place. The GLMM uses a covariance pattern to allow for the multinomial correlations occurring between the binary observations. A cumulative probability, $\mu_{ij}^{[c]}$, is defined as the probability that observation i is in a category less than or equal to j:

$$\begin{aligned} &\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}, \\ &\log(\boldsymbol{\mu}^{[c]}/(1-\boldsymbol{\mu}^{[c]})) = \mathbf{X}\boldsymbol{\alpha}, \\ &\mathrm{var}(\mathbf{e}) = \mathbf{R}. \end{aligned}$$

$\boldsymbol{\mu}$ is the vector of expected multinomial probabilities corresponding to the n extended observations. If there are four categories then we may write

$$\boldsymbol{\mu} = (\mu_{11}, \mu_{12,}\mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}, \ \cdots\ , \mu_{n1}, \mu_{n2}, \mu_{n3})',$$

where

$$\mu_{ij} = \text{probability observation } i \text{ is in category } j.$$

$\boldsymbol{\mu}^{[c]}$ is a vector containing the cumulative probabilities obtained by partitioning the four categories in the three possible places:

$$\boldsymbol{\mu}^{[c]} = (\mu_{11}^{[c]}, \mu_{12}^{[c]}, \mu_{13}^{[c]}, \mu_{21}^{[c]}, \mu_{22}^{[c]}, \mu_{23}^{[c]}, \ \dots\ , \mu_{n1}^{[c]}, \mu_{n2}^{[c]}, \mu_{n3}^{[c]})',$$

where

$$\mu_{ij}^{[c]} = \text{probability}(y_i \leq j) = \sum_{k=1}^{j} \mu_{ik}.$$

So we may equivalently write

$$\begin{aligned}\boldsymbol{\mu}^{[c]} = (&\mu_{11}, \mu_{11} + \mu_{12}, \mu_{11} + \mu_{12} + \mu_{13}, \mu_{21}, \mu_{21} + \mu_{22},\\ &\mu_{21} + \mu_{22} + \mu_{23}, \ \dots\ , \mu_{n1}, \mu_{n1} + \mu_{n2}, \mu_{n1} + \mu_{n2} + \mu_{n3})'.\end{aligned}$$

$\boldsymbol{\alpha}$ is again a vector containing the fixed effects. It has the same form as given in Section 2.1 except $c - 1$ intercept terms are included corresponding to each of the $c - 1$ partitions of the data. Thus, if a model fitting two treatments and three visits as fixed were fitted to the earlier example data, we could write

$$\boldsymbol{\alpha} = (I_1, I_2, I_3, T_\text{A}, T_\text{B}, V_1, V_2, V_3)'.$$

$\mathbf{X}$ is again a design matrix for the fixed effects. However, it now has more rows than previously to correspond to the extended number of observations. For our data $\mathbf{X}$ would be

$$\mathbf{X} = \begin{array}{c} \begin{matrix} I_1 & I_2 & I_3 & T_\text{A} & T_\text{B} & V_1 & V_2 & V_3 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Residual matrix

The residual variance matrix, $\mathbf{R}$, needs to take into account the multinomial correlations that occur within the binary vectors used for each observation. From the multinomial distribution it is known that covariances for the observation vectors, $(y_{i1}, y_{i2}, \ldots, y_{i,c-1})'$, are

$$\begin{aligned}\operatorname{cov}(y_{ij}, y_{ik}) &= \mathrm{E}(y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik}) \\ &= \mu_{ij}(1 - \mu_{ij}), && j = k, \\ &= \mathrm{E}(y_{ij}y_{ik}) - \mathrm{E}(y_{ij})\mathrm{E}(y_{ik}) = -\mu_{ij}\mu_{ik}, && j \neq k.\end{aligned}$$

($\mathrm{E}(y_{ij}y_{ik}) = 0$ when $j \neq k$ because either y_{ij} or y_{ik} has to be zero.)

Thus, within-observation covariance matrices, $\mathbf{R}_i$, can be defined for each of the n original observations. If $c = 4$ we can write the covariance terms corresponding to each pair of partitions for observation i as

$$\mathbf{R}_i = \begin{pmatrix} \mu_{i1}\left(1 - \mu_{i1}\right) & -\mu_{i1}\mu_{i2} & -\mu_{i1}\mu_{i3} \\ -\mu_{i1}\mu_{i2} & \mu_{i2}(1 - \mu_{i2}) & -\mu_{i2}\mu_{i3} \\ -\mu_{i1}\mu_{i3} & -\mu_{i2}\mu_{i3} & \mu_{i3}(1 - \mu_{i3}) \end{pmatrix}.$$

The $\mathbf{R}_i$ matrices form blocks along the diagonal of the full residual matrix, $\mathbf{R}$. For example, in this fixed effects model $\mathbf{R}$ is

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_5 \end{pmatrix},$$

where

$$\mathbf{0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

As in the GLMM definition (Section 3.2), $\mathbf{R}$ can be arranged as a product of a correlation matrix, $\mathbf{P}$, and the matrix of expected Bernoulli variances, $\mathbf{B} = \operatorname{diag}\{\mu_{ij}(1 - \mu_{ij})\}$:

$$\mathbf{R} = \mathbf{B}^{1/2}\mathbf{P}\mathbf{B}^{1/2}.$$

Note that the $\mathbf{A}$ matrix of constants used for the GLMM is now not required, since the data are in Bernoulli form and $\mathbf{A}=\mathbf{I}$. For our example data $\mathbf{P}$ may be written as

$$\mathbf{P}=\begin{pmatrix}\mathbf{P}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{33} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{44} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{55}\end{pmatrix},$$

where

$$\begin{aligned}\mathbf{P}_{ii} &= \text{matrix blocks of 'within-observation' correlations}\\ &=\begin{pmatrix}1 & -\mu_{i1}\mu_{i2}/b_{i12} & -\mu_{i1}\mu_{i3}/b_{i13}\\ -\mu_{i1}\mu_{i2}/b_{i12} & 1 & -\mu_{i2}\mu_{i3}/b_{i23}\\ -\mu_{i1}\mu_{i3}/b_{i13} & -\mu_{i2}\mu_{i3}/b_{i23} & 1\end{pmatrix},\\ b_{ikl} &= [\mu_{ik}(1-\mu_{ik})\mu_{il}(1-\mu_{il})]^{1/2}.\end{aligned}$$

4.2 Mixed ordinal logistic regression

The fixed effects ordinal logistic model can be easily extended to a mixed ordinal logistic regression model by adding random effects terms and allowing covariance patterns in the residual matrix.

In Section 4.2.1 the ordinal mixed model will be specified. The residual matrix for mixed categorical models has a more complex form than for GLMMs and will be defined in Section 4.2.2. As in GLMMs, there can be benefits in reparameterising random effects models as covariance pattern models and this will be discussed in Section 4.2.3. A quasi-likelihood function for the model is defined in Section 4.2.4, and model fitting methods are discussed in Section 4.2.5.

4.2.1 Definition of the mixed ordinal logistic regression model

The ordinal mixed model can now be specified as

$$\begin{aligned}&\mathbf{y}=\boldsymbol{\mu}+\mathbf{e},\\ &\log(\boldsymbol{\mu}^{[c]}/(1-\boldsymbol{\mu}^{[c]}))=\mathbf{X}\boldsymbol{\alpha}+\mathbf{Z}\boldsymbol{\beta},\\ &\boldsymbol{\beta}\sim \mathrm{N}(\mathbf{0},\mathbf{G}),\\ &\mathrm{var}(\mathbf{e})=\mathbf{R}.\end{aligned}$$

$\boldsymbol{\beta}$ is a vector containing the random effects. Thus, if a model fitting two treatments and three visits as fixed and patients as random were fitted to the earlier example data, we could write

$$\boldsymbol{\beta} = (P_1, P_2)'.$$

$\mathbf{Z}$ is the design matrix for the random effects and has additional rows than previously to correspond to the extended number of observations. For our data, $\mathbf{Z}$ would be

$$\mathbf{Z} = \begin{array}{c} \begin{matrix} P_1 & P_2 \end{matrix} \\ \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array}$$

The G matrix

$\mathbf{G}$ is again a matrix of variance parameters corresponding to the random effects and coefficients and has the same form as given in Section 2.1. In the model we have considered, fitting one random effect (patient) $\mathbf{G}$ would have the form

$$\mathbf{G} = \begin{pmatrix} \sigma_{\mathrm{p}}^2 & 0 \\ 0 & \sigma_{\mathrm{p}}^2 \end{pmatrix},$$

where

σ_{p}^2 = patient variance component.

4.2.2 Residual variance matrix

The residual variance matrix, $\mathbf{R}$, needs to take into account, first, the multinomial correlations that occur within the binary vectors used for each observation (see Section 4.1) and, second, any covariance patterns defined at the residual level.

As in Section 4.1, we arrange $\mathbf{R}$ as $\mathbf{B}^{1/2}\mathbf{P}\mathbf{B}^{1/2}$, a product of a correlation matrix, $\mathbf{P}$, and the matrix of expected Bernoulli variances, $\mathbf{B}$.

In a covariance pattern model the correlation matrix, $\mathbf{P}$, will include off-diagonal blocks of correlation parameters. For example, if a general covariance pattern was used to model our example data, then we would require a separate block of correlations for each pair of time points and $\mathbf{P}$ would have the form

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{12} & \mathbf{P}_{22} & \mathbf{P}_{23} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{13} & \mathbf{P}_{23} & \mathbf{P}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{44} & \mathbf{P}_{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{12} & \mathbf{P}_{55} \end{pmatrix}.$$

Note that the diagonal matrix blocks will differ for each observation as indicated at the end of Section 4.1. The $\mathbf{P}_{mn}$ are $(c-1)\times(c-1) = 3\times 3$ submatrices of parameters corresponding to the correlation between observations at visits m and n on the same patient. They replace the single correlation values used in GLMMs, and here we assume they take the form

$$\mathbf{P}_{mn} = \begin{pmatrix} p_{mn,11} & p_{mn,12} & p_{mn,13} \\ p_{mn,12} & p_{mn,22} & p_{mn,23} \\ p_{mn,13} & p_{mn,23} & p_{mn,33} \end{pmatrix},$$

with a separate correlation parameter used for each pair of partitions. Thus, six parameters are used for each $\mathbf{P}_{mn}$ matrix here. This is the parameterisation used by Lipsitz *et al.* (1994) who has written an accompanying `SAS` macro for fitting the model in this form. However, because this model requires more covariance parameters than GLMMs, particularly when the number of categories is high (increased by a factor of $c(c-1)/2$), more complex covariance patterns should be used with caution.

Simpler parameterisation for covariance pattern models

Here, we suggest an alternative simpler parameterisation with just one parameter corresponding to each of the parameters in the original covariance pattern (i.e. one for compound symmetry, three here for a general pattern). The correlation matrix for our example using a general covariance pattern would be

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \theta_{12}\mathbf{P}_{12} & \theta_{13}\mathbf{P}_{13} & \mathbf{0} & \mathbf{0} \\ \theta_{12}\mathbf{P}_{12} & \mathbf{P}_{22} & \theta_{23}\mathbf{P}_{23} & \mathbf{0} & \mathbf{0} \\ \theta_{13}\mathbf{P}_{13} & \theta_{23}\mathbf{P}_{23} & \mathbf{P}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{44} & \theta_{12}\mathbf{P}_{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \theta_{12}\mathbf{P}_{12} & \mathbf{P}_{55} \end{pmatrix}.$$

The $\mathbf{P}_{ij}$ submatrices have a similar form to that given earlier for the $\mathbf{P}_{ii}$ except that they now use the expected values corresponding to observations i and j:

$$\mathbf{P}_{ij} = \begin{pmatrix} 1 & -\mu_{i1}\mu_{j2}/b_{ij,12} & -\mu_{i1}\mu_{j3}/b_{ij,13} \\ -\mu_{i1}\mu_{j2}/b_{ij,12} & 1 & -\mu_{i2}\mu_{j3}/b_{ij,23} \\ -\mu_{i1}\mu_{j3}/b_{ij,13} & -\mu_{i2}\mu_{j3}/b_{ij,23} & 1 \end{pmatrix},$$

$$b_{ij,kl} = [\mu_{ik}(1-\mu_{ik})\mu_{jl}(1-\mu_{jl})]^{1/2}.$$

The θ_{mn} define the covariance pattern and can be parameterised to fit most of the covariance patterns described in Section 6.2. For example, for a compound symmetry pattern, the θ_{mn} would all have the same value and we could write

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \theta\mathbf{P}_{12} & \theta\mathbf{P}_{13} & \mathbf{0} & \mathbf{0} \\ \theta\mathbf{P}_{12} & \mathbf{P}_{22} & \theta\mathbf{P}_{23} & \mathbf{0} & \mathbf{0} \\ \theta\mathbf{P}_{13} & \theta\mathbf{P}_{23} & \mathbf{P}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{44} & \theta\mathbf{P}_{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \theta\mathbf{P}_{12} & \mathbf{P}_{55} \end{pmatrix}.$$

We have not explored the use of this simpler parameterisation in our worked examples as it is not easily fitted using SAS. However, it may be preferable to the correlation matrix suggested by Lipsitz *et al.*, since fewer parameters are used.

Dispersion parameter

As in GLMMs, variance at the residual level can be increased (or decreased) by using a dispersion parameter. The residual variance is multiplied by the dispersion parameter, ϕ, so that

$$\mathbf{R} = \phi\mathbf{B}^{1/2}\mathbf{P}\mathbf{B}^{1/2}.$$

We suggest that it is usually beneficial to fit a dispersion parameter in random effects models as in GLMMs; however, this is not possible in some software packages.

4.2.3 Likelihood and quasi-likelihood functions

The model we have defined, based on binary observations, is now in the form of a GLMM and a quasi-likelihood function can be defined in the same way as described in Section 3.2.2. A general form for the log quasi-likelihood for a GLMM that may contain random effects, coefficients and/or covariance patterns is again

$$\log\{QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y})\} = \log\{QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{R}}; \mathbf{y} \mid \boldsymbol{\beta})\} - 1/2 \log|\mathbf{G}| - 1/2\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta} + K,$$

where

$$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\mathbf{G}}, \boldsymbol{\gamma}_{\mathbf{R}}).$$

This function will correspond to a true log-likelihood function whenever the residuals from the original observations are uncorrelated (i.e. no $\boldsymbol{\gamma}_{\mathbf{R}}$ parameters are included) since $QL(\boldsymbol{\alpha}, \boldsymbol{\gamma}_{\mathbf{R}}; \mathbf{y} \mid \boldsymbol{\beta})$ will then follow a multinomial distribution.

4.2.4 Model fitting methods

Now that the model is in the form of a GLMM, it can be fitted using the approaches suggested in Section 3.2.3. However, it is now necessary to accommodate the multinomial within-observation covariances, and this adds a further degree of complexity to the computation. Several published examples have used generalised estimating equations to fit covariance pattern models (e.g. Lipsitz *et al.*, 1994; Liang *et al.*, 1992; Kenward *et al.*, 1994). Lipsitz *et al.* provide a `SAS` macro, and for this reason we have used their approach to analyse some of the examples in this book. The pseudo-likelihood approach can be used and random effects models are available in the `SAS` procedure `PROC GLIMMIX`. However, the procedure is not at present adapted to fit covariance pattern models. Both Hedeker and Gibbons (1994) and Goldstein (2003) have also suggested approaches for fitting random effects (and coefficients) models.

Alternatively, a Bayesian approach (see Sections 2.3 and 3.2.3) can be used for analysing random effects and coefficients models. For this approach, it is not formally necessary to redefine the observations in the extended binary form. A method such as MCMC can be used to simulate the posterior distribution from the categorical mixed model by assuming a multinomial distribution for the μ_{ij} (i.e. the probabilities of y_i being in each of the categories). Non-informative prior distributions can again be used for all parameters: for example, normal distributions with very large variances for fixed effects and inverse gamma distributions with very small parameters for variance components.

4.3 Mixed models for unordered categorical data

So far in this chapter we have only considered models for ordered categorical data. Although less frequent, unordered categorical variables are sometimes encountered in medicine. Blood group and colour are examples, since there is no natural ordering of their categories. A mixed model for these types of data can be defined in a very similar way to the ordinal mixed model. Again, the data can be re-expressed in extended binary form so that they are in the form of a

GLMM. The main difference from the ordinal mixed model is that the multinomial probabilities, $\boldsymbol{\mu}$, are now linked to the model parameters using 'generalised logits' rather than the logits of the cumulative probabilities used for ordinal data. The generalised logits can be calculated as the logs of the ratios of the probabilities of being in each category to that of being in the last category, that is by $\log(\boldsymbol{\mu}/\boldsymbol{\mu}_{\mathrm{L}})$, where $\boldsymbol{\mu}_{\mathrm{L}}$ is a vector containing the multinomial probabilities of each observation being in the last category. The model can be specified by

$$\begin{aligned}\mathbf{y} &= \boldsymbol{\mu} + \mathbf{e},\\ \log(\boldsymbol{\mu}/\boldsymbol{\mu}_{\mathrm{L}}) &= \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta},\\ \boldsymbol{\beta} &\sim \mathrm{N}(\mathbf{0}, \mathbf{G}),\\ \mathrm{var}(\mathbf{e}) &= \mathbf{R}.\end{aligned}$$

If there were four categories, we could write

$$\boldsymbol{\mu}_4 = (\mu_{14}, \mu_{14}, \mu_{14}, \mu_{24}, \mu_{24}, \mu_{24}, \ldots)',$$

and the vector of generalised logits as

$$\log(\boldsymbol{\mu}/\boldsymbol{\mu}_4) = (\mu_{11}/\mu_{14}, \mu_{12}/\mu_{14}, \mu_{13}/\mu_{14}, \mu_{21}/\mu_{24}, \mu_{22}/\mu_{24}, \mu_{23}/\mu_{24}, \ldots)'.$$

The choice of the last category for the denominator is arbitrary. Any of the categories can, in fact, be used and sometimes convergence will be more likely if the largest category is chosen. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are again vectors containing the fixed and random effects. However, a separate parameter is now needed for each category (except the last) because the proportional odds assumption used for ordinal data does not hold. We illustrate this model using the following hypothetical dataset, which contains the first five observations from a repeated measures trial in which **y** is an unordered categorical variable.

Patient	Visit	Treatment	y
1	1	A	2
1	2	A	1
1	3	A	4
2	1	B	3
2	2	B	1

In a simple model ignoring the effect of visits and fitting treatments as fixed and patients as random, we could write

$$\begin{aligned}\boldsymbol{\alpha} &= (I_1, I_2, I_3, T_{\mathrm{A},1}, T_{\mathrm{A},2}, T_{\mathrm{A},3}, T_{\mathrm{B},1}, T_{\mathrm{B},2}, T_{\mathrm{B},3})',\\ \boldsymbol{\beta} &= (P_{1,1}, P_{1,2}, P_{1,3}, P_{2,1}, P_{2,2}, P_{2,3})',\end{aligned}$$

where

I_j = intercept of the jth category,
$T_{k,j}$ = effect for treatment k, category j,
$P_{i,j}$ = effect for patient i, category j.

Each treatment and patient effect now has a separate parameter corresponding to each category of the data (except the last). The **X** and **Z** design matrices also have extra columns corresponding to the extra parameters and have the form

$$\mathbf{X} = \begin{array}{c} \begin{array}{ccccccccc} I_1 & I_2 & I_3 & T_{A,1} & T_{A,2} & T_{A,3} & T_{B,1} & T_{B,2} & T_{B,3} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\mathbf{Z} = \begin{array}{c} \begin{array}{cccccc} P_{1,1} & P_{1,2} & P_{1,3} & P_{2,1} & P_{2,2} & P_{2,3} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

4.3.1 The G matrix

G now contains blocks of variance parameters to allow for the fact that separate effects are specified for each partition and that these effects are correlated within each patient. A separate covariance parameter can be specified for each of the parameters corresponding to each pair of partitions. For our example data we may write

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{\mathrm{p}} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\mathrm{p}} \end{pmatrix},$$

where

$$\mathbf{G}_{\mathrm{p}} = \begin{pmatrix} \sigma^2_{\mathrm{p},11} & \theta_{\mathrm{p},12} & \theta_{\mathrm{p},13} \\ \theta_{\mathrm{p},12} & \sigma^2_{\mathrm{p},22} & \theta_{\mathrm{p},23} \\ \theta_{\mathrm{p},13} & \theta_{\mathrm{p},23} & \sigma^2_{\mathrm{p},33} \end{pmatrix},$$

$\sigma^2_{p,jj}$ = patient variance component for category j,
$\theta_{p,jk}$ = patient covariance parameter of the category pair j, k.

Alternatively, a model with a simpler parameterisation could be proposed. For example, we could make the assumption that each category had the same variance component and that the random effects for the different partitions were uncorrelated within patients. We could then write

$$\mathbf{G}_{\mathrm{p}} = \sigma^2_{\mathrm{p}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where

σ^2_{p} = patient variance component.

4.3.2 The R matrix

This matrix has the same form as the ordinal mixed model (see Section 4.2.2).

4.3.3 Fitting the model

The model can be fitted using similar techniques to those described for ordinal mixed models (Section 4.2.5). The method defined by Lipsitz *et al.* (1994) can also be used with unordered categorical data, and their SAS macro contains an option for specifying that the data are unordered.

4.4 Practical application and interpretation

In this section, some points relating to the practical application and interpretation of categorical mixed models will be considered. However, we should point out that

experience with these models is limited, and therefore some of the issues are still far from resolved.

4.4.1 Expressing fixed and random effects results

The fixed and random effects estimates are given in terms of logits which become more interpretable when exponentiated to give odds ratios (see Section 3.3.4). For ordinal data, these can be interpreted strictly as the ratio of the odds of being in a lower category, whichever partition of the category is chosen. However, usually it is appropriate to interpret the odds ratios obtained as the ratio of the odds of being in a lower category, averaged over all the possible partitions of the categories. Note that some software may parameterise the model so that the ratio of the odds of being in a higher category is obtained. Confidence intervals for the odds ratios can be obtained by exponentiating confidence intervals calculated on the linear scale in the same way as for GLMMs (see Section 3.3.9).

4.4.2 The proportional odds assumption

The fixed and random effects, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, are assumed to be the same at each partition in models for ordinal data. For example, if there are four categories, an equal $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are assumed whether a 1/2–4, 1–2/3–4 or 1–3/4 partition is made. This means that the odds for effects are proportional across all partitions. This assumption could be tested by fitting a separate set of fixed and random effects at each partition, $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$, and testing the equality of the effects at different partitions. When there is a significant difference between fixed effects estimates for each partition (i.e. the proportional odds assumption does not hold), it may be informative to analyse each partition separately using binary GLMMs. However, often the 'average' $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ over all partitions will be of greatest interpretational value even if the results differ significantly between each partition.

4.4.3 Number of covariance parameters

The number of covariance parameters required by a covariance pattern model is increased by a factor of $(c-1)\times(c-2)/2$ over an equivalent GLMM (unless the alternative simpler parameterisation is used, see Section 4.2.3). Thus, the model can use a large number of covariance parameters and this can sometimes lead to inaccurate estimates or convergence problems. Therefore, the more complex patterns should be used only cautiously, particularly in small datasets or when the number of categories is high. It may also be worth considering combining any categories, which have small frequencies with neighbouring categories.

Alternatively, if there are a large number of categories, say about five or more, it might be worth trying a normal mixed model and checking the resulting residuals. If the model assumptions are approximately satisfied, this type of model would also have the advantage of being simpler to interpret.

4.4.4 Choosing a covariance pattern

As in GLMMs, approximate likelihood ratio tests based on quasi-likelihood values could be used to compare models fitting different covariance parameters. However, a quasi-likelihood value is not always produced by software procedures. For those without access to software providing quasi-likelihoods, we suggest favouring the most simple patterns (e.g. compound symmetry or first-order autoregressive) and using more complex patterns only with larger datasets where there is a strong suggestion that the covariance pattern deviates from a simpler pattern.

4.4.5 Interpreting covariance parameters

It is difficult to interpret the size of covariance parameters fitted in the **R** matrix, since blocks containing $(c-1) \times (c-2)/2$ parameters are estimated rather than individual parameters. The diagonal terms representing correlation between observations in the same categories are the most helpful. However, if the alternative simpler parameterisation is used (see Section 4.2.3), there is only one parameter per block and the parameters can be interpreted in the same way as those from ordinary repeated measures analyses (see Section 6.2).

4.4.6 Checking model assumptions

As in the GLMM, it is assumed that the random effects are normally distributed and uncorrelated. This is difficult to assess but, as for normal data (see Section 2.4.6), fixed effects and their standard errors are not expected to be sensitive to a lack of normality in the random effects. Plots of random effects can be used to identify any outlying effect categories.

4.4.7 The dispersion parameter

The same considerations apply as for Bernoulli data described in Section 3.3.7.

4.4.8 Other points

Points covered in Section 3.3 relating to uniform fixed and random effects categories (3.3.2), negative variance components (3.3.3), bias in fixed and random effects standard errors (3.3.6), significance testing (3.3.8), confidence intervals (3.3.9) and determining whether the simulated posterior distribution has converged (3.3.11) also apply to categorical mixed models.

4.5 Example

In this example, we will consider the analysis of an adverse event, 'cold feet', which was recorded at each visit in the hypertension study introduced in Section 1.3. In Section 3.4, we considered 'cold feet' as a binary variable. Here we will analyse it on its original ordinal scale of 1–5: 1 = none, 2 = occasionally, 3 = on most days, 4 = most of the time, 5 = all of the time. The frequencies of each category by treatment and visit are shown in Table 4.1. Two sets of analyses will be used to illustrate the use of both random effects and covariance pattern models. The first will analyse the last values recorded for each patient and consider the data as a multi-centre design. The second will consider all the post-treatment values and analyse the data as a repeated measures design, ignoring the effects of centres.

Multi-centre analysis

Analyses were carried out using the following models.

Model	Fixed effects	Random effects	Method
1	Baseline, treatment		Maximum likelihood
2a	Baseline, treatment	Centre, treatment·centre	Pseudo-likelihood
2b	Baseline, treatment	Centre, treatment·centre	MCMC

Model 1 does not include fixed centre effects as this would lead to uniform treatment effects as occurred in the binary analysis of cold feet. Therefore, this model is not recommended for estimating treatment effects (see Section 3.4.2, Model 2). Models 2a and 2b take into account the random variation in the treatment effect between centres and results and can be related with more confidence to the 'population' of potential centres. Baseline cold feet is also measured on the same scale of 1–5 and is fitted as a categorical effect.

Table 4.1 Frequencies of 'cold feet' severity by treatment and visit.

		Category				
Treatment	**Visit**	**1**	**2**	**3**	**4**	**5**
A	3	83	4	6	0	4
	4	72	5	6	3	3
	5	70	3	5	2	3
	6	63	5	3	3	2
B	3	69	9	5	2	6
	4	65	7	10	3	3
	5	54	10	8	6	8
	6	55	5	8	4	7
C	3	79	2	7	1	1
	4	85	1	4	0	1
	5	82	3	3	2	1
	6	78	4	3	1	1
Total		855	58	68	27	40

Results

Estimates of the variance components and treatment effects are shown in Table 4.2. Note that the models are parameterised in terms of ratios of lower to higher categories of cold feet. Thus, a positive difference relates to *less* frequent cold feet on the first treatment than on the second. The centre and centre

Table 4.2 Estimates of variance components and fixed effects (on the logit scale).

	Treatment effects (SE, p)		
Model	**A − B**	**A − C**	**B − C**
1	0.65 (0.36, p = 0.07)	−0.87 (0.47, p = 0.06)	−1.53 (0.45, p = 0.0009)
2a	0.66 (0.37, p = 0.07)	−0.87 (0.47, p = 0.07)	−1.53 (0.45, p = 0.0009)
2b	0.70 (0.40, p = 0.09)	−0.94 (0.51, p = 0.06)	−1.63 (0.49, p = 0.006)

	Variance components (SEs, centiles)	
Model	**Centre**	**Treatment·centre**
1	–	–
2a	0.064	0.000
2b	0.075	0.057

treatment variance components in Models 2a and 2b are small compared with those obtained in the binary analyses of cold feet. This is because the dispersion parameter is fixed at one as there is not currently an option to fit a dispersion parameter when modelling ordinal data in `PROC GLIMMIX`. Since there are uniform centre and centre treatment effects, this is likely to have caused a downward bias in the centre and centre treatment variance component as these models will be unable to compensate for the effects of random effects shrinkage (see Section 3.3.5). A downward bias in the centre·treatment variance component will cause the treatment effect standard errors in Model 2a to be underestimated to some extent. However, this is not expected to be a problem in Model 2b, which is fitted using MCMC.

All of the models showed that cold feet were more frequent on treatment B than on treatment C. The results are very similar between Models 1 and 2a. Although Model 2a has estimated a small positive centre variance component, it leads to no noticeable recovery of information on the treatment effects. However, the results for Model 2b fitted using MCMC show more marked differences. This is likely to be mainly due to the positive centre·treatment variance component which causes differences in the treatment effect estimates and an increase in their standard errors, reflecting the additional variation occurring across the centres. As the MCMC approach does not use linear approximations in obtaining the parameter estimates (as pseudo-likelihood does), we would suggest that these results are preferable to those obtained using Model 2a.

Baseline cold feet effects were highly significant in each analysis, increasing the efficiency of the models. However, the effects are not monotonic (see `SAS` output). This happens because of a very low number in some categories at baseline – the numbers in the five categories are 243, 13, 17, 2 and 8. It is the category with only two patients that causes the loss of monotonicity. To avoid this, the data could have been reanalysed with the last two categories combined. However, given the size of the dataset, there are likely only to be small differences in the overall treatment effect results. Alternatively, the baseline category could have been treated as a quantitative covariate.

Odds ratios may be obtained by reversing the signs of the treatment differences on the linear scale and taking their exponentials (Table 4.3). They can be interpreted as the ratio of the odds of having more frequent cold feet. For example, the odds ratio corresponding to the difference A – B is based on the probability of more cold feet on treatment A compared with treatment B. In this context, more cold feet can be considered as category 5 (all the time) versus the other categories; as categories 4 or 5 (most or all of the time) versus the rest; as categories 3, 4 or 5 (on most days or worse) versus the rest; or as categories 2–5 (occasionally or worse) versus category 1 (none). Note that it is an inherent assumption of this model that the same odds ratio applies to every partition between the categories (see Section 4.4.2).

Table 4.3 Odds ratios and 95% confidence intervals from Model 2a.

Effect	Odds ratio
A/B	0.52 (0.25, 1.06)
A/C	2.38 (0.94, 5.88)
B/C	4.55 (1.89, 11.11)

Repeated measures analysis

We will consider analysing the data using a variety of covariance patterns (Model 1 – uncorrelated; Model 2 – compound symmetry; Model 3 – Toeplitz; Model 4 – general) using the SAS macro written by Lipsitz *et al.* (1994). Each model will fit baseline cold feet, treatment and visit effects as fixed effects. Treatment by visit effects were found to be non-significant on initial analysis and therefore have been excluded from each model. Model 4 using a general pattern did not converge even when categories 4 and 5 were combined. This is likely to be due to the large number of covariance parameters that needed to be estimated by the Model (60).

We will first consider the correlation parameter estimates arising from the models. These are rather more difficult to interpret than parameter estimates from covariance patterns in normal mixed models or GLMMs, because there is now a 4×4 matrix block of parameters representing the correlation between a pair of visits. Recall from Section 4.2.2 that

$$\mathbf{R} = \mathbf{B}^{1/2}\mathbf{P}\mathbf{B}^{1/2}.$$

In this example, the correlation matrix, $\mathbf{P}$, has the block diagonal form

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} & \mathbf{P}_{14} & \mathbf{0} & \mathbf{0} & . & . \\ \mathbf{P}_{12} & \mathbf{P}_{22} & \mathbf{P}_{23} & \mathbf{P}_{24} & \mathbf{0} & \mathbf{0} & . & . \\ \mathbf{P}_{13} & \mathbf{P}_{23} & \mathbf{P}_{33} & \mathbf{P}_{34} & \mathbf{0} & \mathbf{0} & . & . \\ \mathbf{P}_{14} & \mathbf{P}_{24} & \mathbf{P}_{34} & \mathbf{P}_{44} & \mathbf{0} & \mathbf{0} & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{55} & \mathbf{P}_{12} & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{12} & \mathbf{P}_{66} & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \end{pmatrix},$$

where the $\mathbf{P}_{ii}$ are the multinomial 'within-observation' correlation matrices and the $\mathbf{P}_{mn}$ give the correlations between observations on the same patient between visits m and n. The $\mathbf{P}_{mn}$ matrices obtained from Models 1–3 are shown in Table 4.4.

Statistical comparisons between the models using likelihood ratio tests were not readily available because quasi-likelihood values were not produced by the SAS macro. On informal examination, the positive diagonal terms in $\mathbf{P}_{ii}$ matrix blocks

Table 4.4 $\mathbf{P}_{mn}$ correlation submatrices.

Model													
		All visit pairs (i.e. all $\mathbf{P}_{mn}$, $m \neq n$)											
1	**Partition**	**1**	**2**	**3**	**4**								
	1	0.00											
	2	0.00	0.00										
	3	0.00	0.00	0.00									
	4	0.00	0.00	0.00	0.00								
		All visit pairs (i.e. all $\mathbf{P}_{mn}$, $m \neq n$)											
2	**Partition**	**1**	**2**	**3**	**4**								
	1	0.38											
	2	0.01	0.27										
	3	−0.07	−0.02	0.09									
	4	−0.12	−0.06	0.10	0.12								
		Visit separation = 1 (i.e. $\mathbf{P}_{12}, \mathbf{P}_{23}, \mathbf{P}_{34}$)				**Visit separation = 2 (i.e. $\mathbf{P}_{13}, \mathbf{P}_{24}$)**				**Visit separation = 3 (i.e. $\mathbf{P}_{14}$)**			
3	**Partition**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
	1	0.43				0.32				0.41			
	2	0.00	0.30			0.02	0.21			0.03	0.33		
	3	−0.10	−0.02	0.23		−0.10	−0.03	0.02		0.05	−0.03	−0.18	
	4	−0.20	−0.12	0.15	0.07	−0.06	−0.02	0.10	0.05	0.02	−0.03	−0.07	0.35

for Models 2 and 3 indicate that the repeated observations on the same patient are correlated. Therefore, Model 1, which has zero correlations and assumes that the observations are independent, should be rejected. The banded pattern (Model 3) does not show marked differences between the correlations depending on visit separation. Therefore, we might choose to base our conclusions on Model 2 with a simpler covariance pattern.

Treatment effect estimates are shown with 'model-based' standard errors in Table 4.5. In Model 1, which naively assumes that the repeated observations are independent, the standard error estimates are noticeably smaller than for Models 2 and 3.

The overall treatment effects in Model 2 were highly significant ($p = 0.0007$). Cold feet were significantly more likely on treatment B than on treatment A ($p = 0.03$), and on treatment B than on treatment C ($p = 0.0002$). The coefficients

Table 4.5 Treatment effect estimates.

Model	Treatment effect (SE)		
	A – B	**A – C**	**B – C**
1 Uncorrelated	0.72 (0.20)	−0.54 (0.25)	−1.26 (0.23)
2 Compound symmetry	0.65 (0.29)	−0.61 (0.35)	−1.26 (0.33)
3 Toeplitz	0.65 (0.29)	−0.57 (0.35)	−1.21 (0.33)

Table 4.6 Odds ratios calculated from Model 2.

Effect	Odds ratio
A/B	0.52 (0.30, 0.92)
A/C	1.84 (0.93, 3.65)
B/C	3.52 (1.84, 6.73)

for the treatment effects are difficult to interpret directly, but, as before, by exponentiation we can calculate odds ratios and 95% confidence intervals (Table 4.6). Confidence intervals are calculated using the 'model-based' standard errors and the $z_{0.975}$ statistic. The exact DF for t statistics was not available from the SAS macro. However, the patient DF of over 300 can be taken as a conservative estimate and the t statistic is then well approximated by the $z_{0.975}$ statistic.

SAS code and output

Multicentre analysis

Model 1

Variables
cf = post-treatment cold feet (1–5)
cf1 = pre-treatment cold feet (1–5)

```
PROC GLIMMIX; CLASS centre treat cf1;
MODEL cf=cf1 treat/ DIST=MULT DDFM=KR S;
ESTIMATE 'A-B' treat 1 -1 0/ CL OR;
ESTIMATE 'A-C' treat 1 0 -1/ CL OR;
ESTIMATE 'B-C' treat 0 1 -1/ CL OR;
```

(Note PROC GENMOD or PROC LOGISTIC could have equivalently been used to fit this model, although the signs for mean effects may differ between the procedures.)

Only output relating to the fixed effects estimates is given:

Parameter Estimates

Effect	cf	treat	cf1	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	1			-1.3254	0.8152	273	-1.63	0.1051
Intercept	2			-0.7739	0.8090	273	-0.96	0.3396
Intercept	3			0.07031	0.8003	273	0.09	0.9301
Intercept	4			0.6673	0.7995	273	0.83	0.4047
cf1			1	4.0832	0.7794	273	5.24	<.0001
cf1			2	3.1632	1.0050	273	3.15	0.0018
cf1			3	1.2127	0.8428	273	1.44	0.1513
cf1			4	2.9671	1.4723	273	2.02	0.0449
cf1			5	0	.	.	.	.
treat		A		-0.8745	0.4709	273	-1.86	0.0644
treat		B		-1.5292	0.4536	273	-3.37	0.0009
treat		C		0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
cf1	4	273	13.30	<.0001
treat	2	273	5.90	0.0031

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
A-B	0.6547	0.3640	273	1.80	0.0732	0.05	-0.06190	1.3713
A-C	0.8745	0.4709	273	-1.86	0.0644	0.05	-1.8016	0.05252
B-C	-1.5292	0.4536	273	-3.37	0.0009	0.05	-2.4222	-0.6363

Label	Exponentiated Estimate	Exponentiated Lower	Exponentiated Upper
A-B	1.9246	0.9400	3.9405
A-C	0.4171	0.1650	1.0539
B-C	0.2167	0.08873	0.5293

Model 2a

```
PROC GLIMMIX; CLASS centre treat cf1;
MODEL cf=cf1 treat/ DIST=MULT DDFM=KR S;
RANDOM centre centre*treat;
ESTIMATE 'A-B' treat 1 -1 0/ CL OR;
ESTIMATE 'A-C' treat 1 0 -1/ CL OR;
ESTIMATE 'B-C' treat 0 1 -1/ CL OR;
```

Only the output relating to the variance components and fixed effects estimates is shown:

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
centre	0.06431	0.1636
centre*treat	0	.

Solutions for Fixed Effects

Effect	cf	treat	cf1	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			1	-1.2812	0.8326	273	-1.54	0.1250
Intercept			2	-0.7274	0.8272	273	-0.88	0.3800
Intercept			3	0.1197	0.8160	273	0.15	0.8835
Intercept			4	0.7211	0.8109	273	0.89	0.3747
cf1			1	4.0224	0.7776	273	5.17	<.0001
cf1			2	3.1525	0.9933	273	3.17	0.0017
cf1			3	1.1650	0.8459	273	1.38	0.1696
cf1			4	2.9810	1.5950	273	1.87	0.0627
cf1			5	0	.	.	.	.
treat		A		-0.8689	0.4741	273	-1.83	0.0679
treat		B		-1.5273	0.4569	273	-3.34	0.0009
treat		C		0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
cf1	4	273	12.48	<.0001
treat	2	273	5.85	0.0032

```
                              Estimates

                Standard
Label Estimate Error  DF  t Value Pr > |t| Alpha  Lower  Upper

A-B    0.6584  0.3653 273   1.80  0.0726 0.05 -0.06075 1.3775
A-C   -0.8689  0.4741 273  -1.83  0.0679 0.05 -1.8023  0.06452
B-C   -1.5273  0.4569 273  -3.34  0.0009 0.05 -2.4268 -0.6278

                Exponentiated    Exponentiated    Exponentiated
     Label        Estimate            Lower            Upper

     A-B            1.9317           0.9411           3.9651
     A-C            0.4194           0.1649           1.0666
     B-C            0.2171          0.08832           0.5338
```

Model 2b

Variables

`alpha01-alpha04` = 4 intercepts
`alpha11-alpha14` = parameters for four binary baseline cold feet variables
`alpha2, alpha3` = parameters for treatment differences A − C and B − C
`cf11-cf14` = values of four binary baseline cold feet variables (recorded in dataset)
`cfm1-cfm5` = values of five binary cold feet variables (recorded in dataset)
`treata, treatb` = values of for treatment differences A − C and B − C
`b_centre, b_ct` = centre and centre·treatment parameters
`v2, v3` = centre and centre·treatment variance component parameters

```
ODS GRAPHICS ON;
PROC MCMC DATA=C OUTPOST=post NMC=200000 THIN=10 SEED=7899;
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
    POSTINTERVALS;
ARRAY cfm[5] cfm1 cfm2 cfm3 cfm4 cfm5;
PARMS alpha01 0 alpha02 0.1 alpha03 0.2 alpha04 0.3 alpha11 0
      alpha12 0 alpha13 0 alpha14 0 alpha2 0 alpha3 0
      v2 1 v3 1;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_centre ~ NORMAL(0, VAR = v2) SUBJECT=centre;
RANDOM b_ct ~ NORMAL(0, VAR = v3) SUBJECT=centre_treat;
mu1 = alpha01 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
```

```
mu2 = alpha02 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
mu3 = alpha03 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
mu4 = alpha04 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + + b_ct;
p1  = LOGISTIC(mu1);
p2  = LOGISTIC(mu2)-LOGISTIC(mu1);
p3  = LOGISTIC(mu3)-LOGISTIC(mu2);
p4  = LOGISTIC(mu4)-LOGISTIC(mu3);
p5  = 1-LOGISTIC(mu4);
ARRAY p[5] p1 p2 p3 p4 p5;
MODEL cfm ~ MULTINOM(p);
RUN;
ODS GRAPHICS OFF;
```

Some of the code has a similar form to that used to analyse cold feet as a binary variable in Section 3.4. The main differences here are the need to define the following:

- dummy binary variables for the (four) intercepts and for baseline cold feet,
- each of the five predicted probabilities needed to define the multinomial distribution, `p1-p5`, based on differences in the predicted cumulative probabilities (note the `SAS LOGISTIC` function is $(1+\exp(-\text{mu}))^{-1}$),
- a multinomial distribution for cold feet.

Note this procedure requires that the starting values for the intercept parameters, `alpha01-alpha04`, that are set in ascending order. If this is not done, the initial values taken by `SAS` do not lead to valid probabilities for the multinomial distribution and the simulation fails.

Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
alpha01	20000	-1.3968	0.9145	-1.9922	-1.3719	-0.7876
alpha02	20000	-0.8030	0.9077	-1.3995	-0.7781	-0.2004
alpha03	20000	0.1043	0.9001	-0.4896	0.1254	0.7111
alpha04	20000	0.7841	0.9013	0.1909	0.8027	1.3918
alpha11	20000	4.2863	0.8621	3.7004	4.2556	4.8365
alpha12	20000	3.5175	1.1067	2.7634	3.4794	4.2410

```
alpha13   20000  1.2757    0.9227    0.6466   1.2511   1.8694
alpha14   20000  3.4559    1.7705    2.2722   3.3860   4.5678
alpha2    20000 -0.9373    0.5077   -1.2696  -0.9306  -0.5948
alpha3    20000 -1.6324    0.4917   -1.9484  -1.6189  -1.2998
v2        20000  0.1460    0.1937    0.0298   0.0753   0.1832
v3        20000  0.1195    0.1661    0.0228   0.0573   0.1463

                          Posterior Intervals

Parameter  Alpha     Equal-Tail Interval       HPD Interval

alpha01    0.050     -3.2644      0.3424    -3.2145     0.3790
alpha02    0.050     -2.6352      0.9326    -2.6066     0.9542
alpha03    0.050     -1.7079      1.8449    -1.6099     1.9266
alpha04    0.050     -1.0260      2.5218    -0.9903     2.5487
alpha11    0.050      2.6878      6.0725     2.6667     6.0444
alpha12    0.050      1.4188      5.7848     1.2882     5.6249
alpha13    0.050     -0.4474      3.1922    -0.5123     3.1119
alpha14    0.050      0.2110      7.1732     0.0615     6.9542
alpha2     0.050     -1.9497      0.0409    -1.9623     0.0235
alpha3     0.050     -2.6391     -0.7098    -2.6171    -0.6923
v2         0.050      0.00662     0.6837     0.00120    0.5232
v3         0.050      0.00546     0.5901     0.00151    0.4459
```

The above tables provide estimates and probability intervals for the four intercept parameters (`alpha01-alpha04`), the four binary baseline cold feet parameters (`alpha11-alpha14`), the treatment differences A − C and B − C (`alpha2` and `alpha3`), and the centre and centre treatment variance components (`v2` and `v3`). Because cold feet was coded from 1 (none) to 5 (all the time), the probability of having less cold feet is being modelled. Therefore negative estimates for A − C and B − C indicate that there is more severe cold feet on treatments A and B than on treatment C.

The mean and SE for A − B, and odds ratios, *p*-values and 95% confidence intervals for all treatment differences, may be obtained using similar code to that used for the MCMC analyses in Sections 2.5 and 3.4. Similar code may also be used to assess convergence.

Repeated measures analysis

Variables

`cf` = post - treatment cold feet (15),
`cf1` = pre - treatment cold feet (15),
`treat` = treatment,
`visit` = visit.

Model 1

```
PROC GENMOD; CLASS treat visit cf1;
MODEL cf=cf1 treat visit/ TYPE3 WALD DIST=MULT;
ESTIMATE 'A-B' treat 1 -1 0/ ALPHA=0.05 EXP;
ESTIMATE 'A-C' treat 1 0 -1/ ALPHA=0.05 EXP;
ESTIMATE 'B-C' treat 0 1 -1/ ALPHA=0.05 EXP;
```

Model Information

Data Set	WORK.A
Distribution	Multinomial
Link Function	Cumulative Logit
Dependent Variable	cf
Number of Observations Read	1047
Number of Observations Used	1047

Class Level Information

Class	Levels	Values
treat	3	A B C
visit	4	1 2 3 4
cf1	5	1 2 3 4 5

Response Profile

Ordered Value	cf	Total Frequency
1	1	854
2	2	58
3	3	68
4	4	27
5	5	40

PROC GENMOD is modelling the probabilities of levels of cf having LOWER Ordered Values in the response profile table. One way to change this to model the probabilities of HIGHER Ordered Values is to specify the DESCENDING option in the PROC statement.

Parameter Information

Parameter	Effect	treat	visit	cf1
Prm1	cf1			1
Prm2	cf1			2
Prm3	cf1			3

Prm4	cf1		4
Prm5	cf1		5
Prm6	treat	A	
Prm7	treat	B	
Prm8	treat	C	
Prm9	visit		1
Prm10	visit		2
Prm11	visit		3
Prm12	visit		4

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Log Likelihood		-608.1612	
Full Log Likelihood		-608.1612	
AIC (smaller is better)		1242.3224	
AICC (smaller is better)		1242.6748	
BIC (smaller is better)		1306.7203	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence	Limits	Chi-Square	Pr>ChiSq
Intercept1		1	-2.7439	0.5214	-3.7658	-1.7219	27.69	<.0001
Intercept2		1	-2.1331	0.5175	-3.1473	-1.1189	16.99	<.0001
Intercept3		1	-0.9766	0.5043	-1.9650	0.0118	3.75	0.0528
Intercept4		1	-0.1334	0.4910	-1.0957	0.8289	0.07	0.7859
cf1	1	1	5.3827	0.4758	4.4500	6.3153	127.96	<.0001
cf1	2	1	4.3937	0.5774	3.2620	5.5253	57.90	<.0001
cf1	3	1	2.4437	0.4974	1.4688	3.4186	24.14	<.0001
cf1	4	1	3.8132	0.8041	2.2371	5.3892	22.49	<.0001
cf1	5	0	0.0000	0.0000	0.0000	0.0000	.42	.4112
treat	A	1	-0.5394	0.2509	-1.0311	-0.0476	4.62	0.0316
treat	B	1	-1.2628	0.2328	-1.7191	-0.8065	29.42	<.0001
treat	C	0	0.0000	0.0000	0.0000	0.0000	.42	.4112
visit	1	1	0.2255	0.2572	-0.2785	0.7296	0.77	0.3805
visit	2	1	0.2157	0.2589	-0.2918	0.7232	0.69	0.4049
visit	3	1	-0.1988	0.2472	-0.6834	0.2858	0.65	0.4214
visit	4	0	0.0000	0.0000	0.0000	0.0000	.42	.4112
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Wald Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr>ChiSq
cf1	4	215.75	<.0001
treat	2	32.27	<.0001
visit	3	4.08	0.2535

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence	Limits	L'Beta Estimate	Standard Error	Alpha
A-B	0.6734	0.5788	0.7556	0.7234	0.2069	0.05
Exp(A-B)				2.0614	0.4266	0.05
A-C	0.3683	0.2629	0.4881	-0.5394	0.2509	0.05
Exp(A-C)				0.5831	0.1463	0.05
B-C	0.2205	0.1520	0.3086	-1.2628	0.2328	0.05
Exp(B-C)				0.2829	0.0659	0.05

Label	L'Beta Confidence Limits		Chi-Square	Pr>ChiSq
A-B	0.3178	1.1290	12.22	0.0005
Exp(A-B)	1.3742	3.0924		
A-C	-1.0311	-0.0476	4.62	0.0316
Exp(A-C)	0.3566	0.9535		
B-C	-1.7191	-0.8065	29.42	<.0001
Exp(B-C)	0.1792	0.4464		

The only part of this output which is useful for this example is found from the column headed L'Beta Estimate onwards. These give the estimates and confidence limits on the linear scale along with the corresponding exponentiated values that give us the odds ratios. The first three columns of numbers are obtained by applying the inverse link to the L'Beta estimates and did not appear in earlier versions of SAS.

Models 2 and 3 The SAS macro written by Lipsitz *et al.* (1994) was used to perform these analyses (see Section 9.1). This macro and the SAS code used can be obtained from web page www.wiley.com/go/brown/applied_mixed. However, we note that the macro no longer appears run successfully for this example using SAS/STAT 12.3.

5

Multi-centre trials and meta-analyses

In this chapter, we consider the analysis of data that are collected from several centres or trials. Such datasets in which observations have a natural grouping can be described as *hierarchical*. Use of a random effects model to analyse a hierarchical dataset often leads to results that can be generalised more widely. Section 5.1 provides an introduction to multi-centre trials; the implications of fitting different models are considered in Section 5.2; a worked example is given in Section 5.3; some general points specific to hierarchical datasets are made in Section 5.4; and sample size estimation methods are introduced in Section 5.5. Meta-analysis is considered in Section 5.6 and an example follows in Section 5.7.

5.1 Introduction to multi-centre trials

5.1.1 What is a multi-centre trial?

A multi-centre trial is carried out at several centres because insufficient patients are available for the study at any one centre, or with the deliberate intention of assessing the effectiveness of treatments in several settings. Sometimes, there will be extra variability in treatment effect estimates, which can be due to differences between the centres (e.g. different investigators, types of patients, climates). This extra variation can be taken into account in the analysis by including centre and centre·treatment effects as random effects in the model. Such variation is likely to be most noticeable in trials that do not compare drugs. For example, in a trial to compare surgical procedures, there may be varying levels of experience available at each centre with the different procedures. This will lead us to expect a positive variance component for the centre·treatment effects.

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

5.1.2 Why use mixed models to analyse multi-centre data?

When centre and centre·treatment effects are fitted as random, allowance is made for variability in the magnitude of the treatment effects between centres. However, deciding whether centre and centre·treatment effects should be fixed or random is often the subject of debate. In practice, the choice will depend on whether treatment estimates are to relate only to the set of centres used in the study or, more widely, to the circumstances and locations of which the trial centres can be regarded as a sample. In the former case, local treatment estimates for the sampled set of individual centres are obtained by fitting centre and centre·treatment effects as fixed. To obtain global treatment estimates, centre and centre·treatment effects should be fitted as random. When this is done, the standard error of treatment differences is increased to reflect the heterogeneity of the treatment effects across centres.

If the centre·treatment term is omitted, there is a choice of whether to fit centre effects as fixed or random. Taking centre effects as random can increase the accuracy of treatment estimates, since information from the centre error stratum is used in addition to that from the residual stratum. Thus, it is nearly always beneficial to fit centres as random, regardless of whether a local or global interpretation is required. The amount of extra information will depend on the degree of treatment imbalance within the centres and the relative sizes of the variance components.

In the analysis of multi-centre trials, it is important to check whether results from any particular centre are outlying. If this occurs, it may be an indication that a centre has not followed the protocol correctly. In the fixed effects model, spurious outlying estimates caused by random variation may occur, particularly in small centres. In contrast, the shrunken estimates of centre and centre·treatment effects obtained by the random effects model do not have this problem.

5.2 The implications of using different analysis models

In this section, we look more closely at the implications of fitting centre and centre·treatment effects as fixed or random. We will consider four different models, and in each of them treatment effects and baseline effects (if available) are fitted as fixed.

5.2.1 Centre and centre·treatment effects fixed

Treatments effects

These are estimated with equal weight given to results from each centre regardless of size. If centre sizes vary greatly, this can cause results to differ markedly from

analyses not fitting centre·treatment effects as fixed. Another potential difficulty with this method is that treatment effects cannot be estimated at all unless all treatments are received at every centre. For example, if no patients received treatment A at one centre, then all comparisons involving treatment A would be non-estimable. In practice, we note, however, that this problem could be resolved by the amalgamation of centres with a small number of patients.

Treatment standard errors

These are based on the residual (within-centre) variation. The variance $(\text{SE})^2$ of a treatment difference is given by

$$\text{var}(t_i - t_j) = \sigma^2(1/n_i + 1/n_j),$$

where σ^2 is the residual variance and n_i and n_j are the number of patients receiving treatments i and j. When an equal number of patients, r, receive each treatment at each of c centres, we have $n_i = rc$ and the variance can be written as follows:

$$\text{var}(t_i - t_j) = 2\sigma^2/rc.$$

This variance will always be less than or equal to that for the model fitting centre·treatment effects as random (see Section 5.2.3).

Outlying centres

These can be determined from the centre and centre·treatment effect means. However, the estimates can be misleading for small centres, which could appear outlying owing only to random variation.

Inference

This strictly applies only to those centres that were included in the trial.

5.2.2 Centre effects fixed, centre·treatment effects omitted

This fixed effects model is often used when the centre·treatment effects in the previous model are non-significant. However, in practice, there is often a lack of power to detect small centre·treatment effects, and hence variability in the treatment effects across centres is often ignored.

Treatment estimates

These take account of differing centre sizes and are not estimated with equal weight given to results from each centre as in the previous model.

Treatment standard errors

These are based on the residual (within-centre) variation. The variances of treatment differences are given by the formula used in the previous model.

Outlying centres

These can be determined from the centre means. However, again, these can be misleading for small centres that could be apparently outlying owing only to random variation.

Inference

This strictly applies only to those centres that were included in the trial. However, extrapolation of the inferences regarding the effect of treatment to other centres seems more reasonable when an assumption has been made that the treatment effect does not depend on the centre in which it has been applied.

5.2.3 Centre and centre·treatment effects random

Unlike the commonly used fixed effects approach of dropping a non-significant centre·treatment interaction, centre·treatment effects are retained in the random effects model, provided the centre·treatment variance component, σ^2_{ct}, is positive. Thus, variation in treatment effects across centres is allowed even though its existence may not have been 'proven' by a significance test.

Treatment estimates

These take account of differing centre sizes. They are estimated using information from the centre·treatment error stratum and also from the centre stratum if treatment effects are not balanced across centres. They are estimable even if some treatments are not received at every centre.

Treatment standard errors

Treatment standard errors are based on the centre·treatment variation. If an equal number of patients receive each treatment at every centre, the variance of treatment effect differences can be obtained as

$$\text{var}(t_i - t_j) = 2(\sigma^2/rc + \sigma^2_{ct}/c).$$

Thus, the variance is directly related to the size of the centre·treatment variance component, σ^2_{ct}, and the number of centres sampled, c. It is always expected to be greater than the fixed effects model variance, $2\sigma^2/rc$, provided σ^2_{ct} is not allowed to be negative.

The variance is not as easily specified when centre sizes are unequal. However, if the proportion of patients allocated to each treatment is the same across centres, the variance is of the form

$$\text{var}(t_i - t_j) = \text{var}_{\text{FE}}(t_i - t_j) + k\sigma^2_{\text{ct}},$$

where $\text{var}_{\text{FE}}(t_i - t_j)$ is the fixed effects model variance and k is a positive constant. Thus, the variance is again always expected to be greater than or equal to that in the fixed effects model.

In most situations, though, the proportions of patients receiving each treatment will differ to some extent from centre to centre. When this is the case, a general formula for the variance cannot be specified. The effect of the centre·treatment variance component will, however, still cause an increase in the variance of the treatment differences.

Outlying centres

These can be determined using the shrunken centre and centre·treatment estimates. Shrinkage is greater for small centres, and therefore spurious outlying estimates will not be obtained for small centres.

Inference

This relates to the 'population' of possible centres from which those in the trial can be regarded as a random sample.

5.2.4 Centre effects random, centre·treatment effects omitted

This model is useful when local estimates are required, since smaller treatment standard errors are often obtained compared with a fixed effects model by recovering extra information from the centre error stratum.

Treatment estimates

These take account of differing centre sizes. Additional information on treatments is recovered from the centre error stratum.

Treatment standard errors

These are based on the residual (within-centre) variation. When the proportion of patients allocated to each treatment is the same across centres, the variances of treatment differences are given by the formula used in the first fixed effects model. When it is not, the variance is expected to be less, since extra information is recovered on treatment effects from the centre error stratum.

Outlying centres

These can be determined using the shrunken centre estimates. Shrinkage is greater for small centres, and therefore spurious outlying estimates are not obtained for small centres.

Inference

This relates to the only sampled centres. Only under the strong assumption that there is no centre·treatment interaction, inferences can be applied to the population of centres from which those in the trial can be considered as a sample.

5.3 Example: a multi-centre trial

We have already considered in some detail the analysis of a multi-centre trial of treatments for hypertension in Sections 1.3 and 2.5. Here, we discuss the interpretation of the results from these analyses in detail and consider estimates of treatment effects by centre.

Results from fixed and random effects' analyses of DBP are summarised in Table 5.1. An initial fixed effects model including centre·treatment effects was also fitted. However, overall treatment effects were not estimable in this model, because all treatments were not received at every centre (see Table 1.1 in Section 1.3). This model gave a non-significant p-value for centre·treatment

Table 5.1 Results from fixed and random effects analyses of diastolic blood pressure.

Model	Fixed effects	Random effects	Method
1	Baseline, treatment, centre		OLS
2	Baseline, treatment	Centre	REML
3	Baseline, treatment	Centre, treatment·centre	REML

		Treatment effects (SEs)		
Model	Baseline	A – B	A – C	B – C
1	0.22 (0.11)	1.20 (1.24)	2.99 (1.23)	1.79 (1.27)
2	0.22 (0.11)	1.03 (1.22)	2.98 (1.21)	1.95 (1.24)
3	0.28 (0.11)	1.29 (1.43)	2.93 (1.41)	1.64 (1.45)

	Variance components		
Model	Centre	Treatment·centre	Residual
1	–	–	71.9
2	7.82	–	70.9
3	6.46	4.10	68.4

effects ($p = 0.19$), and thus the usual 'fixed effects' approach would have been to remove centre·treatment effects to give Model 1.

The centre·treatment variance component is positive in Model 3, and this leads to an increase in the treatment standard errors over the fixed effects models (as indicated by the variance formulae given in Section 5.2). However, note that the baseline standard error is similar between the models. This is because baseline effects are estimated at the residual error level and not at the centre·treatment level. Results from Model 3 can be related to the potential population of centres. Since there are 29 centres, there are no problems arising from an inadequate number of DF for the variance components, and we can be confident in presenting these results if global inference is required.

The centre variance component is positive in Models 2 and 3, and therefore some information on treatments will be recovered from the centre error stratum. However, the standard errors in Model 2 are only slightly smaller than in Model 1, indicating that only a small amount of information has been recovered. Also, some of the improvement in the standard error may be due to the smaller residual variance that has resulted from the use of this model.

Plots of the centre and centre·treatment effects from Model 3 were used in Section 2.5 to assess the normality of the random effects and to check whether any centres were outlying. In addition, we can now take differences between the centre·treatment effects to calculate treatment effect estimates for each centre. In Model 3, these will be shrunken towards the overall treatment mean. We illustrate this by calculating the treatment difference A−C for just the first eight centres in the study. Unshrunken fixed effects estimates are also calculated for comparison using results from the initial model (fitting treatment, centre and centre·treatment effects as fixed).

		Treatment estimates (SE)	
Centre	**Number of patients**	**Fixed model**	**Random model**
1	39	3.81(3.34)	3.19 (2.94)
2	10	−5.67(6.80)	1.56 (3.40)
3	8	25.66(7.63)	6.05 (3.40)
4	12	−0.12(5.90)	2.42 (3.39)
5	11	14.12(7.21)	4.56 (3.40)
6	5	2.89(8.37)	3.03 (3.39)
7	18	7.38(4.82)	4.22 (3.31)
8	6	−4.68(8.33)	2.15 (3.39)

It can be seen that, in general, shrinkage is towards the overall treatment difference of 2.92 (although this is not the case for all centres, because the models make different adjustments for baseline effects). The relative shrinkage

(i.e. (fixed estimate−random estimate)/(fixed estimate)) is usually greatest for the smaller centres. The standard errors of the random effects estimates are smaller than those of the fixed effects estimates because the random effects model utilises information on the treatment effects in the full sample as well as information from the centre of interest. By contrast, the fixed effects standard errors do not utilise the full sample information and are larger because they are calculated using only information from the centre of interest. This also causes the fixed effects standard errors to vary greatly between the centres because they are directly related to the centre sizes. It is difficult to determine whether any of the centres are outlying using the fixed effects estimates because they need to be considered bearing in mind centre size. For example, at centre 3 a very large treatment difference is given by the fixed estimate, but the shrunken random estimate appears acceptable. We note that the standard errors of the random effects estimates have increased by around 20% from those reported in the first edition of this book, by use of the Kenward–Roger option.

SAS code and output

Variables

`centre` = centre number,
`treat` = treatment (A, B, C),
`patient` = patient number,
`dbp` = diastolic blood pressure at last attended visit,
`dbp1` = baseline diastolic blood pressure.

The SAS code to produce the main results is given at the end of Section 2.5. Here, we give the code for obtaining the shrunken and unshrunken treatment effects at the first eight centres. `PROC MIXED` is used first to fit Model 3. `ESTIMATE` statements are included to calculate the shrunken treatment differences at the first eight centres. Next, a fixed effects model is fitted, which again uses `ESTIMATE` statements to calculate (unshrunken) treatment differences. Two datasets, '`random`' and '`fixed`', are then created to extract and label the random and fixed effects estimates. The rest of the code listed is concerned with merging and printing the two sets of estimates.

```
PROC MIXED; CLASS centre treat;
TITLE 'RANDOM EFFECTS MODEL';
MODEL dbp = dbp1 treat/DDFM = KENWARDROGER;
RANDOM centre centre*treat;
ESTIMATE 'A-C,1' treat 1 0 -1; centre*treat 1 0 -1;
ESTIMATE 'A-C,2' treat 1 0 -1| centre*treat 0 0 0 1 0 -1;
ESTIMATE 'A-C,3' treat 1 0 -1| centre*treat
          0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,4' treat 1 0 -1| centre*treat
          0 0 0 0 0 0 0 0 0 1 0 -1;
```

```
ESTIMATE 'A-C,5' treat 1 0 -1| centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,6' treat 1 0 -1| centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,7' treat 1 0 -1| centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0-1;
ESTIMATE 'A-C,8' treat 1 0 -1| centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ODS OUTPUT ESTIMATES = random;

PROC MIXED; CLASS centre treat;
TITLE 'FIXED EFFECTS MODEL';
MODEL dbp = dbp1 treat centre centre*treat;
ESTIMATE 'A-C,1' treat 1 0 -1 centre*treat 1 0 -1;
ESTIMATE 'A-C,2' treat 1 0 -1 centre*treat 0 0 0 1 0 -1;
ESTIMATE 'A-C,3' treat 1 0 -1 centre*treat 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,4' treat 1 0 -1 centre*treat
         0 0 0 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,5' treat 1 0 -1 centre*treat
         0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,6' treat 1 0 -1 centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,7' treat 1 0 -1 centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ESTIMATE 'A-C,8' treat 1 0 -1 centre*treat
         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1;
ODS OUTPUT ESTIMATES = fixed;

DATA random; SET random;
centre=substr(label,5,2)*1;
est=estimate;
se=stderr;
KEEP centre est se;

DATA fixed; SET fixed;
centre=substr(label,5,2)*1;
estf=estimate;
sef=stderr;
KEEP centre estf sef;

* Summarise original dataset 'a' to obtain centre means;
PROC SORT DATA=a; BY centre;
PROC MEANS NOPRINT DATA=a; BY centre; VAR dbp; OUTPUT
OUT=freq N=freq;

DATA b; MERGE fixed random freq; BY centre;
IF centre<=8;
shrink=abs(estf-est);
```

```
PROC PRINT SPLIT='*' noobs; VAR centre freq estf sef est se;
LABEL centre='**CENTRE'
freq='PATIENTS*AT*CENTRE'
estf='FIXED*MODEL*ESTIMATE'
sef='FIXED*MODEL*SE'
est='RANDOM*MODEL*ESTIMATE'
se='RANDOM*MODEL*SE'
shrink='SHRINKAGE';
FORMAT estf est se sef 8.2;
```

RANDOM EFFECTS MODEL
The Mixed Procedure
Model Information

Data Set	WORK.A
Dependent Variable	dbp
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Prasad-Rao-Jeske-Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

Dimensions	
Covariance Parameters	3
Columns in X	5
Columns in Z	108
Subjects	1
Max Obs Per Subject	288

Number of Observations	
Number of Observations Read	288
Number of Observations Used	288
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	2072.30225900	
1	3	2055.64188178	0.00000322
2	1	2055.63936685	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
centre	6.4628
centre*treat	4.0962
Residual	68.3677

Fit Statistics

-2 Res Log Likelihood	2055.6
AIC (smaller is better)	2061.6
AICC (smaller is better)	2061.7
BIC (smaller is better)	2065.7

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	284	6.16	0.0137
treat	2	25	2.16	0.1364

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
A-C,1	3.1914	2.9446	6.68	1.08	0.3160
A-C,2	1.5596	3.4033	2.48	0.46	0.6838
A-C,3	6.0493	3.3976	2.31	1.78	0.1999
A-C,4	2.4208	3.3894	2.82	0.71	0.5297
A-C,5	4.5609	3.3985	2.43	1.34	0.2910
A-C,6	3.0296	3.3878	2.17	0.89	0.4591
A-C,7	4.2176	3.3132	3.55	1.27	0.2799
A-C,8	2.1511	3.3887	2.17	0.63	0.5859

FIXED EFFECTS MODEL

The Mixed Procedure

Model Information

Data Set	WORK.B
Dependent Variable	dbp
Covariance Structure	Diagonal
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Residual

Class Level Information

Class	Levels	Values
centre	29	1 2 3 4 5 6 7 8 9 11 12 13 14 15 18 23 24 25 26 27 29 30 31 32 35 36 37 40 41
treat	3	A B C

Dimensions

Covariance Parameters	1
Columns in X	113
Columns in Z	0
Subjects	1
Max Obs Per Subject	288

Number of Observations

Number of Observations Read	288
Number of Observations Used	288
Number of Observations Not Used	0

Covariance Parameter Estimates

Cov Parm	Estimate
Residual	69.2614

Fit Statistics

-2 Res Log Likelihood	1558.1
AIC (smaller is better)	1560.1
AICC (smaller is better)	1560.1
BIC (smaller is better)	1563.5

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	208	0.99	0.3198
treat	2	208	1.24	0.2905
centre	28	208	1.98	0.0038
centre*treat	48	208	1.20	0.1884

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
A-C,1	3.8081	3.3364	208	1.14	0.2550
A-C,2	-5.6720	6.8036	208	-0.83	0.4054
A-C,3	25.6559	7.6275	208	3.36	0.0009
A-C,4	-0.1189	5.8972	208	-0.02	0.9839
A-C,5	14.1230	7.2085	208	1.96	0.0514
A-C,6	2.8891	8.3700	208	0.35	0.7303
A-C,7	7.3811	4.8201	208	1.53	0.1272
A-C,8	-4.6825	8.3284	208	-0.56	0.5746

The table of shrunken and unshrunken treatment estimates is given within the main text.

5.4 Practical application and interpretation

In this section we consider some general points relating specifically to analyses of multi-centre data.

5.4.1 Plausibility of a centre·treatment interaction

One approach to analysing multi-centre trials of drug treatments works from the premise that it is not plausible for a treatment effect to vary across centres. If it does, it is deemed a fault with the study design. If a significant centre·treatment interaction is not detected, then the design is assumed to be sound and global inference is made from a model not allowing for any variation in treatment effects between centres. However, a drug effect can sometimes vary owing to differences in the centre populations even when they are defined within the constraints of the protocol. For example, one drug may work better on severely ill patients than another but less well on moderately ill patients. Thus, a centre containing more severely ill patients could produce larger treatment effects than centres containing a more even mixture of patients. For this reason, it is our belief that centre·treatment effects are always plausible, and that if global inference is required from a multi-centre trial, then the random effects model is likely to be the most appropriate.

Interactions are often even more plausible in trials not involving drugs. For example, in a trial of surgical techniques, one centre may have much more

expertise with one technique than another. In this type of trial, a random effects model should almost always be used in order to provide global inference.

5.4.2 Generalisation

Consideration of different interpretations of results from fixed effects and random effects analyses brings the issue of generalisation to the fore. Results are often generalised from the situation in which they were sampled to other situations. However, strictly speaking, results should only be generalised when the study sample has been taken at random from the whole population of interest. Since centres are rarely sampled at random, even the global results from a multi-centre trial cannot be formally generalised to the population of possible centres.

There are analogies, though, with a single-centre study. In such studies, patients are not usually selected at random from the potential population of patients available at the centre. However, results are usually seen as some indication of those expected in the future both in the same centre and elsewhere.

Thus, in practice, generalisation needs to be by degree and will always to some extent involve subjective judgements, for example of how well the centres (or patients) sampled represent their full potential populations. Multi-centre studies would usually be considered more generalisable than single-centre studies even though the centres are not randomly sampled, and even if a fixed effects model is employed.

5.4.3 Number of centres

The accuracy with which variance components are estimated is dependent on the number of centres used. If the centre·treatment variance component is inaccurate, then this will have a direct effect on the accuracy of the treatment standard errors (calculated as $2(\sigma^2/rc + \sigma^2_{ct}/c)$ when data are balanced). Thus, if a study uses only a few centres (say less than about five), σ^2_{ct} and hence the treatment effect standard error may not be accurate. In this situation a random effects analysis may be inadvisable, although it could used to provide a rough idea of the global treatment estimates. However, the main conclusions from the study should be based on local results obtained from a fixed effects model, usually omitting the centre·treatment interaction term.

5.4.4 Centre size

Sometimes, a trial contains several centres that stop participating in the trial after recruiting only a few patients. Since little information is available for measuring effects at such centres, a strategy that is often adopted is to combine such centres into one centre in the analysis. This has most use when a fixed effects model is used,

because the centre·treatment interaction can then be assessed more effectively. However, in a mixed model, it usually makes little difference whether such centres are combined together or fitted separately.

5.4.5 Negative variance components

If centre·treatment effects are retained but the centre variance component estimate is negative, then centre effects could be either removed altogether from the model or retained with their variance component constrained to zero. Although the same effect estimates will be obtained using either approach, the DF used by the significance tests will differ. The latter option of retaining the centre effect DF is perhaps the most satisfactory because centre·treatment effects are retained. If both centre and centre·treatment variance components are negative, then they can be removed from the model and the data analysed as a simple between-patient trial. Inference can still be made globally to the population of centres, since there is no variation in the treatment effect across centres.

5.4.6 Balance

In models including treatment·centre effects balance will only be achieved in the unlikely situation where there are an equal number of patients per treatment per centre (this condition is also required to achieve balance across random effects). Treatment mean estimates from either a fixed or random effects model will then equal the raw treatment means. In the more usual situations where there are unequal numbers of patients per treatment per centre, treatment means will differ between fixed effects and random effects models.

In models omitting centre·treatment effects, balance across random effects is achieved when treatments are allocated in equal proportions at each centre (even if the overall centre sizes vary). Treatment estimates are then the same, regardless of whether centre effects are taken as fixed or random. However, if treatments are not allocated in equal proportions at each centre, treatment estimates will differ between the models because information is combined from both the centre and residual error strata in the random effects model.

5.5 Sample size estimation

When designing a multi-centre trial with the intention of estimating global treatment effects, sample size estimates can be calculated in a way that takes into account variation of treatment effects between centres. Here, we will obtain sample sizes that can be used for trials, based on considering differences between pairs of treatments.

5.5.1 Normal data

In Section 5.2, the variance of the difference between a pair of treatments in a balanced dataset was given as

$$\text{var}(t_i - t_j) = 2(\sigma^2/rc + \sigma^2_{\text{ct}}/c),$$

where

r = number of patients per treatment per centre (replicates),
c = number of centres,
t_i = the ith treatment effect,
σ^2 = residual variance,
σ^2_{ct} = centre·treatment variance component.

Estimates for the number of centres (c) and number of patients per treatment per centre (r) can be obtained from the usual sample size estimation equation:

$$\Delta = (t_{\text{DF},1-\alpha/2} + t_{\text{DF},\beta}) \times \text{SE}(t_i - t_j),$$

where

α = significance,
β = power,
Δ = difference to be detected,
τ = number of treatments,
DF = $(c-1)\times(\tau-1)$, the centre·treatment DF.

One difficulty is that estimates of both the patients and centre·treatment variance components are required. Unless multi-centre data are available from a previous study, it is likely that only an estimate of the between-patient residual variance will be available. However, it may still be preferable to use the above formulae with a guessed value for the treatment·centre variance component, rather than assuming it as zero.

In this section, we consider the situation where an equal number of patients will be used per treatment per centre. Some inflation to the calculated sample sizes will be appropriate when there will be varying numbers per centre, but these calculations will provide a reasonable first 'ballpark' estimate. r is taken to be the average number of replicates, $\Sigma r_i/c$. There are three ways in which a sample size can be calculated:

1. ***Number of centres (c) specified*** This approach would be applicable if a decision had been made to use a specific number of centres. After substitution of the formula for $\text{SE}(t_i - t_j)$ in the sample size estimation equation, with some reorganisation we find that the number of patients per replicate (i.e. per treatment per centre) required is given by

$$r = \frac{2(t_{\text{DF},1-\alpha/2} + t_{\text{DF},\beta})^2\sigma^2}{c\Delta^2 - 2(t_{\text{DF},1-\alpha/2} + t_{DF,\beta})^2\sigma^2_{\text{ct}}}.$$

Therefore, $\tau \times r \times c$ patients are required in total. If this formula gives a negative value for r, then it is not possible to detect the specified difference with the required power unless more centres are used. Either c should be increased or, alternatively, the power could be decreased or Δ increased.

2. ***Number of patients per centre ($\tau \times r$) specified*** This approach might be appropriate if the duration of the trial is limited and there is only time to recruit a specified number of patients per centre. The number of centres required is given by

$$c = \frac{2(t_{\mathrm{DF},1-\alpha/2} + t_{\mathrm{DF},\beta})^2(\sigma^2 + r\sigma_{\mathrm{ct}}^2)}{r\Delta^2}.$$

Obviously, $\mathrm{DF} = (c-1) \times (\tau - 1)$ will not be known in advance. z-values from the normal distribution can be used instead of values from the t distribution to obtain an initial estimate of c. A more accurate value can then be calculated by using the DF obtained for this value of c in the above formula and re-estimating c. This can be repeated until convergence is obtained, but changes are usually minimal after the first iteration.

3. ***Neither number of centres nor average patients per centre specified*** In this situation, an optimal sample size can only be calculated by specifying the relative cost of sampling centres compared with sampling patients. The cost of sampling centres will depend on the type of centre being used. For example, the cost of a centre in an international study would be extremely expensive, but centres would be much cheaper in a study using local practitioners. The cost of sampling patients relates to the amount to be paid to the investigator per patient plus the cost of monitoring, validating and processing each patient's data. If we denote the relative cost by g, then the total cost is proportional to $c \times r \times \tau + c \times g$. This is minimised when

$$r = \sqrt{\frac{g\sigma^2}{\tau\sigma_{\mathrm{ct}}^2}}.$$

c is then obtained by substituting r into the formula given earlier:

$$c = \frac{2(t_{\mathrm{DF},1-\alpha/2} + t_{\mathrm{DF},\beta})^2(\sigma^2 + r\sigma_{\mathrm{ct}}^2)}{r\Delta^2}.$$

Sometimes, the values calculated might appear impracticable. For example, if the relative cost, g, of sampling a centre were set to be not much higher than that of sampling a patient (i.e. g close to one), then the number of centres estimated would likely be very high. In this situation, g has clearly been set too low and should be increased.

Example

We will calculate sample sizes for a new hypertension study to compare three treatments. Assuming that DBP is again the primary endpoint, the variance

components obtained in Section 2.5 for centre·treatment effects ($\sigma^2_{ct} = 4.10$) and the residual ($\sigma^2 = 68.4$) can be used to estimate sample sizes. A difference of 5 mmHg is to be detected at the 5% significance level with 90% power.

1. ***Number of centres specified*** It has been decided that the new study will use four centres. Therefore, the DF for the t distribution is $(4-1)\times(3-1)=6$, and we have

$$\begin{aligned} r &= \frac{(2\times(t_{6,0.975}+t_{6,0.90})^2\times 68.4)}{(4\times 5^2 - 2\times(t_{6,0.975}+t_{6,0.90})^2\times 4.10)} \\ &= \frac{(2\times(2.45+1.44)^2\times 68.4)}{(4\times 5^2 - 2\times(2.45+1.44)^2\times 4.10)} \\ &= -86.0. \end{aligned}$$

   Since r is negative, it is not possible to obtain the required power when only four centres are used. If the number of centres is increased to six, then DF $= (6-1)\times(3-1)=10$ and

$$\begin{aligned} r &= \frac{(2\times(t_{10,0.975}+t_{10,0.90})^2\times 68.4)}{(6\times 5^2 - 2\times(t_{10,0.975}+t_{10,0.90})^2\times 4.10)} \\ &= \frac{(2\times(2.23+1.37)^2\times 68.4)}{(6\times 5^2 - 2\times(2.23+1.37)^2\times 4.10)} \\ &= 40.5. \end{aligned}$$

   Thus, $41\times 3 = 123$ patients would be required in each of the six centres, and the total number of patients is $123\times 6 = 738$.

2. ***Number of patients per centre specified*** It has been decided that an average of only 15 patients per centre will be used so that the study can be completed quickly. Using $r=5$, we obtain an initial estimate of c using z statistics as

$$c = 2\times(1.96+1.28)^2\times(68.4+5\times 4.10)/(5\times 5^2) = 14.9.$$

   With 15 centres, the t distribution DF for use in the original formula would be $(15-1)\times(3-1)=28$. On recalculating using t_{28} statistics c becomes 16.1, which should be rounded up to 17. Therefore, 17 centres should be used with five patients per treatment group ($17\times 5\times 3 = 255$ patients in total).

3. ***Neither number of centres n or average patients per centre specified*** The study will use centres from different countries, and therefore the cost of sampling centres compared with patients is high. If we set the relative cost of sampling centres compared with sampling patients at $g=100$, then

$$r = \sqrt{[100\times 68.4/(3\times 4.10)]} = 23.6.$$

Rounding r to 24, we obtain an initial estimate of c:

$$c = 2 \times (1.96 + 1.28)^2 \times (68.4 + 24 \times 4.10)/(24 \times 5^2) = 5.8.$$

This value of c is rounded to six. As c is small, it should be recalculated more accurately using $\mathrm{DF} = (6-1) \times (3-1) = 10$ as

$$\begin{aligned} c &= 2 \times (t_{10,0.975} + t_{10,0.90})^2 \times (68.4 + 24 \times 4.10)/(24 \times 5^2) \\ &= 2 \times (2.23 + 1.37)^2 \times (68.4 + 24 \times 4.10)/(24 \times 5^2) = 7.21. \end{aligned}$$

However, we can be more precise, because the necessity to have an integer number of centres means that in moving from 7.21 centres to eight centres, our value for r can be recalculated. In this instance, it is reduced appreciably from 23.6 to 16.6. We can therefore obtain our desired power from 17 patients per treatment per centre using eight centres. We could also consider the alternative of using seven centres and recalculating r. This gives $r = 23.6$. We can therefore compare the cost of seven centres with 24 patients per treatment per centre, with the design with eight centres.

5.5.2 Binary data

Consideration of the sample size requirements in a mixed models situation is far more complicated with non-normal distributions than for the normal case. The relationships between variances on the natural scale and on the transformed scale are complex and do not readily fit into the sample size formulae we have met already. For thoroughly reliable power calculations it is necessary to undertake simulation studies, and software has been developed to facilitate this. One such example is MLPowSim Software Package (www.bristol.ac.uk/cmm/software/mlpowsim/mlpowsim-manual.pdf), (Browne, Lahi and Parker, 2009).

Simpler, more approximate, methods may also be helpful though. A general approach that has been advocated for determining sample size is to base the calculations on the standard sample size formulae for the fixed effects model and then to inflate this sample size by a so-called design effect (Snijders, 2005). This can be used when a similar study has been conducted previously using a mixed models analysis. If the standard errors of the treatment effect of interest are denoted by $\mathrm{SE_M}$ for the mixed model and $\mathrm{SE_F}$ for a corresponding fixed effects model, then the design effect is defined by $(\mathrm{SE_M}/\mathrm{SE_F})^2$. If N denotes the sample size from standard formulae, then $N\times$ design effect is the approximate required sample size for the mixed model. This approach has its limitations. We have seen in the previous section that it is possible to have situations where the centre·treatment interaction is so large that it is impossible to have the desired power, whatever inflation is given to the overall sample size. Nevertheless, for some situations, this approach may be helpful.

In general, we believe that it is more satisfactory to use the formulae provided for normal data in Section 5.2.1 with input parameters on the logit scale, to obtain an approximate sample size.

Example

We will again calculate sample sizes for a new hypertension study, but this time we will compare two treatments. However, now we assume (unrealistically!) that the incidence of the adverse event, cold feet, is the primary endpoint. A doubling in the proportion of cold feet is to be detected from 0.1 to 0.2, and thus the required difference in logits (log(odds(0.2))−log(odds(0.1))) is 0.81. The study should have sufficient power to detect a difference at the 5% significance level with 80% power.

We will use the approximate result that the variance(log(odds(p))) = $1/Np + 1/N(1-p)$ if we have a binomial sample of size N. In looking at the distributional variance of individuals, $N=1$ and so, at an average proportion of 0.15, the variance on the logit scale is $1/0.15 + 1/0.85 = 7.84$. We will use our results of an analysis of our earlier example, but this time without fitting a dispersion parameter, which the sample size calculations do not incorporate. The resultant centre·treatment variance component is 0.547.

We assume that there are no stipulations for the number of centres or patients and seek to design the cheapest trial. If we set the relative cost of using a centre compared to recruiting a patient at $g = 50$, then

$$r = \sqrt{\frac{50 \times 7.84}{2 \times 0.547}} = 18.9,$$

which we take as 19.

From this, we obtain the number of centres required as

$$c = \frac{2 \times 10.5 \times (7.84 + 19 \times 0.547)}{19 \times 0.81^2} = 30.7$$

Thus, 31 centres with 19 patients per treatment are required.

This may be unrealistic, but we should appreciate that these calculations are based on the strong assumption about the value of the centre·treatment variance component. It has great uncertainty attached to it, and we might wish to recalculate on the assumption that it has been overestimated.

Alternatively, we could calculate sample size for a fixed number of centres. If we assume 20 centres are to be used, the number of patients required per treatment per centre is

$$r = \frac{2 \times 10.5 \times 7.84}{20 \times 0.81^2 - 2 \times 10.5 \times 0.547} = 100.7$$

Thus, 20 centres with 101 patients per treatment are required. However, reducing the number of centres below 18 leads to a negative value for r, and again we may consider recalculating on the assumption that the centre·treatment variance component has been overestimated.

5.5.3 Categorical data

Sample size estimation is always difficult when the variable of interest is categorical. If there are more than about five categories, the formula for continuous data is likely to provide a reasonable approximation. In other situations, the best approach might be to partition the categories and use the formula for binary data.

Precision of sample size estimates

The use of sample size calculations gives very precise answers, but they are based on assumed values for variance components that may be quite imprecise and specified differences that may be somewhat arbitrary. We do not therefore recommend slavish adherence to the precise numbers obtained from the formulae. It is sensible to undertake a kind of sensitivity analysis to see the extent to which the sample size depends on the assumptions made, particularly for binary data. We believe that the correct use of sample size calculations is to obtain reliable ballpark figures.

5.6 Meta-analysis

This type of analysis is increasingly used to combine results from several clinical trials, which assess the same treatments in order to provide a more precise overall estimate of the treatment effects. When the original data are available, an identical hierarchical structure to the multi-centre trial arises with trials replacing centres. The implications of fitting trial and trial·treatment effects as fixed or random are then the same as in multi-centre analyses (see Section 5.2).

If treatment estimates are to relate only to the trials included, then *local* treatment estimates are obtained by fitting trial and treatment·trial effects as fixed (although in practice the trial·treatment interaction is usually removed if non-significant). If they are to relate more widely to the circumstances and locations sampled by the trials, *global* estimates can be obtained by fitting trial and treatment·trial effects as random. When this is done, the standard errors of treatment estimates are increased to reflect the heterogeneity across trials. Trial·treatment variance components are often relatively larger than centre·treatment variance components, because different protocols are used by the different trials. Thus, there are often more noticeable increases in the treatment standard errors in meta-analyses than in multi-centre trials.

As with multi-centre trials, taking trial effects as random has the theoretical advantage of increasing the accuracy of treatment estimates. This is because information from the trial error stratum is used in addition to that from the residual stratum. Sometimes, there are factors that differ at the trial level, which can help to explain differences in results between trials. For example, race or type of clinic may affect the treatment effect size. These variables can be included as covariates in a mixed model and may reduce the trial·treatment variability (and hence lead to more precise treatment estimates).

Outlying trials can be checked for, using the shrunken estimates of trial and trial·treatment effects. Since shrinkage is greater when there are fewer observations per trial, spurious outlying estimates caused by random variation are less likely to occur. However, this would not be the case in a fixed effects model where there is no shrinkage of the trial and trial·treatment estimates. Often, the estimates of treatment effects from individual trials are themselves of interest. The shrunken estimates that utilise information from all trials are more robust than estimates from a fixed effects model, although it has to be recognised that there may be difficulty in conveying the concept of shrunken estimates to a medical researcher!

Meta-analysis tends to be carried out most frequently with binomial data and most published work has related to this. Commonly, this is based on frequencies for the main outcome variable(s) as individual data is commonly unavailable. Although meta-analysis can be used with normal data, in practice individual trials are often adequate to achieve the desired power while this is not always the case with binary outcomes. Achieving adequate power is therefore less of a motivation with normal data. However, in a pharmaceutical company, it may still be advantageous to undertake a meta-analysis on normal data arising from a series of trials in the same drug programme. Also, systematic reviews are being conducted increasingly commonly for an expanding number of treatments and outcome variables. Readers wishing to learn more about meta-analysis are advised to consult Whitehead (2002) or visit the Cochrane Collaboration website (www.cochrane.org).

There is one important conceptual difference between the way trial effects are handled by most meta-analysts compared to the method described earlier. In practice, meta-analysis often takes place in two stages. In the first stage, treatment differences within trials are calculated. The second stage combines these contrasts in either a fixed effects or a random effects analysis, dependent on which is deemed appropriate. Thus, trial effects are implicitly treated as fixed effects, and only the trial·treatment effect is potentially considered as random. In contrast, as noted above, fitting both trial and trial·treatment effects as random allows the recovery of between trial information, thus breaking the meta-analysis principle of concurrent control. In practice, fitting the trial effect as fixed or random usually makes little difference to the conclusions.

5.7 Example: meta-analysis

This example considers meta-analysis data that are taken from Thompson and Pocock (1991). The data come from nine trials comparing a diuretic treatment with a control treatment in relation to the incidence of pre-eclampsia. The number of women with pre-eclampsia within each trial and treatment group is shown in Table 5.2.

Table 5.2 Frequencies of pre-eclampsia/numbers randomised in trials included in meta-analysis.

Trial	Diuretic	Control	Odds ratio
Weseley	14/131	14/136	1.04
Flowers	21/385	17/134	0.40
Menzies	14/57	24/48	0.33
Fallis	6/38	18/40	0.23
Cuadros	12/1011	35/760	0.25
Landesman	138/1370	175/1336	0.74
Krans	15/506	20/524	0.77
Tervila	6/108	2/103	2.97
Campbell	65/153	40/102	1.14
Total	291/3759	345/3183	0.72

5.7.1 Analyses

The five analysis models shown in Table 5.3 are considered. Models 1 and 2 were fitted using fixed effects models (GLMs). In these models, the data are analysed in binomial form (i.e. using the trial·treatment frequencies). However, identical

Table 5.3 Variance component and treatment odds ratio estimates.

Model	Method (data form)	Fixed effects	Random effects
1	GLM (binomial)	Treatment, trial, trial·treatment	–
2	GLM (binomial)	Treatment, trial	–
3	P-L (Bernoulli)	Treatment	Trial, trial·treatment
3(a)	P-L (binomial)	Treatment	Trial, trial·treatment
3(b)	MCMC (binomial)	Treatment	Trial, trial·treatment
4	P-L (Bernoulli)	Treatment, trial	Trial·treatment

	Variance components				
Model	Trial	Trial·treatment	Dispersion parameter	Treatment odds ratio (95% CI)	p – value
1			1^F	0.63 (0.47–0.84)	0.002
2			1^F	0.66 (0.56–0.79)	< 0.0001
3	1.42	0.16	0.98	0.60 (0.34–1.05)	0.07
3(a)	1.43	0.16	1^F	0.60 (0.34–1.05)	0.07
3(b)	1.50	0.18	1^F	0.60 (0.34–1.04)	0.07
4		0.16	0.99	0.60 (0.34–1.05)	0.07

Note: ^{F} = parameter is fixed.

results would have been obtained had the data been analysed in Bernoulli form (i.e. using one observation per patient). Model 3 is a random effects model, and the data are analysed in Bernoulli form using a GLMM. Using this form allows the 'shrunken' treatment effects at each trial to be estimated using `PROC GLIMMIX`. However, very similar results can be obtained by analysing the data as binomial frequencies, since there are no baseline values and no categories are uniform. We illustrate this by fitting Model 3(a) which is identical to Model 3, except that the data are in binomial form and the dispersion parameter is fixed at one (allowing variance at the residual level to be modelled by the trial·treatment variance component). Model 3(b) is the same random effects model as Model 3(a) but is fitted using MCMC (see Section 2.3.5). Model 4 has been added in this edition to show the results from the standard meta-analysis approach.

Odds ratios and confidence intervals are calculated by exponentiating the treatment difference estimates and their confidence intervals on the logit scales (see Section 3.3.9). In recent versions of `SAS`, this can be done using relevant options. The treatment effect is tested using asymptotic Wald Chi-squared tests in Models 1 and 2, and an F test in Models 3 and 3(a) (see Section 3.3.8 for details on GLM and GLMM significance testing). In Model 3(b), twice the probability of the treatment difference being greater than zero is taken to provide a 'Bayesian' p-value (see Section 2.3.3).

In Models 3 and 3(a), the Kenward–Roger method has been used to adjust for the fixed effects standard error bias. This has had the effect of both increasing the standard error and modifying the DF for the F test as compared with the results presented in the earlier editions of this book.

5.7.2 Results

The results are shown in Table 5.3. Results from Models 1 and 2 do not take account of any additional variation in the treatment effect between trials and, therefore, should be formally related only to the trials included. The trial·treatment interaction was highly significant in Model 1 ($p = 0.0006$), and this would cast doubt on any formal extrapolation of the results from these models. In Model 1, the overall treatment effect is calculated as an unweighted average of the treatment effects at each trial (see Section 5.2). Such an estimate is clearly inappropriate, since the trial sizes differ widely. This problem does not arise in Model 2, where centre·treatment effects are omitted. Note that in the above models, the 95% confidence intervals are based on exponentiating the treatment estimate $\pm 1.96 \times \text{SE}$ because of the asymptotic normality of the estimate.

Models 3, 3(a) and 3(b) take account of the extra variation in the treatment effect between trials by fitting trial and trial·treatment effects as random. In these models, we are assuming that the random effects are normally distributed and, as the variance components are estimated, the confidence intervals are more appropriately estimated for Models 3 and 3(a) as treatment estimate $\pm t \times \text{SE}$. The DF

for t are automatically taken to be those arising from the Kenward–Roger method when this option is specified in the `MODEL` statement. Model 3(a) gives very similar results to Model 3, indicating that it makes little difference here whether the data are analysed in Bernoulli or binomial form. The trial variance component is fairly large in all of the models, indicating that the overall incidence of pre-eclampsia varies greatly between trials. Thus, it is likely that quite different inclusion criteria were used for the trials or that pre-eclampsia was defined differently by the different practitioners. The positive trial·treatment components indicate some variation in the treatment effect across trials. This is reflected in the size of the treatment confidence intervals, which are wider than those in Models 1 and 2. The results from these analyses can be generalised with some confidence to the full population of pre-eclampsia sufferers.

The Bayesian analysis (Model 3(b)) gives almost identical treatment ORs and confidence intervals to the pseudo-likelihood analyses (Models 3 and 3(a)). The differences in variance component estimates between Models 3, 3(a) and 3(b) are not unexpected, since we have taken them to be the medians of the marginal posterior distributions, whereas in Model 3 the estimates are the values that maximise the pseudo-likelihood surface.

The use of fixed trial effects in Model 4 has not changed the estimates in Table 5.3 from those in Model 3(a) where the trial effect was random. Similarly, the corresponding shrunken estimates and standard errors in Table 5.4 would change by no more than one unit in the second decimal place (results not shown).

5.7.3 Treatment estimates in individual trials

Another advantage of using a random effects model is that shrunken estimates of the treatment effect at each trial can be obtained. Because shrinkage is greater when there are fewer observations per trial, any spurious outlying estimates

Table 5.4 Shrunken and unshrunken treatment effect estimates (standard errors) at each trial.

Trial	Shrunken (Model 3)	Unshrunken (Model 1)	Number of patients
Weseley	−0.14(0.37)	0.04(0.40)	267
Flowers	−0.82(0.33)	−0.92(0.34)	519
Menzies	−0.91(0.38)	−1.12(0.42)	105
Fallis	−1.04(0.45)	−1.47(0.55)	78
Cuadros	−1.16(0.30)	−1.39(0.34)	1771
Landesman	−0.31(0.12)	−0.30(0.12)	2766
Krans	−0.32(0.33)	−0.26(0.35)	1030
Tervila	0.07(0.52)	1.09(0.83)	211
Campbell	0.03(0.25)	0.14(0.26)	255

caused by random variation will not occur. The shrunken estimates also allow us to check whether results from any particular trial are outlying. If they are, it may be an indication that the trial was not suitable for inclusion in the meta-analysis. The shrunken estimates obtained from Model 3 are given in Table 5.4 along with the (unshrunken) GLM estimates obtained from Model 1 (on the logit scale). These estimates are all shrunken towards the overall treatment estimate of -0.51 ($\exp(-0.51) = 0.60$). The standard errors are smaller for the shrunken estimates because they utilise information from the whole sample, not just that from the individual trials. Greatest shrinkage occurs for the smallest trials. For example, the unshrunken estimate from Tervila appears extreme, but the shrunken estimate is much more reasonable and would not cause us to suspect the quality of the trial.

SAS code and output

Variables
`treat` = treatment,
`trial` = trial number,
`eclam` = number of women with pre-eclampsia,
`n` = number of women in treatment group at trial.

Model 1

```
DATA a; INPUT trial treat eclam n;
CARDS;

1 1 14 131
1 2 14 136
2 1 21 385
2 2 17 134
3 1 14 57
3 2 24 48
4 1 6 38
4 2 18 40
5 1 12 1011
5 2 35 760
6 1 138 1370
6 2 175 1336
7 1 15 506
7 2 20 524
8 1 6 108
8 2 2 103
9 1 65 153
9 2 40 102
;
PROC GENMOD; CLASS trial treat;
```

```
MODEL eclam/n=treat trial treat*trial / DIST=B TYPE3 WALD;
ESTIMATE 'overall' treat 1 -1 / EXP;
```

The output is not shown but has a similar form to that of Model 2.

Model 2

```
PROC GENMOD; CLASS trial treat;
MODEL eclam/n=treat trial/DIST=B TYPE3 WALD;
ESTIMATE 'overall' treat 1 -1 / EXP;
```

Model Information

Data Set	WORK.A
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	eclam
Response Variable (Trials)	n
Number of Observations Read	18
Number of Observations Used	18
Number of Events	636
Number of Trials	6942

Class Level Information

Class	Levels	Values
trial	9	1 2 3 4 5 6 7 8 9
treat	2	1 2

Parameter Information

Parameter	Effect	trial	treat
Prm1	Intercept		
Prm2	treat		1
Prm3	treat		2
Prm4	trial	1	
Prm5	trial	2	
Prm6	trial	3	
Prm7	trial	4	
Prm8	trial	5	
Prm9	trial	6	
Prm10	trial	7	
Prm11	trial	8	
Prm12	trial	9	

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	8	29.3761	3.6720
Scaled Deviance	8	29.3761	3.6720
Pearson Chi-Square	8	28.7996	3.6000
Scaled Pearson X2	8	28.7996	3.6000
Log Likelihood		-1877.3795	

Algorithm converged.

The deviance and Pearson Chi-square are measures of model fit and have similar roles to the residual sum of squares in normal data models.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-0.1137	0.1379	-0.3841	0.1566	0.68	0.4096
treat	1	1	-0.4104	0.0885	-0.5840	-0.2369	21.48	<.0001
treat	2	0	0.0000	0.0000	0.0000	0.0000	.67	.3267
trial	1	1	-1.8457	0.2380	-2.3122	-1.3792	60.14	<.0001
trial	2	1	-2.1344	0.2118	-2.5496	-1.7192	101.52	<.0001
trial	3	1	-0.2363	0.2409	-0.7084	0.2359	0.96	0.3267
trial	4	1	-0.5054	0.2779	-1.0500	0.0393	3.31	0.0690
trial	5	1	-3.2740	0.1958	-3.6577	-2.8903	279.67	<.0001
trial	6	1	-1.7287	0.1420	-2.0071	-1.4503	148.15	<.0001
trial	7	1	-3.0515	0.2150	-3.4730	-2.6300	201.36	<.0001
trial	8	1	-2.9293	0.3830	-3.6800	-2.1787	58.50	<.0001
trial	9	0	0.0000	0.0000	0.0000	0.0000	.67	.3267
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Wald Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
treat	1	21.48	<.0001
trial	8	444.03	<.0001

Asymptotic Wald Chi-squared tests are performed for each fixed effects parameter. These tests should be interpreted cautiously in small datasets.

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
overall	-0.4104	0.0885	0.05	-0.5840	-0.2369	21.48	<.0001
Exp(overall)	0.6634	0.0587	0.05	0.5577	0.7891		

The final line of the above output provides the estimate of the OR and accompanying 95% confidence limits, and is generated by the `EXP` option in the `ESTIMATE` statement. In SAS Version 9.3, there have been some changes in detail from the above output, which is from an earlier version, but the results are identical.

Model 3

Variables
`outcome` = success of treatment (1/0),
`trial` = trial number,
`treat` = treatment group,
`freq` = number of women with this outcome.

```
PROC GLIMMIX;
CLASS trial treat;
NLOPTIONS MAXITER=50;
FREQ freq;
MODEL outcome=treat / DIST=B SOLUTION DDFM=KENWARDROGER;
RANDOM trial treat*trial;
RANDOM _RESIDUAL_;
LSMEANS treat / DIFF PDIFF OR CL;
```

This code analyses the data in Bernoulli form so that there is a 0/1 observation corresponding to each patient in the trial. The `FREQ` statement is used here to indicate the number of times each observation in our dataset is repeated – this saves setting up a large dataset containing 6942 observations, many of which would be the same. In this example, more iterations than the default were required to obtain convergence. This was achieved with the `NLOPTIONS` statement

```
                Fit Statistics
-2 Res Log Pseudo-Likelihood     39379.40
Generalized Chi-Square            6766.61
Gener. Chi-Square / DF               0.98

      Covariance Parameter Estimates
                                Standard
   Cov Parm        Estimate       Error
   trial             1.4246      0.7754
   trial*treat       0.1621      0.1258
   Residual (VC)     0.9750     0.01657
```

Note that the residual covariance parameter corresponds to the dispersion parameter.

Solutions for Fixed Effects

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		−1.8137	0.4294	9.122	−4.22	0.0022
treat	1	−0.5105	0.2293	6.176	−2.23	0.0663
treat	2	0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	6.176	4.96	0.0663

treat Least Squares Means

treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1	-2.3242	0.4307	9.218	-5.40	0.0004	0.05	-3.2949	-1.3535
2	-1.8137	0.4294	9.122	-4.22	0.0022	0.05	-2.7832	-0.8442

treat Least Squares Means

treat	Odds Ratio	Lower Odds Ratio	Upper Odds Ratio
1	0.098	0.037	0.258
2	0.163	0.062	0.430

Differences of treat Least Squares Means

treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
1	2	-0.5105	0.2293	6.176	-2.23	0.0663	0.05

Differences of treat Least Squares Means

treat	_treat	Lower	Upper	Odds Ratio	Lower Odds Ratio	Upper Odds Ratio
1	2	-1.0677	0.04671	0.600	0.344	1.048

Model 3(a)

Variables
As in Models 1 and 2.

```
PROC GLIMMIX;
CLASS trial treat;
MODEL eclam/n=treat /DIST=B SOLUTION DDFM=KENWARDROGER;
RANDOM trial trial*treat;
LSMEANS treat / DIFF PDIFF OR CL;
```

Fit Statistics

-2 Res Log Pseudo-Likelihood	48.30
Generalized Chi-Square	17.00
Gener. Chi-Square / DF	1.06

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
trial	1.4250	0.7753
trial*treat	0.1593	0.1249

Solutions for Fixed Effects

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		-1.8133	0.4293	9.113	-4.22	0.0022
treat	1	-0.5105	0.2287	6.158	-2.23	0.0659
treat	2	0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	6.158	4.98	0.0659

treat Least Squares Means

treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1	-2.3238	0.4306	9.212	-5.40	0.0004	0.05	-3.2944	-1.3531
2	-1.8133	0.4293	9.113	-4.22	0.0022	0.05	-2.7827	-0.8440

Differences of treat Least Squares Means

treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
1	2	-0.5105	0.2287	6.158	-2.23	0.0659	0.05

```
           Differences of treat Least Squares Means
                                        Odds       Lower       Upper
 treat  _treat    Lower    Upper  Ratio  Odds Ratio  Odds Ratio
 1           2  -1.0666  0.04568  0.600       0.344       1.047
```

Model 3(b)

```
PROC MCMC OUTPOST=post3b NMC=2000000 THIN=10
  SEED=7899;
PARMS alpha0 alpha1 v1 v2;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
RANDOM b_trial ~ NORMAL(0, VAR = v1) SUBJECT=trial;
RANDOM b_trial_treat ~ NORMAL(0, VAR = v2) SUBJECT=trial_treat;
mu = alpha0 + alpha1*treat + b_trial + b_trial_treat;
expected = LOGISTIC(mu);
MODEL eclam ~ BINOMIAL(n,expected);

* obtain treatment effect p-value;
DATA p1; SET post3b;
a_b=alpha1;  * treat effect;
* define indicator variables for whether the sampled
   differences are greater than or less than zero;
IF a_b<0 THEN a_b0=1; ELSE a_b0=0;

PROC MEANS NOPRINT DATA=p1; VAR a_b a_b0;
OUTPUT OUT=p2 SUM=dum a_b0_n N=samples mean=a_b_mean
  std=a_b_std;

DATA p3; SET p2;
a_b0_p=a_b0_n/samples;
IF a_b0_p<0.5 THEN a_b_p=a_b0_p*2; ELSE a_b_p=(1-a_b0_p)*2;
PROC PRINT NOOBS DATA=p3; VAR a_b_p;
TITLE 'p-values for treatment comparison';
```

```
                        Posterior Summaries
                             Standard          Percentiles
 Parameter      N    Mean Deviation     25%         50%     75%
 alpha0    200000 -1.8507    0.4951 -2.1585     -1.8504 -1.5376
 alpha1    200000 -0.5073    0.2727 -0.6681     -0.5089 -0.3511
 v1        200000  1.8564    1.4335  1.0142      1.4953  2.2530
 v2        200000  0.2639    0.3036  0.0994      0.1765  0.3124
```

```
                    Posterior Intervals

Parameter  Alpha    Equal-Tail  Interval       HPD    Interval

alpha0     0.050       -2.8547   -0.8644   -2.8448     -0.8562
alpha1     0.050       -1.0424    0.0498   -1.0438      0.0481
v1         0.050        0.4092    5.4249   0.00347      4.3125
v2         0.050        0.0281    1.0709   0.00468      0.7733

          p-value for treatment comparison

                          a_b_p

                        0.06617
```

Model 4

The code for Model 4 is very similar to Model 3(a), differing only in change from the trial effect being fixed rather than random. The output is not shown.

```
PROC GLIMMIX DATA=a;
CLASS trial treat;
MODEL eclam/n=treat trial /DIST=B SOLUTION DDFM=KENWARDROGER;
RANDOM trial*treat;
LSMEANS treat / DIFF PDIFF OR CL;
```

Estimating treatment effect by trial from Models 1 and 3

Model 1

The following `ESTIMATE` statements are added to the earlier code for Model 1.

```
ESTIMATE 'c1' treat 1 -1 trial*treat 1 -1;
ESTIMATE 'c2' treat 1 -1 trial*treat 0 0 1 -1;
ESTIMATE 'c3' treat 1 -1 trial*treat 0 0 0 0 1 -1;
ESTIMATE 'c4' treat 1 -1 trial*treat 0 0 0 0 0 0 1 -1;
ESTIMATE 'c5' treat 1 -1 trial*treat 0 0 0 0 0 0 0 0 1 -1;
ESTIMATE 'c6' treat 1 -1 trial*treat 0 0 0 0 0 0 0 0 0 0 1 -1;
ESTIMATE 'c7' treat 1 -1 trial*treat 0 0 0 0 0 0 0 0 0 0 0 0
  1 -1;
ESTIMATE 'c8' treat 1 -1 trial*treat 0 0 0 0 0 0 0 0 0 0 0 0
  0 0 1 -1;
ESTIMATE 'c9' treat 1 -1 trial*treat 0 0 0 0 0 0 0 0 0 0 0 0
  0 0 0 0 1 -1;
```

Alternatively, the same estimates could have been obtained using the following `SLICE` statement instead of the multiple `ESTIMATE` statements.

```
SLICE trial*treat/ SLICEBY=trial DIFF;
```

Model 3

The following `ESTIMATE` statements are added to the earlier code for Model 3. Note the different form of the `ESTIMATE` statements because in this model `trial*treat` is a random effect, whereas in Model 1 it was a fixed effect.

```
ESTIMATE 'overall' treat 1 -1;
ESTIMATE 'c1' treat 1 -1 | trial*treat 1 -1;
ESTIMATE 'c2' treat 1 -1 | trial*treat 0 0 1 -1;
ESTIMATE 'c3' treat 1 -1 | trial*treat 0 0 0 0 1 -1;
ESTIMATE 'c4' treat 1 -1 | trial*treat 0 0 0 0 0 0 1 -1;
ESTIMATE 'c5' treat 1 -1 | trial*treat 0 0 0 0 0 0 0 0 1 -1;
ESTIMATE 'c6' treat 1 -1 | trial*treat 0 0 0 0 0 0 0 0 0 0
  1 -1;
ESTIMATE 'c7' treat 1 -1 | trial*treat 0 0 0 0 0 0 0 0 0 0
  0 0 1 -1;
ESTIMATE 'c8' treat 1 -1 | trial*treat 0 0 0 0 0 0 0 0 0 0
  0 0 0 0 1 -1;
ESTIMATE 'c9' treat 1 -1 | trial*treat 0 0 0 0 0 0 0 0 0 0
  0 0 0 0 0 0 1 -1;
```

The estimates produced by these analyses are given in Table 5.4.

6

Repeated measures data

In this chapter, we consider mixed models approaches to analysing repeated measures data. An introduction to repeated measures data and its analysis is given in Section 6.1. In Section 6.2, covariance pattern models are described, and two worked examples follow in Sections 6.3 and 6.4. Random coefficients models are described in Section 6.5, which is followed by a worked example in Section 6.6. Methods for sample size estimation are introduced in Section 6.7. In this chapter, we will just be considering simple designs where the repeated measures are defined on a single time scale. Occasionally, designs have a more complex pattern of repeated measurements; for example, repeated measurements may be taken within each of several visits. This design will be considered in Section 8.1.

6.1 Introduction

Any dataset in which subjects are measured repeatedly over time can be described as repeated measures data. Repeated measurements can be made either at predetermined intervals (e.g. at fortnightly visits or at specified times following a drug dose) or in an uncontrolled fashion so that there are variable intervals between the repeated measurements. The type of analysis model chosen will depend on whether the intervals are fixed and, if so, whether they are constant.

6.1.1 Reasons for repeated measurements

There are many reasons for collecting repeated measures data. Some examples are as follows:

- To ensure that a treatment is effective over a specified period. Often, this will be done using a carefully planned trial with fixed timings for visits.

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

- To monitor safety aspects of the treatment over a specified period (repeated efficacy measurements are then incidental).
- To see how long a single dose of a drug takes to become effective by measuring drug concentration or a physiological response at fixed intervals (often over 24 h). In this situation, repeated measurements are taken at a single visit.
- Repeated observations are sometimes inherent in the measurement itself; for example, blood pressure monitors can take measurements as frequently as every 10 s.
- To monitor particular groups of patients over time. Often, these are retrospective studies where repeated measurements have been recorded in an unplanned fashion. For example, repeated observations on patients with a particular disease may be available from a hospital clinic.

6.1.2 Analysis objectives

The objectives in analysing repeated measures data will differ depending on the purpose of the study. Ideally, they should be clarified at the design stage. Some examples of common objectives are as follows:

- To measure the average treatment effect over time.
- To assess treatment effects at each time point and to test whether treatment interacts with time.
- To assess specific features of the treatment response profile, for example, area under the curve (AUC), maximum or minimum value and time to the maximum value.
- To identify any covariance patterns in the repeated measurements.
- To determine a suitable model to describe the relationship of a measurement with time.

Before considering mixed models, we first review some fixed effects approaches to analysing repeated measures data that are sometimes satisfactory.

6.1.3 Fixed effects approaches

It is important that any analysis of a parallel group study compares treatment effects against a background of between-patient variation. This is because treatment effects are contained within patient effects. Each of these fixed effects approaches has the potential to compare treatments in this way.

Analysis of mean response over time

This method is satisfactory when the overall treatment effect is of interest, the times are fixed and there are no missing data. However, it does not give any

information on whether the treatment effect changes over time. When there are missing data, the analysis is only likely to be satisfactory if the response variable does not change with time.

Separate analyses at each time point

Separate analyses are carried out to compare treatments at each time point. Treatment standard errors are then correctly estimated at the between-patient level. One of many drawbacks to this approach is that repeated testing is taking place and therefore a significant treatment effect is more likely to occur at some time point by chance. In addition, the tests will be correlated. There may also be problems of interpretation if a treatment effect is significant at some time points but not at others. Another, admittedly less important, disadvantage is that the treatment standard errors will be less accurate, since they are based only on the observations at one time point, rather than using data from all time points. This analysis strategy is often observed in medical journals, but it is a strategy that should be discouraged.

Analyses of response features

Features summarising each patient's response profile (e.g. AUC, minimum or maximum value and time to maximum value) can be analysed. This approach is satisfactory if these summary features are of particular interest and if there is not a great deal of missing data. When there are missing data, it may not be possible to obtain a satisfactory estimate of some features (e.g. an AUC estimate would be biased if some of the observations were missing, although it is possible that an approach such as interpolation would help overcome this; the maximum value might be unrepresentative if observations around the true maximum were missing). If summary features are used, then restraint should be exercised in their selection to avoid problems of multiple testing.

Analysis of raw data fitting patient effects as fixed

In this model, patient, treatment, time and treatment·time effects can be fitted as fixed effects. However, treatment standard errors should not be obtained from the residual mean square, since this represents 'within-patient' variation. When there are no missing data, standard errors based on between-patient variation can be calculated manually using the patient sum of squares in the ANOVA table (note that many packages such as SAS (`PROC GLM`) will, by default, calculate treatment standard errors from the residual mean square). When there are no missing data, this model will give identical results to an equivalent random effects model fitting patients as random. It also has the advantage of being able to assess the treatment effects over time. However, it is not appropriate when there are missing data unless special adjustments are made.

6.1.4 Mixed models approaches

Mixed models have the following advantages:

- A single model can be used to estimate overall treatment effects and to estimate treatment effects at each time point. Treatment effects are correctly compared against a background of between-patient variation. There is no need to calculate mean values across all time points (to obtain the overall treatment effects) or to analyse time points separately (to obtain treatment effects at each time point). Standard errors for treatment effects at individual time points are calculated using information from all time points and are therefore more robust than standard errors calculated from separate time points.
- The presence of missing data causes no problems, provided they can be assumed missing at random.
- The covariance pattern of the repeated measurements can be determined and taken account of (e.g. the model can tell us whether the measurements across all time points have a constant correlation or whether the pattern of correlations is more complex and varies with time).

There are several ways a mixed model can be used to analyse repeated measures data. The simplest approach is to use a random effects model with patient effects fitted as random. This will allow for a constant correlation between all observations on the same patient. However, often, the correlation between observations on the same patient is not constant. For example, correlation may decrease as visits become more widely separated in time. A covariance pattern model can be used to allow for this or, alternatively, for a more complex pattern of correlation. These models are considered in Section 6.2. When the relationship of the response variable with time is of interest, a random coefficients model is appropriate. Regression slopes or curves are fitted for each patient, and the regression coefficients are allowed to vary randomly between the patients. These models are considered in Section 6.5.

6.2 Covariance pattern models

The basic structure of covariance pattern models has been described in Section 2.1.5. In this section, we consider their use for analysing repeated measures data in more detail and describe some more complex types of covariance pattern.

As described in Section 2.1.5, in covariance pattern models, the covariance structure is defined directly by specifying a covariance pattern rather than by using random effects. Observations within each category of a chosen *blocking effect* (e.g. patients) are assumed to have a specific pattern of covariance, which is defined across a time effect such as period or visit. For example, in a repeated measures trial, a pattern across periods could be specified for covariances between

observations occurring on the same patients. The covariance pattern is defined within the residual matrix, **R**. This matrix is blocked by patients so that only observations on the same patient are correlated. **R** can be written as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{R_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{R_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R_4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R_5} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R_6} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R_7} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{pmatrix}.$$

The $\mathbf{R_i}$ are submatrices of covariances corresponding to each patient and have dimension equal to the number of observations occurring on each patient. The **0**'s represent matrix blocks of zeros denoting zero correlation between observations on different patients. We will now consider different ways to define covariance patterns in the $\mathbf{R_i}$ matrix blocks.

6.2.1 Covariance patterns

A large selection of covariance patterns is available for use in mixed models. Most of the patterns are dependent on measurements being taken at fixed times, and some are also easier to justify when the observations are evenly spaced. There are also patterns where covariances are based on the exact value of time (rather than, say, visit number), and these are most useful in situations where the time intervals are irregular. Some examples of covariance patterns will be given. Still further possible types of covariance patterns can be found in the `SAS PROC MIXED` documentation.

Simple covariance patterns

Some simple covariance patterns for the $\mathbf{R_i}$ matrices for a trial with four time points are shown. In the general pattern (i), sometimes also referred to as 'unstructured', the variances of responses, σ_i^2, differ for each time period i, and the covariances, θ_{jk}, differ between each pair of periods j and k. For the first-order autoregressive model (ii), the variances are equal and the covariances decrease exponentially depending on their separation $|\, j-k\,|$, so $\theta_{jk} = \rho^{|j-k|}\sigma^2$. This is sometimes an appropriate model when periods are evenly spaced. It can then be seen as a 'natural' model from a time series viewpoint. However, it may be justified empirically in circumstances where the observations are not evenly spaced. For example, in monitoring the acute effect of drugs, it is common to take

measurements at short intervals soon after drug administration, when the level of observations may be changing rapidly, with increasingly separated intervals later on as observations change more slowly. Under such circumstances, adjoining observations may well show similar covariances, despite unequal periods, with exponentially decreasing covariances for increasingly separated observation numbers.

For the compound symmetry covariance model (iii), all covariances are equal. The Toeplitz model (iv) uses a separate covariance for each level of separation between the time points. This model is also known as the general autoregressive model.

(i) General

$$\mathbf{R}_i = \begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{12} & \sigma_2^2 & \theta_{23} & \theta_{24} \\ \theta_{13} & \theta_{23} & \sigma_3^2 & \theta_{34} \\ \theta_{14} & \theta_{24} & \theta_{34} & \sigma_4^2 \end{pmatrix}.$$

(ii) First-order autoregressive

$$\mathbf{R}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}.$$

(iii) Compound symmetry

$$\mathbf{R}_i = \begin{pmatrix} \sigma^2 & \theta & \theta & \theta \\ \theta & \sigma^2 & \theta & \theta \\ \theta & \theta & \sigma^2 & \theta \\ \theta & \theta & \theta & \sigma^2 \end{pmatrix}.$$

(iv) Toeplitz

$$\mathbf{R}_i = \begin{pmatrix} \sigma^2 & \theta_1 & \theta_2 & \theta_3 \\ \theta_1 & \sigma^2 & \theta_1 & \theta_2 \\ \theta_2 & \theta_1 & \sigma^2 & \theta_1 \\ \theta_3 & \theta_2 & \theta_1 & \sigma^2 \end{pmatrix}.$$

Before software for fitting covariance patterns was readily available, repeated measures analyses were often performed either using a random effects model or by fitting a multivariate normal distribution to the repeated measurements. A random effects model gives identical results to a compound symmetry pattern model (iii) (provided the patient variance component is not negative). Equality of the correlation terms in the compound symmetry structure was often assessed using a test of sphericity (e.g. Greenhouse and Geisser, 1959). However, if a lack of sphericity was found, there were limited alternative analyses available. If the data were complete, then fitting a multivariate normal distribution would give the same results as using a general pattern (i). This model could require a lot

of covariance parameters. In addition, if there were missing data, then fitting a multivariate normal distribution would not be satisfactory, as most packages cause all information on patients with incomplete data to be lost. Covariance pattern models overcome these limitations by providing a flexible choice of covariance patterns, which can be fitted to either complete or incomplete data.

Different variances for each time point

Sometimes, variability in a measurement will differ between the time points. This was allowed for in the general covariance pattern (i). Some additional patterns allowing for differing variances are given in the following section. In pattern (v), time points have different variances, but observations on the same patient are uncorrelated. This should only be used if preliminary analyses with more parameterised patterns indicate a lack of correlation between the repeated observations. Patterns (vi)–(viii) have similar forms to the autoregressive, compound symmetry and Toeplitz patterns, except that different variances for each time point are now used.

(v) Heterogeneous uncorrelated

$$\mathbf{R}_i = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix}.$$

(vi) Heterogeneous compound symmetry

$$\mathbf{R}_i = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho\sigma_1\sigma_4 & \rho\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}.$$

(vii) Heterogeneous first-order autoregressive

$$\mathbf{R}_i = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho^3\sigma_1\sigma_4 & \rho^2\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}.$$

(viii) Heterogeneous Toeplitz

$$\mathbf{R}_i = \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 & \rho_3\sigma_1\sigma_4 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \rho_1\sigma_2\sigma_3 & \rho_2\sigma_2\sigma_4 \\ \rho_2\sigma_1\sigma_3 & \rho_1\sigma_2\sigma_3 & \sigma_3^2 & \rho_1\sigma_3\sigma_4 \\ \rho_3\sigma_1\sigma_4 & \rho_2\sigma_2\sigma_4 & \rho_1\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}.$$

Separate covariance patterns for each treatment group

Sometimes, measurements on different treatments will have different variances and covariances. For example, it may be the case that measurements are more variable on an active treatment than on a placebo. This can be allowed for by using separate sets of covariance parameters for each treatment group. For example, if the first three patients in a trial received treatments A, B and A and were each measured at three time points, then the **R** matrix for these patients with separate *compound symmetry* structures for each treatment would be

$$\mathbf{R} = \begin{pmatrix} \sigma^2_A & \theta_A & \theta_A & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_A & \sigma^2_A & \theta_A & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_A & \theta_A & \sigma^2_A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_B & \theta_B & \theta_B & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_B & \sigma^2_B & \theta_B & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_B & \theta_B & \sigma^2_B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_A & \theta_A & \theta_A \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_A & \sigma^2_A & \theta_A \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_A & \theta_A & \sigma^2_A \end{pmatrix}.$$

Alternatively, if a *general structure* was used for each treatment group, then the **R** matrix would be

$$\mathbf{R} = \begin{pmatrix} \sigma^2_{A,1} & \theta_{A,12} & \theta_{A,13} & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{A,12} & \sigma^2_{A,2} & \theta_{A,23} & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{A,13} & \theta_{A,23} & \sigma^2_{A,3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{B,1} & \theta_{B,12} & \theta_{B,13} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{B,12} & \sigma^2_{B,2} & \theta_{B,23} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{B,13} & \theta_{B,23} & \sigma^2_{B,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{A,1} & \theta_{A,12} & \theta_{A,13} \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{A,12} & \sigma^2_{A,2} & \theta_{A,23} \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{A,13} & \theta_{A,23} & \sigma^2_{A,3} \end{pmatrix}.$$

If most of the covariances were small or negative, then observations on the same patient could be made *uncorrelated*, while different variances were still allowed for each treatment:

$$\mathbf{R} = \begin{pmatrix} \sigma^2_A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_A & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_B & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_B & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_A & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_A & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_A \end{pmatrix}.$$

Banded covariances

It will sometimes be apparent from the covariance parameter estimates that the correlation between widely separated observations is negligible. In this situation, it may be appropriate to 'band' the $\mathbf{R_i}$ matrices by setting correlations between the observations that are widely separated in time to zero. For example, a general covariance pattern with band size 3 would be

$$\mathbf{R_i} = \begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & 0 \\ \theta_{12} & \sigma_2^2 & \theta_{23} & \theta_{24} \\ \theta_{13} & \theta_{23} & \sigma_3^2 & \theta_{34} \\ 0 & \theta_{24} & \theta_{34} & \sigma_4^2 \end{pmatrix}.$$

Banding can be done for any covariance pattern and has the advantage of reducing the number of covariance parameters that need to be fitted. It is therefore of greatest use for trials with a large number of time points.

Covariance patterns defined using time as a continuous measure

A covariance pattern can be defined according to the exact separation of observations in time. This is the only type of pattern apart from compound symmetry that is appropriate when time points do not occur at predetermined intervals. However, it can also be useful for time points that are fixed but unevenly separated or when there is some flexibility in the timing of fixed interval time points (e.g. an interval of 12–16 days may be allowed for a nominal 2-week interval). There are many ways to define covariances from the time interval. Two examples are:

$$r_{ijk} = \sigma^2 \rho^{d_{ijk}} \quad \text{(power)},$$

$$r_{ijk} = \sigma^2 \exp(d_{ijk}^2/\rho^2) \quad \text{(Gaussian)},$$

where

r_{ijk} = covariance for observations j and k on patient i,
d_{ijk} = distance (usually time) between observations j and k on patient i.

Both these structures cause the covariance to decrease exponentially with the time interval between pairs of observations on the same patient.

6.2.2 Choice of covariance pattern

There are many covariance patterns available, and choosing the most appropriate one is not always easy. It is ideal to select the pattern that best fits the true covariance of data. As well as providing appropriate standard errors for fixed effects estimates, this can yield additional information about the action of treatments or the phenomenon being studied. However, as more covariance parameters are included in the pattern, the chances of overfitting increase

(i.e. the covariance pattern matches the observed pattern but may not be the true pattern). The likelihood (or quasi-likelihood) statistic is a basic measure of model fit, but it will increase as more covariance parameters are added and so therefore cannot be used directly for making comparisons. There are two alternative approaches to making a model choice. One is to compare models based on measures of fit that are adjusted for the number of covariance parameters. Another is to use likelihood ratio tests to find whether additional parameters cause a statistically significant improvement in the model. Our preference is for the second approach, in which covariance parameters are only included if they are proved to be necessary.

Measures of model fit

The likelihood statistic is expected to become larger as more parameters are included in the model. The two statistics described are based on the likelihood but make allowance for the number of covariance parameters fitted. They can be used to make direct comparisons between models that fit the same fixed effects. Akaike's information criterion (AIC) (Akaike, 1974) is given by

$$\text{AIC} = \log(L) - q,$$

where q is the number of covariance parameters. Schwarz's information criterion (SIC) (Schwarz, 1978) takes into account the number of fixed effects, p, the number of observations, N and the number of covariance parameters, q. It is given by

$$\text{SIC} = \log(L) - (q \log(N - p))/2.$$

Models with larger values of AIC and SIC denote better fits. However, it is unclear to us which criterion is preferable. Both the criteria are calculated within `PROC MIXED`, although the figures given correspond to twice those shown here and with the sign reversed.

We anticipate that these criteria will also be approximately satisfactory when based on the quasi-likelihood statistics arising out of generalised linear mixed models (GLMMs) and categorical mixed models.

Statistical comparisons between models

Models can be compared statistically using likelihood ratio tests, provided that they fit the same fixed effects and their covariance parameters are nested. Nesting of covariance parameters occurs when the covariance parameters in the simpler model can be obtained by restricting some of the parameters in the more complex model (e.g. a compound symmetry pattern is nested within a Toeplitz pattern, but it is not nested within a first-order autoregressive pattern). The likelihood ratio test statistic is given by

$$2(\log(L_1) - \log(L_2)) \sim \chi^2_{\text{DF}},$$

where DF = difference in number of covariance parameters fitted.

If the covariance parameters in the models compared are not nested, statistical comparison using a likelihood ratio test is not valid. In this situation, comparisons of each model with a simpler model that is nested within both models could be made and the model giving the most significant improvement selected. Alternatively, Akaike's or Schwarz's criteria could be used.

Again, we anticipate that these tests will be approximately satisfactory when based on the quasi-likelihood or pseudo-likelihood statistics arising out of GLMMs and categorical mixed models but are unaware of formal justification for this.

Which covariance patterns to consider?

It is not usually practical to test a large number of covariance patterns in a single application. A safe strategy would be to start with simple patterns such as the compound symmetry or first-order autoregressive. A general pattern could then be used to give some indication of whether any other more complex patterns are likely to be appropriate. These more complex patterns can be tested and should be accepted only if they lead to a significant improvement in the likelihood. Particularly complex patterns are only likely to be significant in larger datasets where more information on the true covariance pattern is available.

In many datasets, especially those with only a few repeated measurements, estimates of overall treatment effects may differ little between models using different covariance patterns. If identifying the covariance pattern is not of intrinsic interest, a compound symmetry pattern may well prove adequate to estimate treatment effects and standard errors. A rough check can be made of this by comparing the results with those obtained using a general pattern. If the differences are small, then the compound symmetry pattern can be used with reasonable confidence.

6.2.3 Choice of fixed effects

Treatment, time, treatment·time and baseline effects (if recorded) can all be fitted as fixed. Estimates of the overall treatment effect will differ depending on whether treatment·time effects are included in the model. When they are, treatment effects are calculated as the average of the treatment estimates obtained from each time point. When they are omitted, a weighted average of estimates from each time point is obtained. The weights are related to the variances of estimates at each time point, which are in turn related to the number of observations at each time point. The decision as to which estimate to use should rest with the desired interpretation and whether the treatment·time effects are significant. The unweighted estimate may be more appealing if there are missing data at later time points, so that bias towards earlier time points is reduced. However, if a treatment effect appears relatively constant over time, then the weighted estimate has the advantage of being less influenced by potentially inaccurate estimates at

time points with fewer observations. If treatment·time effects are significant, it would be wise to present additional treatment estimates at each time point. Less importance should then be attached to the overall treatment estimate.

The previous discussion has assumed that we are interested in comparing treatments, though, of course, the applications can be much wider. In studies where patients are grouped by a variable other than treatment (e.g. an epidemiological study to compare patients with different disease types), the comparator variable should simply be substituted for the treatment effects.

6.2.4 General points

Missing data

Missing data, in the form of gaps in the series of observations or caused by patient dropout, may occur frequently in repeated measures data (see Section 2.4.7). They are a less serious problem, however, when a mixed model is used, unless there are very substantial between-treatment group differences with respect to the pattern of dropout. The reason missing values are less problematical is that observations at each time point influence estimates of treatment effects at every other time point, owing to the specification of a covariance pattern. Thus, patients whose observations are limited to early time points because of dropout will nevertheless be taken into account when estimates are made of treatment effects at later time points. Clearly, such individuals will not influence these estimates as greatly as individuals whose data are complete; so the pattern of missing data in different treatment groups cannot be completely ignored. There will also be potential biases if patients show patterns of rapid deterioration prior to dropout. In such cases, their early observations may be 'good', leading to a corresponding 'good' influence on the unobserved time points after dropout, when this is clearly inappropriate. *Ad hoc* approaches to imputing missing values, or analyses including and excluding dropouts, may then need to be employed. Alternatively, a method that allows for non-random missing data in repeated measures analysis could be considered (e.g. Diggle and Kenward, 1994). Thus, it is too simplistic to say that missing values do not matter in the mixed models analysis of repeated measures data, but the method is quite robust, even when the data may not be entirely missing at random. The text *Missing Data in Clinical* Studies by Molenberghs and Kenward (2007) is recommended for readers wishing to gain more knowledge of this area.

Significance testing

Fixed effects can be tested using F tests as described in Section 2.4.4. Again, Satterthwaite's DF should be used wherever possible. Within `SAS`, the use of the option `DDFM=KENWARDROGER` within the `MODEL` statement is recommended

for reasons expanded upon in the following section. If suitable software is not available for calculating an appropriate DF, we suggest that overall treatment effects and effects at individual time points should be compared using the patient DF. This is likely to produce a conservative test.

Fixed effects standard error bias

Downward bias of fixed effects standard errors will occur as described in Section 2.4.3 because the covariance parameters are estimated and not known. A 'robust' variance estimate for fixed effects, known as the 'empirical' variance estimator, was mentioned for covariance pattern models in Section 2.4.3. This variance takes into account the observed covariance in the data, which, it is claimed, may help alleviate some of the small sample bias. Our own simulation studies have revealed, however, that, for relatively small sample sizes, the use of the empirical variance estimator leads to the wrong size of test – when the null hypothesis is true, an excess of statistically significant differences are found. Paradoxically, we have found that with larger sample sizes, the empirical variance estimator performs much better than at smaller sample sizes.

An alternative approach to adjust for the downward bias of fixed effects standard errors is to inflate the estimated variance–covariance matrix of the fixed effects, var($\widehat{\boldsymbol{\alpha}}$), following the approach described by Kenward and Roger (1997). This is implemented in SAS using the option `DDFM=KENWARDROGER` (or `DDFM=KR`) within the `MODEL` statement. Since the last edition of this text a modified adjustment, `DDFM=KR(LINEAR)`, has become available in `PROC MIXED`. This is expected to provide an improved approximation for certain types of covariance pattern. In `PROC GLIMMIX`, a further improved adjustment, detailed in Kenward and Roger's (2009) publication, `DDFM=KR2`, is available in addition to the `DDFM=KR(LINEAR)` option, but only in `SAS/STAT` 12.1 onwards and not in SAS 9.3. Amongst the examples considered in this text, these new options provide a different standard error to the `DDFM=KR` option only for the first-order autoregressive structure and for the heterogenous structures ((vi)–(viii)) described in Section 6.2.1.

In contrast to our findings on the empirical variance estimator, our simulation studies showed that the Kenward–Roger method performed satisfactorily down to very small sample sizes (five subjects per treatment group) and in the presence of missing values. Our recommendation is that in most circumstances, and particularly with small sample sizes, the Kenward–Roger method should be the one of choice. There is still the problem, of course, of specifying an appropriate covariance pattern. For large sample sizes, the methods of Section 6.2.2 can be applied in the knowledge that there will be reasonable power for detecting statistically significant improvements in model fit, with more complicated covariance patterns. Alternatively, a pragmatic approach that would lead to a simple, pre-specified analysis plan would be to choose a simple covariance pattern, such as compound symmetry, but use the empirical variance estimator.

This will ensure that the estimated standard errors of the fixed effects reflect the observed covariance pattern of the data. Although we cannot recommend this approach for small sample sizes, we feel it is a viable alternative with larger studies. For smaller sample sizes, there are no ideal solutions. Our preference is to choose a simple, plausible, covariance pattern for the situation (often compound symmetry or Toeplitz) and use the Kenward–Roger method.

The empirical variance estimator and the Kenward–Roger method can be used with both normal and non-normal distributions. Our simulation studies have looked only at the normal distribution, and we cannot legitimately infer the extent to which the findings will generalise to non-normal distributions. The empirical variance is calculated in SAS by using the `EMPIRICAL` option in the `PROC MIXED` statement. For non-normal data, it is calculated by default when the `REPEATED` statement in `PROC GENMOD` is used. The introduction of `PROC GLIMMIX` to SAS offers additional flexibility in the use of empirical variance estimators, though we have no experience in their use and cannot make recommendations. Within the `PROC GLIMMIX` statement, there are five options of the form `EMPIRICAL = <keyword>`. The option `EMPIRICAL = CLASSICAL` produces the usual empirical estimators, while the other four choices are different bias-adjusted estimators.

Model checking

In covariance pattern models, it is assumed that the residuals have a multivariate normal distribution with zero means and covariance matrix $\mathbf{R}$. This can be allowed for by dividing the residuals by the matrix representing the root of the variance matrix, $\mathbf{V}$, that is dividing by $\mathbf{C}$ such that $\mathbf{C}'\mathbf{C} = \mathbf{V}$. Thus, these standardised residuals are calculated by $\mathbf{C}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}} - \mathbf{Z}\hat{\boldsymbol{\beta}})$. The standardised residuals can be calculated in `PROC MIXED` using the `VCIRY` option in the `MODEL` statement. These residuals are then assumed to have a normal distribution, which can be checked using the methods suggested in Chapter 2. If there is evidence of non-normality or outliers, then the suggestions described in this section should be used.

6.3 Example: covariance pattern models for normal data

The hypertension trial analysed in Sections 1.3 and 2.5 is now considered as repeated measures data. DBP recorded at each of the fortnightly post-treatment visits will be analysed, and the effect of centres will be ignored. The primary objective is to obtain an overall estimate of the treatment difference. The number of patients attending at each visit is summarised by treatment as shown in the following table. Visits 3–6 are the four post-treatment visits, and visit 2 values are used as the baseline covariate.

	Treatment			
Visit	**A**	**B**	**C**	**Total**
1	106	101	104	311
2	106	100	102	308
3	100	96	94	290
4	95	91	94	280
5	87	88	93	268
6	83	84	91	258

6.3.1 Analysis models

Treatment, time, treatment·time and baseline effects are fitted as fixed effects in all the models considered.

The patterns listed in the following table are fitted and compared statistically using likelihood ratio tests. The covariance pattern used by Model 6 was suggested by the results from Models 1–5.

Model	**Covariance pattern**
1	Compound symmetry
2	First-order autoregressive
3	Toeplitz
4	General
5	Separate compound symmetry for each treatment group
6	Separate Toeplitz pattern for each treatment group

6.3.2 Selection of covariance pattern

The covariance patterns and measures of model fit resulting from each analysis are shown in Table 6.1. Correlations between visits are positive in all models, indicating that it is important to take account of the correlations between the repeated measurements.

Models 1 and 2 are the simplest covariance patterns. Since they each use two covariance parameters, we choose Model 1, which has the highest likelihood. It seems unlikely that the correlation between periods decays exponentially as they become more widely separated as modelled in Model 2.

Model 3 with a Toeplitz pattern indicates that the correlation between visits may be less when they are not adjacent. The compound symmetry pattern is

Table 6.1 Results from using different covariance patterns.

Model	Covariance parameters (variances and correlation matrix)	$-2\log(L)$ (no. parameters)	Akaike's information criterion (AIC)
1	$76\begin{pmatrix} 1 & & & \\ 0.53 & 1 & & \\ 0.53 & 0.53 & 1 & \\ 0.53 & 0.53 & 0.53 & 1 \end{pmatrix}$	7463.4(2)	7467.4
2	$76\begin{pmatrix} 1 & & & \\ 0.57 & 1 & & \\ 0.57^2 & 0.57 & 1 & \\ 0.57^3 & 0.57^2 & 0.57 & 1 \end{pmatrix}$	7485.3 (2)	7489.3
3	$76\begin{pmatrix} 1 & & & \\ 0.57 & 1 & & \\ 0.48 & 0.57 & 1 & \\ 0.46 & 0.48 & 0.57 & 1 \end{pmatrix}$	7450.6 (4)	7458.6
4	$\begin{matrix} 76 \\ 71 \\ 86 \\ 73 \end{matrix}\begin{pmatrix} 1 & & & \\ 0.52 & 1 & & \\ 0.48 & 0.61 & 1 & \\ 0.46 & 0.50 & 0.61 & 1 \end{pmatrix}$	7442.3 (10)	7462.3
5	A85 $\begin{pmatrix} 1 & & & \\ 0.54 & 1 & & \\ 0.54 & 0.54 & 1 & \\ 0.54 & 0.54 & 0.54 & 1 \end{pmatrix}$ B68 $\begin{pmatrix} 1 & & & \\ 0.39 & 1 & & \\ 0.39 & 0.39 & 1 & \\ 0.39 & 0.39 & 0.39 & 1 \end{pmatrix}$ C76 $\begin{pmatrix} 1 & & & \\ 0.63 & 1 & & \\ 0.63 & 0.63 & 1 & \\ 0.63 & 0.63 & 0.63 & 1 \end{pmatrix}$	7447.5 (6)	7459.5
6	A85 $\begin{pmatrix} 1 & & & \\ 0.58 & 1 & & \\ 0.48 & 0.58 & 1 & \\ 0.50 & 0.48 & 0.58 & 1 \end{pmatrix}$ B68 $\begin{pmatrix} 1 & & & \\ 0.42 & 1 & & \\ 0.33 & 0.42 & 1 & \\ 0.42 & 0.33 & 0.42 & 1 \end{pmatrix}$ C76 $\begin{pmatrix} 1 & & & \\ 0.69 & 1 & & \\ 0.61 & 0.69 & 1 & \\ 0.46 & 0.61 & 0.69 & 1 \end{pmatrix}$	7424.0 (12)	7448.0

nested within the Toeplitz pattern used by Model 1, and the models can therefore be compared statistically using a likelihood ratio test. This test gave $\chi^2_2 = 12.78$, indicating that the Toeplitz structure is a significant improvement ($p = 0.002$).

Model 4 with a general pattern also indicates that the correlation between visits is less when they are not adjacent. We determine whether the extra parameters used lead to a significant improvement over Model 3. The likelihood ratio test gives $\chi^2_6 = 8.25$ ($p = 0.22$), which shows that the use of the extra six parameters in the general pattern is not necessary.

Model 5 has separate compound symmetry patterns for each treatment and indicates that covariances may differ between treatments. The likelihood ratio test shows that this model is a significant improvement over Model 1, $\chi^2_4 = 15.80$ ($p = 0.003$). However, it cannot be compared statistically with Model 3 (Toeplitz), since the two models are not nested.

On the basis of the fact that Model 5 indicates that separate covariances for each treatment group may be necessary and that Model 3 suggests a Toeplitz pattern, Model 6 incorporating both these features was tested. Models 3 and 5 are nested within Model 6, and Model 6 shows significant improvements over both of them, $\chi^2_8 = 26.59 (p = 0.0008)$ and $\chi^2_6 = 23.57 (p = 0.0006)$.

Thus, we have statistically justified the use of a fairly complex covariance pattern. This is likely to be partly because the trial is relatively large, and so the covariance parameters are estimated with a reasonable accuracy. Model 6 has given us statistical evidence that the treatment groups have different variances. In addition, the Toeplitz structures indicate that correlations between repeated measurements are the highest for treatment C and the lowest for treatment B. These differences are likely to be, in some way, due to the different actions of the treatments. In smaller trials, however, it is often not possible statistically to justify any pattern more complex than the compound symmetry or first-order autoregressive. This is not necessarily because a more complex pattern does not exist but because there is insufficient information to determine it.

6.3.3 Assessing fixed effects

The overall treatment effect estimates obtained from Models 1 and 6 are summarised in Table 6.2. The small differences arise from differences in the distribution of missing values between treatments. Thus, in this dataset, selecting the most appropriate covariance pattern has had little effect on the results. However, this is not always the case. In situations where the variance differs between the visits, standard errors can differ more markedly.

In the last row of Table 6.2, results obtained after omitting the treatment·time effects are given. The treatment effects are then calculated as weighted averages of the effects at each time point. This contrasts with the models fitting the

Table 6.2 Comparing treatment effects between Models 1 and 6.

		Model 1 (compound symmetry)	Model 6 (separate Toeplitz pattern for each treatment)
Overall treatment effects	A – B	1.22 (1.03)	1.25 (0.99)
	A – C	3.01 (1.02)	3.04 (1.08)
	B – C	1.79 (1.03)	1.79 (1.00)
Visit 3 treatment effects	A – B	1.35 (1.26)	1.36 (1.26)
	A – C	3.40 (1.25)	3.42 (1.29)
	B – C	2.05 (1.28)	2.06 (1.24)
Visit 4 treatment effects	A – B	0.56 (1.28)	0.56 (1.28)
	A – C	1.86 (1.27)	1.89 (1.30)
	B – C	1.30 (1.28)	1.34 (1.25)
Visit 5 treatment effects	A – B	2.91 (1.30)	3.00 (1.31)
	A – C	4.67 (1.29)	4.77 (1.32)
	B – C	1.76 (1.29)	1.77 (1.26)
Visit 6 treatment effects	A – B	0.05 (1.32)	0.09 (1.33)
	A – C	2.09 (1.30)	2.10 (1.34)
	B – C	2.04 (1.31)	2.01 (1.27)
Treatment·visit *p*-value		0.22 (1.31)	0.11 (1.31)
Overall treatment	A – B	1.23 (1.02)	1.23 (1.02)
effects in model	A – C	3.03 (1.03)	3.01 (1.02)
omitting treatment·visit effects	B – C	1.80 (1.03)	1.78 (1.03)

treatment·time interaction, which causes treatment effects to be unweighted averages of the effects at each time point. There is little difference between the two sets of results because there is only a relatively small amount of missing data. It is a matter of personal choice as to which approach is preferable. There are noticeable differences in the treatment effect estimates between the visits. However, there is an absence of any coherent pattern, and the treatment·time interaction was not significant in any model.

6.3.4 Model checking

Residual plots of scaled residuals are used to detect any outliers or a general lack of normality. The graphs of the scaled residuals in Figure 6.1 neither indicate appreciable departure from normality, nor any definite outlying observations.

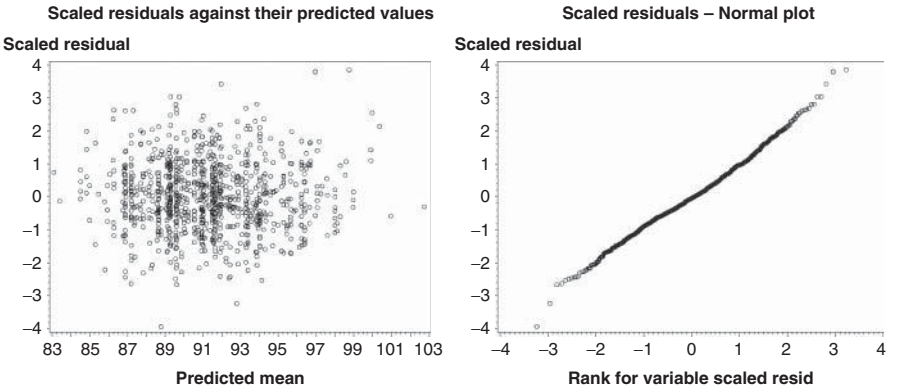


Figure 6.1(a) Residual plots (using `OUTP` and `PROC GPLOT`).

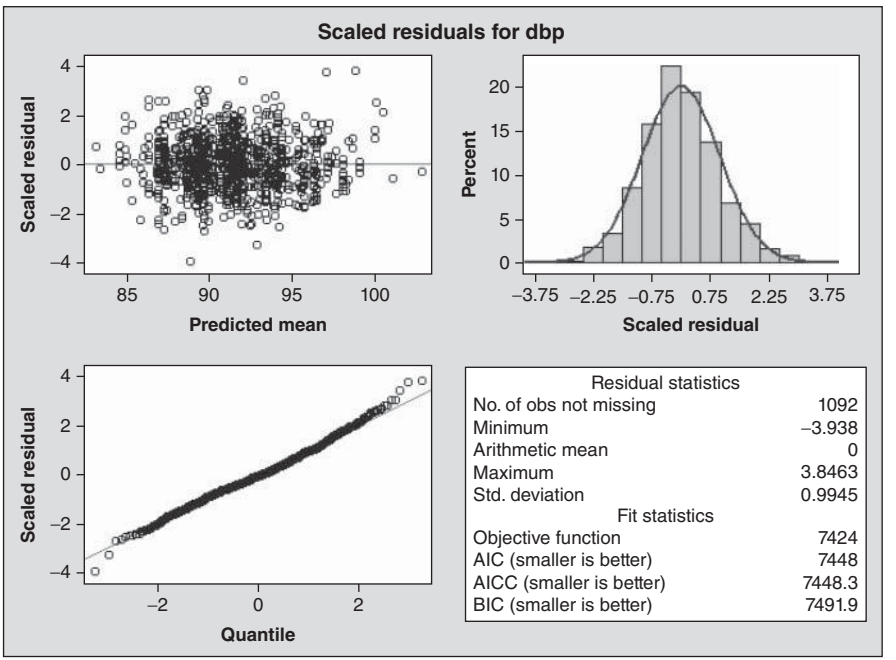


Figure 6.1(b) Residual plots and statistics (via `RESIDUAL` option, `SAS` Version 9).

`SAS` code and output

Variables
`treat` = treatment(A,B.C),
`pat` = patient number,

dbp = diastolic blood pressure,
dbp1 = baseline diastolic blood pressure,
visit = visit number.

SAS code is shown for Model 6. Code for the other models is identical except that different REPEATED statements are used and Model 2 (first order autoregressive) uses the DDFM=KR(LINEAR) option to provide an improved adjustment to the standard errors and significance test DFs (see Section 6.2.4). For the other models, the use of the DDFM=KR(LINEAR) option led to identical standard errors and significance test DFs to the DDFM=KR option. The REPEATED statements for Models 1–5 are given after the Model 6 output.

```
PROC MIXED NOCLPRINT; CLASS pat treat visit;
MODEL dbp = dbp1 treat visit treat*visit/ DDFM=KR;
REPEATED visit/ SUBJECT=pat TYPE=TOEP GROUP=treat
    R=1,3,4 RCORR=1,3,4;
LSMEANS treat/ DIFF PDIFF CL;

ESTIMATE 'a-b,v3' treat 1 -1  0 treat*visit 1 0 0 0 -1  0  0  0  0  0  0  0;
ESTIMATE 'a-c,v3' treat 1  0 -1 treat*visit 1 0 0 0  0  0  0  0 -1  0  0  0;
ESTIMATE 'b-c,v3' treat 0  1 -1 treat*visit 0 0 0 0  1  1  0  0 -1  0  0  0;
ESTIMATE 'a-b,v4' treat 1 -1  0 treat*visit 0 1 0 0  0 -1  0  0  0  0  0  0;
ESTIMATE 'a-c,v4' treat 1  0 -1 treat*visit 0 1 0 0  0  0  0  0  0 -1  0  0;
ESTIMATE 'b-c,v4' treat 0  1 -1 treat*visit 0 0 0 0  0  1  1  0  0 -1  0  0;
ESTIMATE 'a-b,v5' treat 1 -1  0 treat*visit 0 0 1 0  0  0 -1  0  0  0  0  0;
ESTIMATE 'a-c,v5' treat 1  0 -1 treat*visit 0 0 1 0  0  0  0  0  0  0 -1  0;
ESTIMATE 'b-c,v5' treat 0  1 -1 treat*visit 0 0 0 0  0  0  1  0  0  0 -1  0;
ESTIMATE 'a-b,v6' treat 1 -1  0 treat*visit 0 0 0 1  0  0  0 -1  0  0  0  0;
ESTIMATE 'a-c,v6' treat 1  0 -1 treat*visit 0 0 0 1  0  0  0  0  0  0  0 -1;
ESTIMATE 'b-c,v6' treat 0  1 -1 treat*visit 0 0 0 0  0  0  0  1  0  0  0 -1;
```

The NOCLPRINT option suppresses the lengthy printing of patient categories. The use of the R and RCORR options displays the covariance parameters in a more meaningful way as matrices. If no subject numbers are specified, matrices for only the first patient will be printed. We have requested matrices for patients 1, 3 and 4 so that a covariance and correlation matrix is printed for a patient on each treatment.

```
                         Iteration History
Iteration          Evaluations          -2 Res Log Like          Criterion
0                            1            7794.69107620
1                            2            7424.25314231         0.00009243
2                            1            7423.98385033         0.00000131
3                            1            7423.98024667         0.00000000
```

Convergence criteria met.

Estimated R Matrix for pat 1

Row	Col1	Col2	Col3	Col4
1	76.1169	52.7624	46.4925	35.3745
2	52.7624	76.1169	52.7624	46.4925
3	46.4925	52.7624	76.1169	52.7624
4	35.3745	46.4925	52.7624	76.1169

Estimated R Correlation Matrix for pat 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6932	0.6108	0.4647
2	0.6932	1.0000	0.6932	0.6108
3	0.6108	0.6932	1.0000	0.6932
4	0.4647	0.6108	0.6932	1.0000

Estimated R Matrix for pat 3

Row	Col1	Col2	Col3	Col4
1	68.2100	28.9323	22.4760	28.6921
2	28.9323	68.2100	28.9323	22.4760
3	22.4760	28.9323	68.2100	28.9323
4	28.6921	22.4760	28.9323	68.2100

Estimated R Correlation Matrix for pat 3

Row	Col1	Col2	Col3	Col4
1	1.0000	0.4242	0.3295	0.4206
2	0.4242	1.0000	0.4242	0.3295
3	0.3295	0.4242	1.0000	0.4242
4	0.4206	0.3295	0.4242	1.0000

Estimated R Matrix for pat 4

Row	Col1	Col2	Col3	Col4
1	84.9809	49.5186	41.0737	42.2137
2	49.5186	84.9809	49.5186	41.0737
3	41.0737	49.5186	84.9809	49.5186
4	42.2137	41.0737	49.5186	84.9809

Estimated R Correlation Matrix for pat 4

Row	Col1	Col2	Col3	Col4
1	1.0000	0.5827	0.4833	0.4967
2	0.5827	1.0000	0.5827	0.4833
3	0.4833	0.5827	1.0000	0.5827
4	0.4967	0.4833	0.5827	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Variance	pat	treat A	84.9809
TOEP(2)	pat	treat A	49.5186
TOEP(3)	pat	treat A	41.0737
TOEP(4)	pat	treat A	42.2137
Variance	pat	treat B	68.2100
TOEP(2)	pat	treat B	28.9323
TOEP(3)	pat	treat B	22.4760
TOEP(4)	pat	treat B	28.6921
Variance	pat	treat C	76.1169
TOEP(2)	pat	treat C	52.7624
TOEP(3)	pat	treat C	46.4925
TOEP(4)	pat	treat C	35.3745

Fit Statistics

-2 Res Log Likelihood	7424.0
AIC (smaller is better)	7448.0
AICC (smaller is better)	7448.3
BIC (smaller is better)	7491.9

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
11	370.71	<.0001

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	285	28.73	<.0001
treat	2	187	4.04	0.0192
visit	3	437	12.41	<.0001
treat*visit	6	356	1.73	0.1130

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
a-b,v3	1.3578	1.2598	426	1.08	0.2817
a-c,v3	3.4212	1.2866	367	2.66	0.0082
b-c,v3	2.0634	1.2426	390	1.66	0.0976
a-b,v4	0.5557	1.2789	439	0.43	0.6641
a-c,v4	1.8917	1.3002	377	1.45	0.1465
b-c,v4	1.3360	1.2481	393	1.07	0.2851
a-b,v5	3.0026	1.3081	456	2.30	0.0222
a-c,v5	4.7694	1.3221	390	3.61	0.0003
b-c,v5	1.7668	1.2592	400	1.40	0.1614
a-b,v6	0.08613	1.3316	470	0.06	0.9485
a-c,v6	2.0957	1.3397	399	1.56	0.1185
b-c,v6	2.0096	1.2745	410	1.58	0.1156

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
treat	A	92.7437	0.7595	96.2	122.11	<.0001	0.05
treat	B	91.4931	0.6404	93.5	142.87	<.0001	0.05
treat	C	89.6992	0.7631	93.3	117.55	<.0001	0.05

Least Squares Means

Effect	treat	Lower	Upper
treat	A	91.2361	94.2513
treat	B	90.2215	92.7648
treat	C	88.1840	91.2144

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
treat	A	B	1.2506	0.9945	186	1.26	0.2102	0.05
treat	A	C	3.0445	1.0759	191	2.83	0.0052	0.05
treat	B	C	1.7939	0.9975	181	1.80	0.0738	0.05

Differences of Least Squares Means

Effect	treat	_treat	Lower	Upper
treat	A	B	-0.7115	3.2126
treat	A	C	0.9223	5.1667
treat	B	C	-0.1744	3.7623

`REPEATED` statements for Models 1–5

```
1. REPEATED visit/ SUBJECT=pat TYPE=CS R RCORR;
2. REPEATED visit/ SUBJECT=pat TYPE=AR(1) R RCORR;
3. REPEATED visit/ SUBJECT=pat TYPE=TOEP R RCORR;
4. REPEATED visit/ SUBJECT=pat TYPE=UN R RCORR;
5. REPEATED visit/ SUBJECT=pat TYPE=CS GROUP=treat R=1,3,4
         RCORR=1,3,4;
```

Checking model assumptions in Model 6

The following code can be used to obtain the residual and normal plots shown in Figure 6.1. These are based on the scaled residuals. When the `VCIRY` option is used, `SAS` only scales the marginal residuals (produced by the `OUTPM` option). However, in this case, these are the same as the conditional residuals produced by the `OUTP` option since no random effects are fitted.

```
PROC MIXED NOCLPRINT; CLASS treat visit pat;
MODEL dbp = dbp1 treat visit treat*visit/ DDFM=KR VCIRY
         OUTPM=predm;
REPEATED visit/ SUBJECT=pat TYPE=TOEP GROUP=treat;
SYMBOL1 v=circle;
PROC GPLOT DATA=predm;
PLOT scaledresid*pred;
TITLE 'SCALED RESIDUALS AGAINST THEIR PREDICTED VALUES';
PROC RANK DATA=predm OUT=norm NORMAL=TUKEY; VAR scaledresid;
         RANKS s_est;
PROC GPLOT DATA=norm; PLOT scaledresid*s_est;
TITLE 'SCALED RESIDUALS - NORMAL PLOT';
```

Alternatively, for SAS/GRAPH users, the following ODS code can be used to produce the residual plots shown in Figure 6.2. In SAS 9.3, the labelling of the graphs has changed slightly.

```
ODS HTML FILE="<output file.html>" GPATH="<graphs directory>";
ODS GRAPHICS ON;
PROC MIXED NOCLPRINT; CLASS treat visit pat;
MODEL dbp = dbp1 treat visit treat*visit/ RESIDUAL
   DDFM=KR VCIRY;
REPEATED visit/ SUBJECT=pat TYPE=TOEP GROUP=treat;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
```

6.4 Example: covariance pattern models for count data

This was a placebo-controlled trial of an anti-convulsant treatment for epilepsy involving 59 patients. The data are taken from Thall and Vail (1990). The number of epileptic seizures was counted over an 8-week period prior to treatment and then over four 2-week periods following treatment. None of the patients dropped out of the study. A histogram of the number of epileptic episodes is shown in Figure 6.2 by treatment group. These show that many patients have few seizures, while a few have a large number. This distribution indicates that a Poisson error and a log link function may be appropriate. However, it is possible that the small number of very large frequencies will produce outlying residuals. Note that in this example an offset variable (see Section 3.1.2) is not needed because the trial periods are strictly 2 weeks long (it does not matter that the baseline period is longer – this is taken into account by the baseline effect estimate).

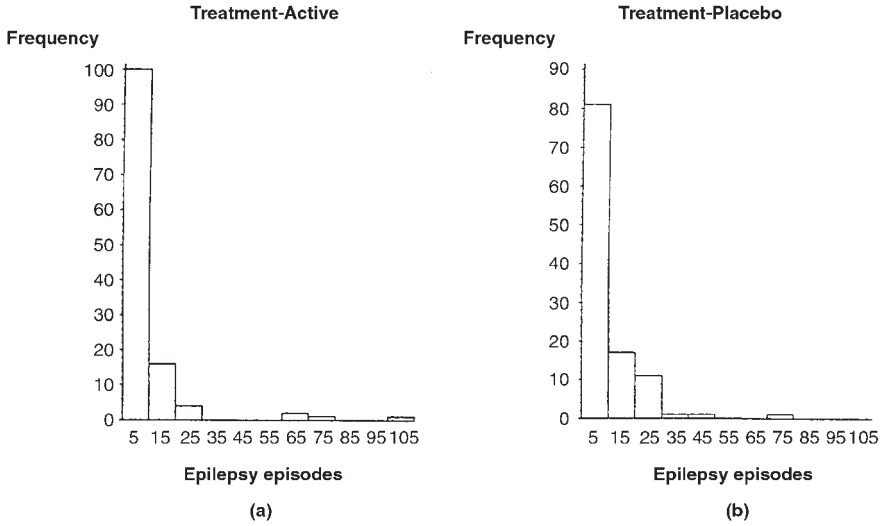


Figure 6.2 Histograms of the number of epilepsy episodes by treatment group.

6.4.1 Analysis models

Models with several different covariance patterns are fitted using pseudo-likelihood (PL). Baseline, treatment, visit and treatment·visit are fitted as fixed in all the models. Models 1–5 were fitted initially. The pattern for Model 6 was suggested by the results from these models.

Model	Covariance pattern
1	Compound symmetry
2	Separate compound symmetry for each treatment group
3	First-order autoregressive
4	Toeplitz
5	General
6	Separate general pattern for each treatment group

Results

Results from the six models are shown in Table 6.3. Correlations in all the models are positive, indicating that observations on the same patient are correlated. We assume that approximate likelihood ratio tests based on $-2\log(\text{PL})$ can be used to compare different models. Models 1 and 3 have the same number of covariance parameters, and therefore their PLs can be compared directly. Since Model 1 (compound symmetry) has the highest PL, it is therefore preferred to Model 3

Table 6.3 Results from all models.

Model	Covariance parameters	−2log(PL[a]) (no. covariance parameters)	Treatment difference (placebo−active) (SE)
1	$\begin{pmatrix} 4.90 & & & \\ 2.08 & 4.90 & & \\ 2.08 & 2.08 & 4.90 & \\ 2.08 & 2.08 & 2.08 & 4.90 \end{pmatrix}$	603.95 (2)	0.106 (0.152)
2	P $\begin{pmatrix} 5.55 & & & \\ 1.07 & 5.55 & & \\ 1.07 & 1.07 & 5.55 & \\ 1.07 & 1.07 & 1.07 & 5.55 \end{pmatrix}$ A $\begin{pmatrix} 4.29 & & & \\ 2.42 & 4.29 & & \\ 2.42 & 2.42 & 4.29 & \\ 2.42 & 2.42 & 2.42 & 4.29 \end{pmatrix}$	591.11 (4)	0.106 (0.151)
3	4.78[b] $\begin{pmatrix} 1.00 & & & \\ 0.43 & 1.00 & & \\ 0.43^2 & 0.43 & 1.00 & \\ 0.43^3 & 0.43^2 & 0.43 & 1.00 \end{pmatrix}$	612.32 (2)	0.106 (0.137)
4	$\begin{pmatrix} 4.90 & & & \\ 2.22 & 4.90 & & \\ 1.81 & 2.22 & 4.90 & \\ 2.18 & 1.81 & 2.22 & 4.90 \end{pmatrix}$	602.93 (4)	0.106 (0.152)
5	$\begin{pmatrix} 4.98 & & & \\ 2.27 & 4.15 & & \\ 2.09 & 2.82 & 7.82 & \\ 1.47 & 1.52 & 2.23 & 2.44 \end{pmatrix}$	566.86 (10)	0.106 (0.150)
6	P $\begin{pmatrix} 4.53 & & & \\ 2.68 & 3.63 & & \\ 1.49 & 2.27 & 10.61 & \\ 0.05 & 0.58 & 1.69 & 1.92 \end{pmatrix}$ A $\begin{pmatrix} 5.62 & & & \\ 1.81 & 4.02 & & \\ 2.97 & 3.26 & 5.52 & \\ 3.05 & 2.28 & 2.93 & 2.85 \end{pmatrix}$	504.69 (20)	0.105 (0.151)

[a] PL = pseudo-likelihood value.
[b] Here the variance and correlation matrix are given.

(autoregressive). Model 2 with separate compound symmetry patterns for each treatment performs significantly better than Model 1 ($\chi^2_2 = 12.84$), and so does Model 5 with a general pattern ($\chi^2_6 = 37.09$). However, Model 4 with a Toeplitz pattern is no better than Model 1 ($\chi^2_2 = 1.02$) and can be rejected. On the basis of the fact that Model 2 indicates that separate covariances for treatments may be necessary and that Model 5 indicates a general covariance pattern, Model 6 combining these features is tested. This shows a significant improvement over both Models 2 and 5 ($\chi^2_{18} = 86.42$ and $\chi^2_{10} = 62.17$, respectively) and is therefore selected.

The baseline effect was highly significant in all models, and its inclusion has increased the precision of the analyses. The treatment·time interaction was not significant in any model ($p > 0.8$ in all models), and so we can be reasonably confident in reporting an overall treatment effect based on the average of all time points. In this example, the overall treatment effect will not be affected by whether the interaction term is removed because the data are complete. No evidence of a significant treatment effect was apparent in any model. Relative rates and 95% confidence intervals can be calculated from the mean treatment difference and SE. In Model 6, the confidence interval for the treatment effect on the linear scale is

$$95\%\ \text{CI} = 0.105 \pm t_{57,0.975} \times 0.151.$$

$t_{57,0.975} = 2.00$, so we obtain

$$95\%\ \text{CI} = 0.105 \pm 2.00 \times 0.151 = (-0.197, 0.407).$$

A comparison of the treatments in terms of a relative rate is obtained by exponentiating the effect estimate:

$$\text{RR} = \frac{\text{Seizure rate on placebo}}{\text{Seizure rate on active}} = \exp(0.105) = 1.11.$$

Confidence intervals for the relative rate are calculated by exponentiating the confidence intervals calculated on the linear scale, $\exp(-0.197, 0.407) = (0.82, 1.50)$.

Plots of the Pearson residuals against their fitted values are used to provide a rough check of model assumptions and to look for outliers (Figure 6.3).

These plots indicate that the Poisson assumption is not appropriate. The model has not adequately coped with observations with very large values. The influence of these observations on the results was assessed by refitting Model 6 with the largest residual removed (patient 25, visit 3). This model gives a treatment effect estimate of 0.029 with a standard error of 0.140. This is noticeably different from the treatment effect of 0.105 (0.151) estimated with the most outlying observation included. The variances for the placebo at visit 3 are also now much less, indicating that the previous estimate was highly influenced by the outlier. Even with the outlier removed, the distribution of residuals remains unsatisfactory. Therefore, we should conclude that the assumption of Poisson

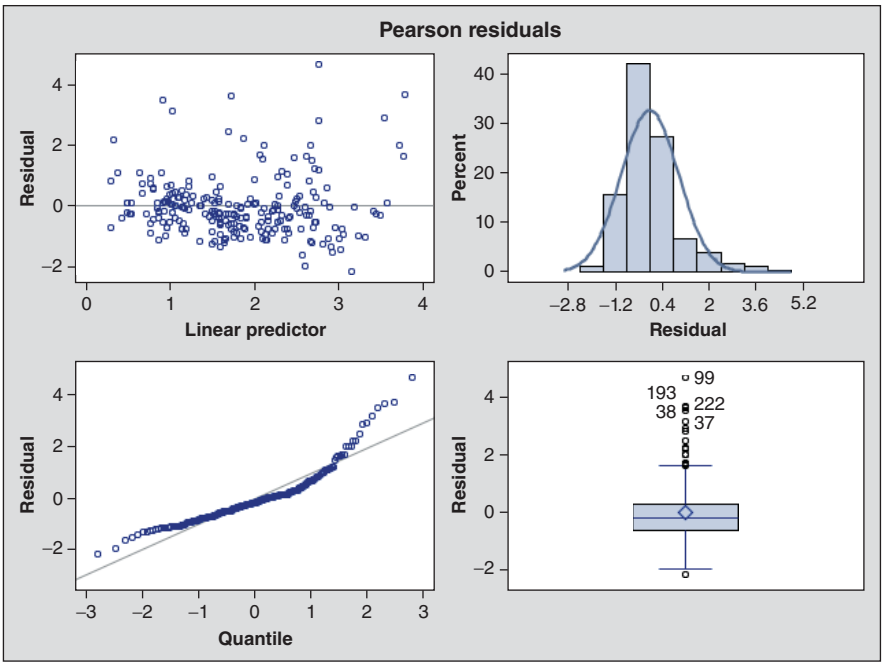


Figure 6.3 Panel of Pearson residuals: scatterplot of residuals against their predicted values; a histogram with normal density; a Q-Q plot and a boxplot of the residuals.

error for these data is not appropriate and consider an alternative analysis method. A transformation of the data is unlikely to overcome this problem since there are a large number of zeros. In the absence of non-parametric mixed models, one possibility is to categorise the number of epilepsy attacks and to use a categorical mixed model. This will be illustrated in the following section.

6.4.2 Analysis using a categorical mixed model

The number of post-treatment epilepsy attacks was categorised into four groups. The groupings are chosen so that each category is of a reasonable size.

Group	Attacks
1	0 (10%)
2	1–3 (29%)
3	4–10 (39%)
4	11 + (22%)

The baseline attack rate was not categorised and was again fitted as log(attacks). Attempts were made to fit compound symmetry, Toeplitz and general covariance patterns to the data using the SAS macro written by Lipsitz *et al.* (1994) (see Section 9.3). However, convergence was achieved only for the compound symmetry model. The treatment·time interaction was not significant, and this effect was removed from the model.

Covariance parameter estimates

Correlations between each pair of partitions are given in the following table. Recall from Section 4.2 that instead of a single correlation parameter to model the compound symmetry pattern, there is now a matrix of rank $(c-1)\times(c-1)$, giving a correlation parameter for each pair of partitions. The correlation values for our compound symmetry model were:

Partition	1	2	3
1	0.67		
2	0.03	0.17	
3	0.00	0.16	0.28

Thus, there appears to be some correlation between observations on the same patients.

Fixed effects estimates

The fixed effect estimates were as follows:

Effect	Estimate	Empirical SE	Model-based SE
Intercept 1	−0.32	0.56	0.52
2	1.95	0.50	0.42
3	4.67	0.65	0.61
Baseline	−0.078	0.01	0.01
Treatment	−0.81	0.35	0.35
Visit 1	−0.34	0.31	0.26
Visit 2	−0.64	0.37	0.26
Visit 3	0.29	0.36	0.25

The three intercept terms (arising from the three possible partitions of the four categories) and the visit terms are of little interest. The large size of the baseline term relative to its standard error indicates that the model has benefited from its

inclusion. The overall treatment effect is significant ($p = 0.02$). This differs from the GLMM analysis and appears to indicate that the analysis of the categorised attack rate is more sensitive. This is likely to be because there are three extremely large values (> 60 attacks) in the active treatment group compared with only one in the placebo group. In the GLMM analysis, these will have a large effect on the variance of the treatment effects, whereas this does not occur in the categorical analysis since they are grouped with other values above 10. Similarly, they will have a reduced influence on the estimated magnitude of the treatment effect.

The coefficient for the treatment effect is difficult to interpret directly, but by exponentiation, we can calculate an odds ratio in an analogous way to Section 3.3.4 where GLMMs were considered. In this case, $\exp(-0.81) = 0.44$, and this is the estimate of the odds ratio for the probability of a 'favourable' outcome on placebo compared with active treatment, whether 'favourable' is defined as 0 attacks, ≤3 attacks or ≤10 attacks. Note that the odds ratio from this model is defined in terms of a 'favourable' outcome, whereas in the GLMM, it is in terms of the rate of epilepsy attacks. Note also that it is an inherent assumption of this model that the same odds ratio applies to every partition between the categories. The 95% confidence intervals can be calculated as before from $\exp(-0.81 \pm t_{57,0.975} \times 0.35) = \exp(-0.81 \pm 2.00 \times 0.35) = (0.22, 0.90)$. (Note that these use a t distribution with the patient DF of 57 as used in the GLMM.)

SAS code and output

Variables
`pat` = patient,
`time` = time (1, 2, 3, 4),
`treat` = treatment (1=anti-convulsant drug, 0=placebo),
`epis` = number of epilepsy attacks,
`lbase` = log(baseline epilepsy attacks).

Model 6 did not converge in `PROC GLIMMIX`. However, convergence was obtained using a previous `GLIMMIX SAS` macro, and this was used to provide the results in Table 6.3.

`SAS` code is given below for Model 2. Code for the other models is identical except that different `RANDOM` statements are used. These statements are printed after the Model 2 output.

Model 2

```
PROC GLIMMIX PLOTS=ALL; CLASS pat time treat;
TITLE 'Compound Symmetry separate variances';
MODEL epis=lbase treat time treat*time/ DIST=P DDFM=KR;
RANDOM time / SUBJECT=pat TYPE=CS RESIDUAL VCORR=1,29
GROUP=treat;
ESTIMATE 'treat' treat 1 -1/ CL;
```

Fit Statistics

-2 Res Log Pseudo-Likelihood	591.11
Generalized Chi-Square	227.00
Gener. Chi-Square / DF	1.00

Estimated V Correlation Matrix for pat 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.3061	0.3061	0.3061
2	0.3061	1.0000	0.3061	0.3061
3	0.3061	0.3061	1.0000	0.3061
4	0.3061	0.3061	0.3061	1.0000

Estimated V Correlation Matrix for pat 29

Row	Col1	Col2	Col3	Col4
1	1.0000	0.5641	0.5641	0.5641
2	0.5641	1.0000	0.5641	0.5641
3	0.5641	0.5641	1.0000	0.5641
4	0.5641	0.5641	0.5641	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate	Standard Error
Variance	pat	treat 0	3.8508	0.6051
CS	pat	treat 0	1.6983	0.7757
Variance	pat	treat 1	1.8701	0.2788
CS	pat	treat 1	2.4198	0.7807

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
lbase	1	56.98	120.61	<.0001
treat	1	56.01	0.49	0.4867
time	3	147	1.19	0.3175
time*treat	3	147	0.24	0.8711

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
treat	0.1060	0.1513	56.01	0.70	0.4867	0.05	-0.1972	0.4091

RANDOM statements for other models

```
1. RANDOM time/ SUBJECT=pat TYPE=CS RESIDUAL VCORR;
3. RANDOM time/ SUBJECT=pat TYPE=AR(1) RESIDUAL VCORR;
4. RANDOM time/ SUBJECT=pat TYPE=TOEP RESIDUAL VCORR;
5. RANDOM time/ SUBJECT=pat TYPE=UNR RESIDUAL VCORR;
6. RANDOM time/ SUBJECT=pat TYPE=UNR RESIDUAL
     VCORR=1,29 GROUP=treat;
```

Residual plots In version 9.3, the previous code, through the `PLOTS=ALL` option will provide several types of residual plots, one of which appears in Figure 6.3.

Note that the residuals will be correlated in this example, which is not ideal. In `PROC GLIMMIX`, there is no equivalent to the `VCIRY` option in `PROC MIXED`, and so direct examination of the residuals to identify potential outliers is the simplest practical method.

Use of the V option Although the V option is perhaps the obvious one to choose in order to produce the covariance parameters shown in Table 6.3, in Version 9.3, it is not functioning correctly, at the time of writing.

Categorical analysis The `SAS` macro written by Lipsitz *et al.* (1994) was used to fit these models (see Section 9.3). This macro and the code used can be obtained from web page www.wiley.com/go/brown/applied_mixed. Note that a similar model to the compound symmetry model could have been fitted with `PROC GLIMMIX` through fitting a patient effect as random and specifying `DIST = MULTINOMIAL`.

6.5 Random coefficients models

6.5.1 Introduction

A random coefficients model is an alternative approach to modelling repeated measures data. The model is devised to describe arithmetically the relationship of a measurement with time. The statistical properties of random coefficients models have already been introduced in Sections 1.4.2 and 2.1.4. In this section, we will consider in more depth the practical details of fitting these models and the situations in which they are most appropriate.

The most common applications are those in which a linear relationship is assumed between the outcome variable of interest and time. The main question of interest is then likely to be whether the rate of change in this outcome variable differs between the 'treatment' groups. Such an example was reported by Smyth *et al.* (1997). They carried out a randomised controlled trial of glutathione versus placebo in patients with ovarian cancer who were being treated with *cis*-platinum. This drug has proven efficacy in the treatment of ovarian cancer but has a number of adverse effects as well. Amongst these is a toxic effect on the kidneys. This effect can be monitored by the creatinine levels in the patients' blood. One of the hoped-for secondary effects of glutathione was to reduce the rate of decline of renal (kidney) function. This was assessed using a random coefficients model, but analysis showed no statistically significant difference between the rates of decline in the two treatment arms. Such an analysis may find widespread application in the analysis of 'safety' variables in clinical trials because it is important to establish what effect new drugs may have on a range

of biochemical and haematological variables. If these variables are measured serially, analysis is likely to be more efficient if based on all observations, using a method that will be sensitive to a pattern of rise or decline in the 'safety' variables. A further example in which the rate of decline of CD4 counts is compared in two groups of HIV-infected haemophiliacs will be presented in detail in Section 6.6.1.

In fitting linear random coefficients models, as described above, we will wish to fit fixed effects to represent the average rate of change of our outcome variables over time (i.e. a time effect) and assess the extent to which treatments differ in the average rate of change by fitting a treatment·time interaction. We will also require fixed effects to represent the average intercepts for each treatment (i.e. a treatment effect). In addition to the fixed effects representing average slopes and intercepts, the random coefficients model allows the slopes and intercepts to vary randomly between patients and cause a separate regression line to be fitted for each patient. This is achieved by fitting patient effects (to allow intercepts to vary) and patient·time as random to allow slopes to vary. These effects are used in the calculation of the standard errors of the time and treatment·time effects, which are our main focus of interest. Our basic model is therefore

Fixed effects: time, treatment and treatment·time,
Random effects: patient and patient·time.

The effects included represent a minimum set of effects that will be considered in the model. Other patient characteristics, such as age and sex and their interactions with time, can readily be incorporated into the model, and we will see later that polynomial relationships and the effect of baseline levels can also be incorporated.

When the repeated measures data are obtained at fixed points in time, there will be a choice between the use of covariance pattern models and random coefficients models. This choice may be influenced by how well the dependency of the observations on time can be modelled and whether interest is centred on the changing levels of the outcome variable over time or on its absolute levels. In many instances, the random coefficients model will be the 'natural' choice, as in the examples presented. If the times of observation are not standardised, or if there are substantial discrepancies between the scheduled times and actual time of observation, then random coefficients models are more likely to be the models of choice.

Modelling non-linear relationships with time

The models considered so far assume a linear relationship with time. In many applications, the linear model described is sufficient for assessing whether there is a time trend or whether the trend is varying across treatment groups. However, it is also possible to model non-linear relationships with time; for example, by using polynomial or exponential functions. We will only consider models that can be fitted in `PROC MIXED` using polynomial functions of time. `PROC NLMIXED` can also be used to fit other types of non-linear functions but that is outwith the coverage in this book.

We suggest that a model based on polynomials of time is built up by adding polynomials of increasingly higher order one at a time, both as fixed effects and random coefficients. If a variance component for a random coefficient is negative, the last random coefficient added to the model should be removed, and no further random coefficients should be added. However, higher order polynomials can still be considered as fixed effects if appropriate. This model building process is illustrated in the worked example in Section 6.6.2 and readers may find it helpful to refer to this example and the example in Section 6.6.1 before considering the material in the remainder of this section.

6.5.2 General points

Negative variance components

The usual action when a negative variance component estimate is obtained for a random coefficient would be to constrain the variance component to zero or to refit the model with the random coefficient removed. In `PROC MIXED`, we have found that non-convergence or a message stating that the **G** matrix is not positive semi-definite are usually indications of a negative variance component. In some cases, `PROC MIXED` will successfully constrain the variance component to be close to zero and provide satisfactory results, but in others, the model will not converge. In this second situation, it appears necessary to remove the random coefficient from the model. This is similar, apart from the degrees of freedom, to setting the variance component at zero. When there are non-linear terms of increasing complexity in the model, a possible course of action would be to remove the random coefficients one by one in decreasing order of complexity until all variance components become positive. We should add a word of caution, however. A negative variance component can sometimes occur if the model has been misspecified. For example, a negative variance component for a linear slope could occur when there is actually a non-linear relationship with time. As in all modelling situations, we need to check that our assumptions are reasonable. Sometimes, it will be worthwhile exploring the addition of other non-linear terms before removing random coefficients with negative variance components. Very occasionally, one may wish to permit a variance component to become negative. For example, with non-linear terms in the model, the marginal variance over time may be better modelled by allowing, say, the quadratic term to be negative. In `SAS`, this can be achieved through the `NOBOUND` option in the `PROC MIXED` or `PROC GLIMMIX` statements.

Use of baseline measurements

If there is a pre-treatment, baseline observation, then there are two distinct ways in which it can be used. In one approach, it can be specified as a fixed effect, so

that it is considered as a covariate in the analysis. However, it will often be more natural to think of such an observation as the first repeated measurement at time zero, with time measured from the start of treatment. Such an example occurs in Section 6.6.1 when we analyse the change in CD4 counts over time.

Non-comparative datasets

Repeated measures data are not always collected to compare specific groups of patients (e.g. treatment groups); data may be collected simply to monitor a group of patients over time. In this situation, the linear model described in Section 6.5 would simplify to:

Fixed effects: time,
Random effects: patient and patient·time,

and interest would lie primarily with the time effect estimate. Its standard error would reflect the variation in time slope occurring between patients.

Shrunken random coefficients estimates

Estimates of the random coefficients (i.e. intercept and time effects for each patient) are not usually of interest. However, it is possible to estimate them, and they will be shrunken towards the overall intercept and time effects. This avoids the potential problem of unrealistic slope estimates that may occur when there are only a few observations per patient.

Significance testing

The points relating to significance testing given in Section 2.4.4 also apply to random coefficients models. Time and treatment·time effects can be tested using *F* tests based on the Kenward–Roger method. If software is not available for calculating Satterthwaite's DF, then the patient DF can be used as a conservative estimate of the denominator DF (since time and treatment·time effects are contained within the random patient·time coefficients whose DF are equal to the patient DF).

If required, the significance of the variance components arising from the random coefficients can also be tested using likelihood ratio tests to compare models including and not including the corresponding random coefficients. However, the degrees of freedom used for the test now needs to take into account the number of additional covariance parameters required to add the random coefficient. For example, when a random slope is added, there are two additional covariance parameters – a variance component for the slopes and a correlation between the random slopes and intercepts. The conventional likelihood ratio test would use 2 degrees of freedom. However, ideally, the test should take into account that the variance component for the random slope is truncated at zero. The *p*-value is then based on a 50:50 mixture distribution of chi-squared

distributions with 1 and 2 degrees of freedom.

$$p = \frac{1}{2}P(\chi_1^2) + \frac{1}{2}P(\chi_2^2).$$

Similarly, a third random term in the model would lead to three additional parameters being included in the model, and a p-value based on a 50:50 mixture distribution of chi-squared distributions with 2 and 3 degrees of freedom would be appropriate. More justification for the use of these mixture distributions is given in Section 2.4.4 and in Verbeke and Molenberghs (2000, Section 6.3.4).

However, even if the variance component is non-significant, it is usually desirable to retain the associated random coefficient in the model, provided its variance component is positive.

Model checking

The residuals can be checked for normality by plotting them against their predicted values and by using normal plots. In addition, plots of the residuals against time will help check whether their variance is constant over time. The random coefficients are assumed to have a multivariate normal distribution with zero mean and covariance matrix **G**. This assumption is difficult to check formally. However, estimates of the fixed effects and their standard errors are expected to be robust to any non-normality in the random effects (see Section 2.6.4). Bivariate plots of the random coefficients (e.g. patient residuals against patient·time residuals) can be helpful for showing up any outlying random effect categories, and their influence may be assessed by removing them from the model. However, if predictions of random effects are of interest, it should be borne in mind that these will be more sensitive to any non-normality in their distribution.

6.5.3 Comparisons with fixed effects approaches

In this section, we consider two fixed effects approaches that have been used for modelling linear relationships with time for repeated measures data and show how they differ from the random coefficients model.

The first approach is an extremely statistically naive one, but one which appears from time to time in the medical literature. Treatment, time and treatment·time are fitted as fixed effects, and the effects of patient and patient·time are totally ignored. Thus, all observations are treated as independent, and the standard errors of intercept and slope·effects will be erroneously small because they take no account of between-patient variability.

The second fixed effects approach is more robust. A two-stage model is used: first, time slopes are calculated for each patient; then, an analysis is performed on the slope estimates (e.g. Rowell and Walters, 1976). This has an advantage over the first approach in that random variation in slope effects is allowed between patients. However, a drawback is that slopes estimated for patients with only a few

observations can be unrealistic, and the slopes from all patients are given equal weight regardless of their accuracy. A suitably weighted analysis could, of course, be considered, but it will usually be simpler and more efficient to apply a random coefficients model.

When a random coefficients model is used, the problems encountered with the fixed effects approaches do not occur. The standard errors of intercept and slope effects take into account between-patient variability; shrunken random coefficients estimates avoid the problem of unrealistic slope estimates that may occur when there are only a few observations per patient, and slopes from individual patients are appropriately weighted.

6.6 Examples of random coefficients models

In this section, we will present two examples. In the first of these, we will fit (after a transformation) a linear model. This corresponds to the most common type of random coefficients model that is likely to be encountered in practice. In the second example, we examine the use of polynomial models. Only in this example do we go through in detail the model checking procedures, which we recommend be undertaken in all analyses.

6.6.1 A linear random coefficients model

One of the measures of disease severity in patients with HIV infections has been the CD4 count. This was found to decline with the time since infection with HIV. Many patients with haemophilia became HIV positive, often because of receiving infected blood products, but it was reported that treatment of their haemophilia with high-purity monoclonally immunopurified factor VIII concentrates could slow or halt their rate of decline in CD4 count. In Britain, such patients were treated for some years with high-purity factor VIII concentrate, but some centres had been supplied with concentrate that was monoclonally immunopurified, while others used concentrates that were ion-exchange purified. This permitted a 'natural experiment' that was monitored prospectively from the time of a patient's transfer from intermediate-purity factor VIII concentrate to high-purity concentrate, for a period of 3 years (Hay *et al.*, 1998).

Various transformations of the CD4 count have been proposed in the literature. In this case, in agreement with most authors, a square root transformation was found to give approximately normal distributions and to produce linear rates of change over time. CD4 counts were taken at entry to the study (i.e. at the time of change to one of the forms of high-purity factor VIII concentrate) and at approximately 6-monthly intervals thereafter for a period of 3 years. The exact times when the samples were taken were used in the following analysis.

Of the 116 patients with severe haemophilia A who entered the study, 79 received the monoclonally immunopurified product, while 37 received ion-exchange-purified factor VIII. By the end of the study, 25 patients had died, 15 in the monoclonal group and 10 in the ion-exchange group, and three patients were lost to follow-up. One patient died before any post-treatment CD4 counts were obtained, but the data from all other patients were used in the analysis.

The median CD4 counts at entry to the study were $0.30 \times 10^9/1$ in the monoclonal group and $0.16 \times 10^9/1$ in the ion-exchange group. From a design viewpoint, this difference was unhelpful but was a consequence of the absence of randomisation. One of the centres using the ion-exchange product had patients who had been infected earlier, with consequently lower CD4 counts. The final CD4 counts had median values of 0.16 versus $0.08 \times 10^9/1$. The median reductions were $0.08 \times 10^9/1$ in the monoclonal group and $0.03 \times 10^9/1$ in the ion-exchange group.

In the analysis presented below using a random coefficients model, no covariates are fitted. The terms fitted are therefore

Fixed effects: time, treatment and treatment·time,
Random coefficients: patient and patient·time.

The results of fitting the model are shown in Table 6.4. The principal interest will be in the magnitude of the treatment·time interaction and its level of significance because this indicates whether the two treatments differ in their effect on the rate of decline of CD4 counts. The estimate is similar in magnitude to its standard error, indicating an absence of statistical significance ($p = 0.25$). The treatment estimate is just over twice its standard error, and this is statistically significant at conventional levels ($p = 0.03$). This test is of little interest to us, however, as this simply confirms that the CD4 counts in the monoclonal group are higher than those in the ion-exchange group, as was noted in the CD4 levels prior to the commencement of either form of high-purity factor VIII concentrate.

The terms in the **G** matrix are positive along the diagonal, but there is a negative covariance term. The diagonal terms show that there is between-patient variation in the slopes and intercepts, as we would expect. The negative covariance term is

Table 6.4 Results from analysis of change in CD4 counts.

	Fixed effects (SE)	**Covariance parameters in G matrix**	**Residual**
Intercept	0.523 (0.026)	$\begin{pmatrix} 0.0509 & -0.0003 \\ -0.0003 & 0.0027 \end{pmatrix}$	0.0073
Treatment	−0.102 (0.047)		
Time	−0.050 (0.008)		
Treatment·time	0.016 (0.014)		

not surprising because of the well-known negative correlation between estimates of slopes and intercepts in regression analysis. However, it is perfectly possible in random coefficients models for covariance terms to be positive on occasions.

In reporting a random coefficients analysis, it is helpful to give the point estimates for average change in each treatment group, together with their standard errors or confidence intervals. If the analysis is being undertaken using SAS, these are not immediately available but can soon be obtained. The SAS output, given in more detail at the end of this chapter, has the following section:

Solution for Fixed Effects

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		0.5228	0.02633	112	19.86	<.0001
treat	1	-0.1025	0.04708	112	-2.18	0.0316
treat	2	0	.	.	.18	.2418
TIME		-0.04962	0.007597	103	-6.53	<.0001
TIME*treat	1	0.01583	0.01356	107	1.17	0.2459
TIME*treat	2	0	.	.	.18	.2418

The line labelled TIME gives the mean rate of decline and standard error in the 'reference' category for treatment, which in this case is treatment group 2 (monoclonal). Thus, the mean rate of decline is 0.050(CD4 count)$^{1/2}$ per year with a standard error of 0.008. To obtain the corresponding figures for the ion-exchange group, the program needs to be rerun with the labelling of the groups reversed to give

Solution for Fixed Effects

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		0.4203	0.03903	113	10.77	<.0001
treat	1	0.1025	0.04708	112	2.18	0.0316
treat	2	0	.	.	.24	.2418
TIME		-0.03379	0.01124	108	-3.01	0.0033
TIME*treat	1	-0.01583	0.01356	107	-1.17	0.2459
TIME*treat	2	0	.	.	.28	.2418

It will be seen that the interaction term remains unchanged, apart from its sign, but the row for TIME now gives the rate of decline and standard error for the ion-exchange group. This is 0.034(CD4 count)$^{1/2}$ per year with a standard error of 0.011.

Examinations of residuals are not shown, but they indicated that the model produced an acceptable fit.

6.6.2 A polynomial random coefficients model

In this example, antibody levels to a herpes virus were measured in 45 children with one of two types of cancer: solid lump tumour (18) or leukaemia (27). The measurements were taken during hospital visits for courses of chemotherapy treatment. The duration of treatment ranged from 1 month to 3 years (median 12 months), and the intervals between treatments differed between the children. The aim of the study was to establish whether virus antibody levels were affected by chemotherapy treatment and whether this change was related to cancer type. It is known that the herpes virus antibody is present in nearly all children (all children in this study had it), and its average level decreases as children become older. Virus antibody levels for individual patients are plotted on the next page. These indicate that levels of the antibody fluctuate widely in some children (see Figure 6.4). The relationship of antibody level with time can be assessed by using a random coefficients model. This will help to determine whether chemotherapy is having an adverse effect on virus levels.

Building a polynomial model

As the relationship of virus level with time may be non-linear, a model using polynomials of time will be considered. However, note that in many applications, the linear model described in Section 6.6.1 is deemed sufficient to assess whether there is a time trend or whether the trend is varying across treatment groups.

The model is built up by adding polynomials of increasing order one by one into the model, both as fixed effects and as random coefficients. Random coefficients will be retained, provided their variance components are positive. However, fixed effects polynomials are added until they are non-significant. The age distribution will change at different treatment durations depending on which patients have observations present. To help overcome this variation, age at the start of treatment is included in all models. The following linear model (Model 1) is fitted initially.

Fixed effects: type (solid lump or leukaemia), age and time,
Random coefficients: patient (intercepts) and patient·time (slopes).

Results from this analysis are shown in Table 6.5. The variance components corresponding to the two random coefficients are both positive. This indicates that there is more variation between the regression lines for each patient than expected by chance, that is patients vary in their rates of virus decay. Next, quadratic time effects are added into the model (Model 2):

Fixed effects: type, age, time and time^2,
Random coefficients: patient, patient·time and patient·time^2.

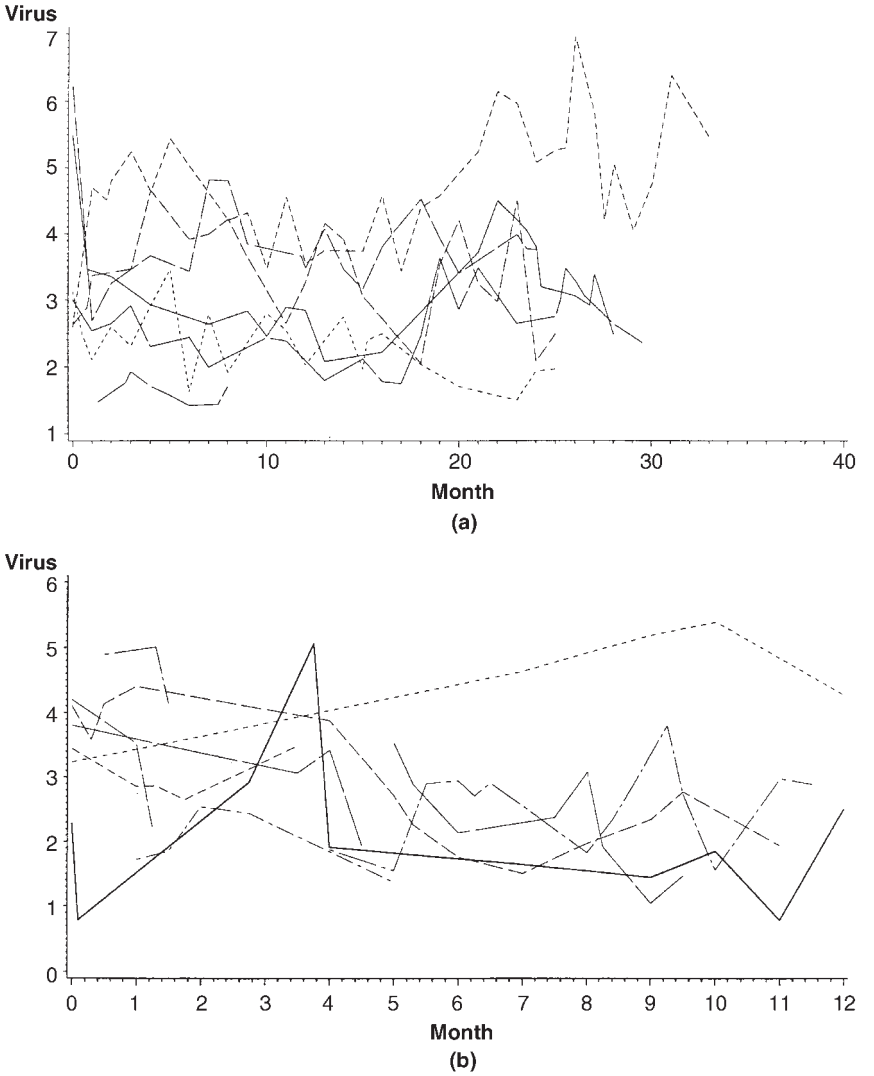


Figure 6.4 Examples of individual patient profiles. (a) Type = AL. (b) Type = ST.

The three variance components obtained from this model were all positive, indicating more variation between the quadratic curves for each patient than expected by chance.

When cubic time effects were added as random coefficients, the model did not converge (using `PROC MIXED`). Results provided at the last iteration indicated

Table 6.5 Results from Models 1–3.

Model	Fixed effects		G matrix and residual
1 (linear)	Intercept	3.65 (0.24)	$\begin{pmatrix} 0.44 & \\ 0.013 & 0.0042 \end{pmatrix}$
	Type	−0.23 (0.25)	
	Age	−0.046 (0.038)	
	Time	−0.032 (0.014)	0.56
2 (quadratic)	Intercept	3.70 (0.25)	$\begin{pmatrix} 0.59 & & \\ -0.043 & 0.025 & \\ 0.0016 & -0.0007 & 0.00002 \end{pmatrix}$
	Type	−0.08 (0.26)	
	Age	−0.051 (0.039)	
	Time	−0.081 (0.031)	
	Time^2	0.0025 (0.0011)	0.53
3 (cubic)	Intercept	3.74 (0.25)	$\begin{pmatrix} 0.60 & & \\ -0.045 & 0.024 & \\ 0.0017 & -0.0007 & 0.00002 \end{pmatrix}$
	Type	−0.060 (0.26)	
	Age	−0.049 (0.039)	
	Time	−0.118 (0.036)	
	Time^2	0.0065 (0.0026)	0.53
	Time^3	−0.000 11 (0.000 06)	

that the patient·time^3 variance component estimate was becoming negative. Attempts to confirm this by using the `NOBOUND` option in the `PROC MIXED` statement also failed to obtain convergence. Thus, only random coefficients up to the quadratic term are considered. However, a fixed cubic effect can still be included (Model 3):

Fixed effects: type, age, time, time^2 and time^3,
Random coefficients: patient (intercepts), patient·time and patient·time^2.

The fixed cubic effect was almost significant ($p = 0.06$) when reported in the first edition of this book. The inclusion of the `KENWARDROGER` option to reduce standard error bias has increased this to $p = 0.10$. We would usually disregard the cubic term on this basis, but for consistency with the first edition, it will be retained.

Tests of interactions with type and age

Interactions were tested between time effects (time, time^2 and time^3) and cancer type and found to be non-significant. Therefore, we can be reasonably confident in assuming that changes in the virus antibody level are similar for the two types of cancer.

The age effect was not significant in any model. However, since it is known that average antibody levels decrease with age, it is important to retain it. Interactions with age were found to be non-significant, and so we assume that the pattern of antibody change over time is unrelated to age.

Checking model assumptions

The random coefficients model assumes that the residuals have a normal distribution and that the random coefficients have a multivariate normal distribution. We will look at plots of the residuals and random effects for Model 3.

Residuals Residuals are checked by plotting them against their predicted values and by using a normal plot (Figure 6.5). A plot against time is also used to determine whether the residual variance varies over time.

These plots show one potential outlying observation with a residual of over three (patient 67, month 0). On closer examination, it appeared possible that this was caused by a recording error (the recorded value of 7.22 was well above all other values for this patient, and it seemed likely that the true value was 1.22). On removal of this observation and refitting Model 3, the fixed effects results changed fairly noticeably (Table 6.6). Since a recording error appeared likely, we will base our conclusions on the analysis with the outlier removed.

Random coefficients The three sets of random coefficients are plotted against each other to check for any outlying patients (Figure 6.6). These plots indicate no marked deviation from bivariate normal distributions. However, as for random effects models (see Section 2.4.6), the plots will not always show up deviations from normality. The patient time and patient·time^2 residuals are highly correlated, as expected from their correlation of -0.98 (calculated from the covariance parameters by $-0.000\,709/(0.0236 \times 0.000\,0224)^{1/2}$).

Plot of predicted virus antibody level

The results become more meaningful when the predicted antibody level is plotted against time. This is done for Models 2 and 3 with the outlying observation removed (Figure 6.7). The quadratic model (Model 2) curve is also plotted to make the conclusions more robust since the cubic coefficient is of dubious significance.

The curves for the two models differ markedly for higher values of time. However, only a small proportion of observations was made after 24 months (5%), and the cubic coefficient will be based largely on these. The models are therefore only really plausible up to 24 months, for which they are quite similar. The virus antibody levels decrease most rapidly initially and then flatten. The divergent curves illustrate how overinterpretation might occur when part of them is only based on a small amount of data.

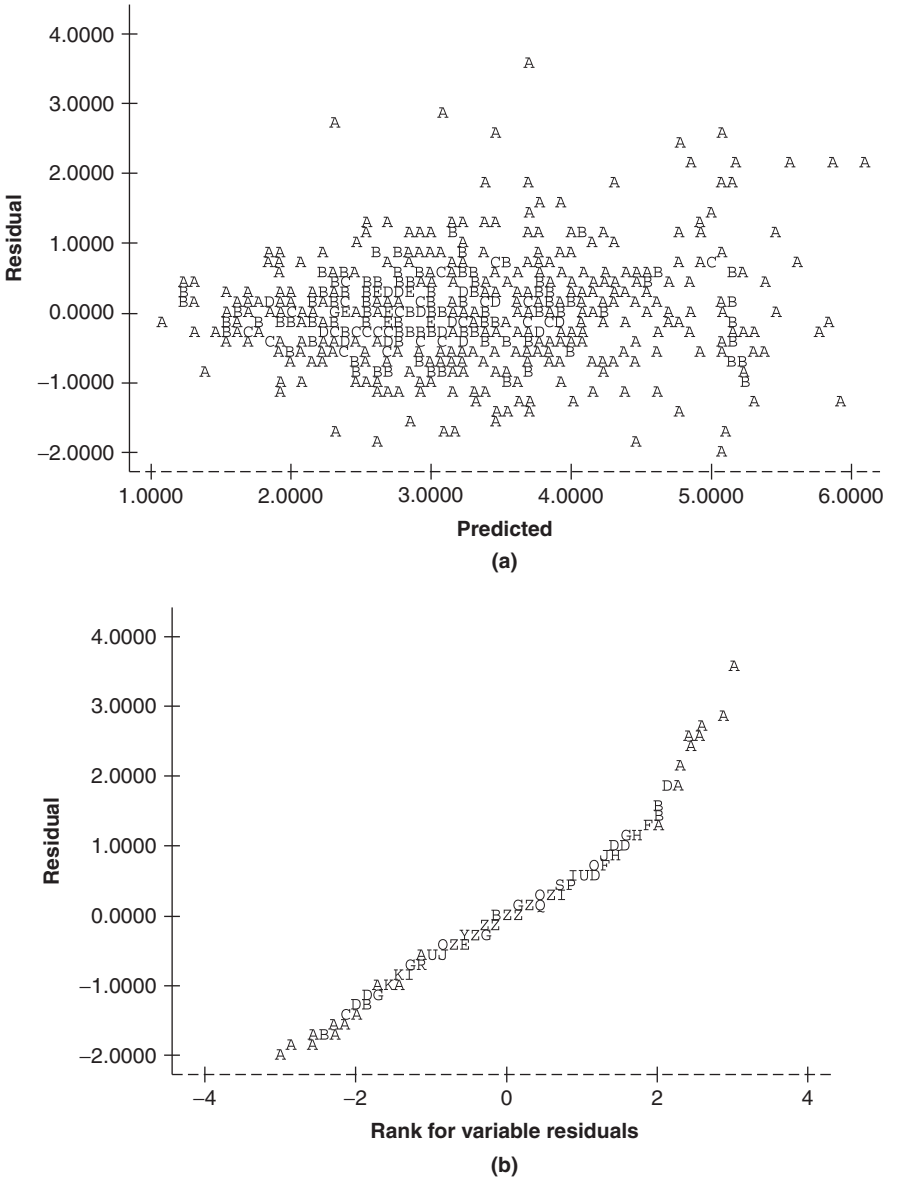


Figure 6.5 Plots of (a) residuals against their predicted values (b) normal plot (c) residuals against time (months). A = 1 obs., B = 2 obs., etc.

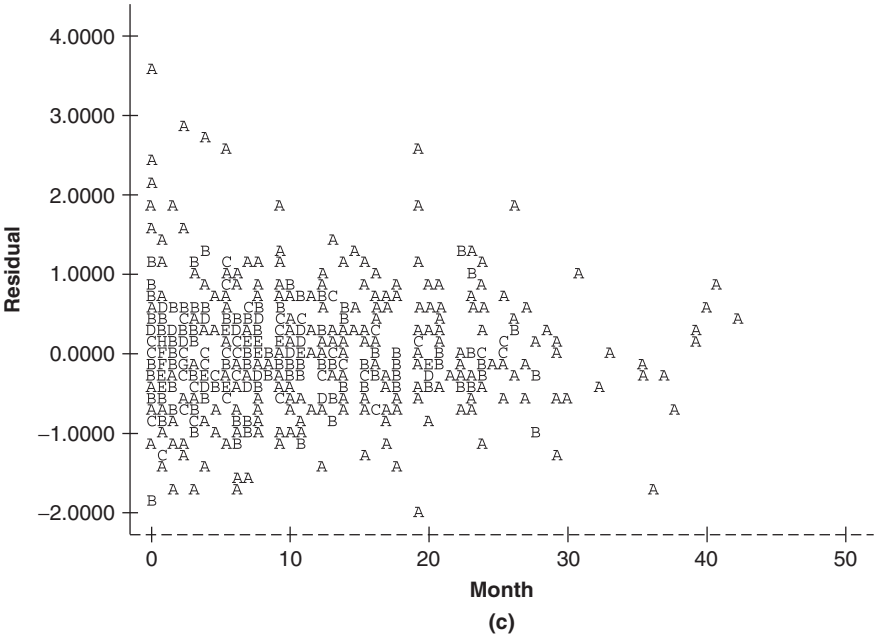


Figure 6.5 (*Continued*)

Table 6.6 Results from Model 3 with and without the outlier.

Model	Fixed effects		G matrix and residual
3 – (without outlier)	Intercept	3.75 (0.26)	$\begin{pmatrix} 0.60 & & \\ -0.025 & 0.016 & \\ 0.0011 & -0.0005. & 0.00001 \end{pmatrix}$
	Type	−0.11 (0.26)	
	Age	−0.059 (0.041)	
	Time	−0.097 (0.033)	
	Time^2	0.0055 (0.0024)	0.50
	Time^3	−0.000 09 (0.000 06)	
3	Intercept	3.74 (0.25)	$\begin{pmatrix} 0.60 & & \\ -0.045 & 0.024 & \\ 0.0017 & -0.0007 & 0.00002 \end{pmatrix}$
	Type	−0.060 (0.26)	
	Age	−0.049 (0.039)	
	Time	−0.118 (0.036)	
	Time^2	0.0065 (0.0026)	
	Time^3	−0.000 11 (0.000 06)	0.53

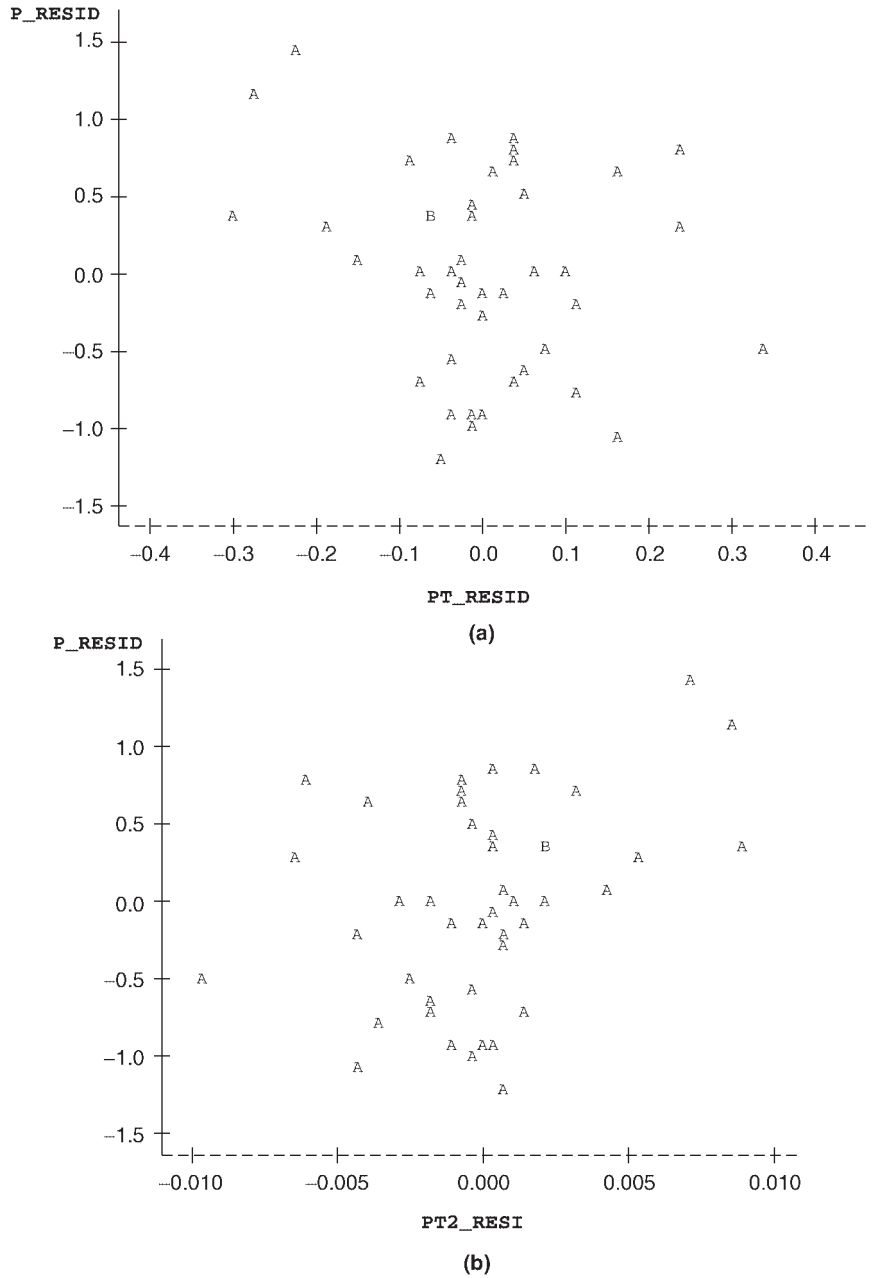


Figure 6.6 Plot of random coefficients: (a) patient coefficients vs. patient·time coefficients (b) patient coefficients vs. patient·time^2 coefficients (c) patient·time coefficients vs. patient·time^2 coefficients. A = 1 obs., B = 2 obs., etc.

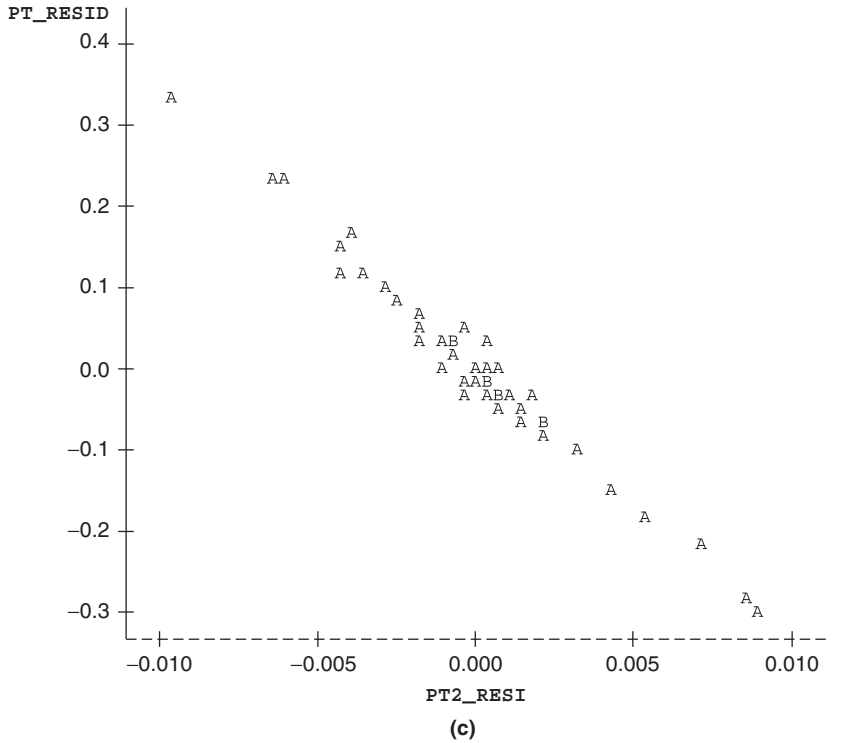


Figure 6.6 (*Continued*)

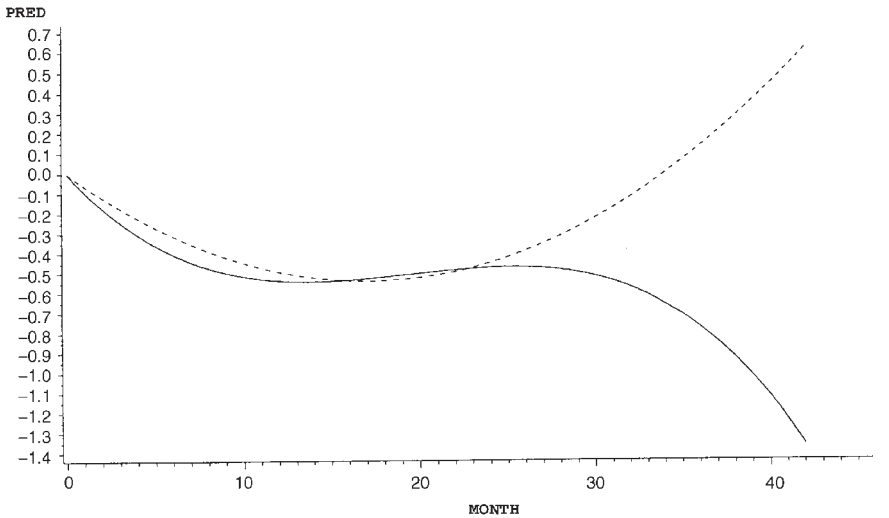


Figure 6.7 Predicted virus antibody level vs. time (Models 2 and 3). Curve: – – cubic; – – – –, quadratic.

Table 6.7 Results from Model 1 with time centred about its mean of 10 months.

Model	Fixed effects		G matrix and residual
1	Intercept	3.65 (0.24)	$\begin{pmatrix} 0.44 & \\ 0.013 & 0.0042 \end{pmatrix}$
	Type	−0.23 (0.25)	
	Age	−0.046 (0.038)	
	Time	−0.032 (0.014)	0.56
1 – with altered time origin	Intercept	3.34 (0.27)	$\begin{pmatrix} 1.13 & \\ 0.056 & 0.0042 \end{pmatrix}$
	Type	−0.23 (0.25)	
	Age	−0.046 (0.038)	
	Time	−0.032 (0.014)	0.56

Illustration of invariance to time origin

We pointed out in Section 2.1.4 that the fixed effects and **V** matrix estimates were invariant to the origin used for time. We illustrate this property by fitting Model 1 with time centred about its mean of 10 months (i.e. using time = time − 10). Results from this model are shown in Table 6.7.

Thus, the variance parameters connected with the patient random coefficients (intercepts) and the fixed intercept effect have altered to adjust for the change in time origin. However, the fixed effects and the other variance parameters are unaltered. The terms in $\mathbf{V} = \text{var}(\mathbf{y})$ will also be found to be identical between the models when calculated for specific values of time.

SAS code and output

Linear random coefficients model (6.6.1)

Variables
`patient` = patient number,
`treat` = type of factor VIII,
`time` = time in years from start of treatment,
`cd4_sqrt` = square root of CD4 count.

```
PROC MIXED NOCLPRINT DATA=cd4;
CLASS patient treat;
MODEL cd4_sqrt = treat time treat*time / SOLUTION
   DDFM=KENWARDROGER;
RANDOM INT time / SUBJECT=patient TYPE=UN SOLUTION;
TITLE 'SQUARE ROOT OF CD4 COUNTS OVER TIME';
TITLE3 'RANDOM COEFFICIENTS MODEL';
```

The term INT in the RANDOM statement is a SAS-reserved term for an intercept effect.

The use of the RANDOM statement to fit patient and patient·time effects as random coefficients is not immediately obvious. Specification of patient as a SUBJECT effect (SUBJECT = patient) blocks the **G** matrix by patients and causes interactions between the effects specified (INT and time) and patient to be fitted as random coefficients (hence patient and patient·time are fitted). The TYPE = UN option causes the random coefficients specified to have a multivariate normal distribution (i.e. a general covariance structure). In this example, the distribution will be bivariate normal as only two random coefficients are specified.

SQUARE ROOT OF CD4 COUNTS OVER TIME
RANDOM COEFFICIENTS MODEL
Iteration History

Iteration	Evaluations	-2 Res Log Likelihood	Criterion
0	1	30.49214157	
1	3	-828.61051673	0.00283032
2	1	-831.97977353	0.00024121
3	1	-832.24370131	0.00000248
4	1	-832.24627574	0.00000000

Convergence criteria met.
Covar iance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	PATIENT	0.05090
UN(2,1)	PATIENT	-0.00026
UN(2,2)	PATIENT	0.002740
Residual		0.007347

UN(1,1) and UN(2,2) are the variance component estimates for the patient and patient·time random coefficients. UN(2,1) is the covariance between the random coefficients. Note that the relative sizes of the patient·time variance component cannot be compared directly with the residual because it involves time. In this analysis, all the variance components are positive. However, in the situation where a variance component is negative, SAS would not converge and the variance component estimates output from the final iteration would usually show that one variance component estimate was becoming very close to zero.

Fit Statistics

-2 Res Log Likelihood	−832.2
AIC (smaller is better)	−824.2
AICC (smaller is better)	−824.2
BIC (smaller is better)	−813.3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	862.74	<.0001

Solution for Fixed Effects

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		0.5228	0.02633	112	19.86	<.0001
treat	1	−0.1025	0.04708	112	−2.18	0.0316
treat	2	0	.	.	5.55	5.5555
TIME		−0.04962	0.007597	103	−6.53	<.0001
TIME*treat	1	0.01583	0.01356	107	1.17	0.2459
TIME*treat	2	0	.	.	5.55	5.5555

Solution for Random Effects

Effect	PATIENT	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	101	0.1983	0.06494	435	3.05	0.0024
TIME	101	0.008769	0.03049	270	0.29	0.7739
Intercept	102	0.2988	0.05803	506	5.15	<.0001
TIME	102	0.03689	0.02791	336	1.32	0.1872
Intercept	103	0.05650	0.05946	453	0.95	0.3426
TIME	103	0.02604	0.03494	202	0.75	0.4570
Intercept	104	0.2385	0.05872	500	4.06	<.0001
TIME	104	−0.00048	0.02808	331	−0.02	0.9864

etc.

Note that the output immediately above has been generated by the use of the SOLUTION option in the RANDOM statement. The intercept and time terms do not give the intercepts and slopes directly. To achieve this, these terms would need to be added to the relevant fixed effects estimates.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	112	4.74	0.0316
TIME	1	107	37.81	<.0001
TIME*treat	1	107	1.36	0.2459

Polynomial random coefficients model (6.6.2)

Variables

`pat` = patient number,
`virus` = herpes antibody level,
`type` = illness type: A = acute leukaemia; S = solid lump tumour,
`month` = time, months since start of treatment,
`month2` = month^2,
`month3` = month^3.

SAS code for cubic model selected

```
PROC MIXED; CLASS type pat;
MODEL virus=age type month month2 month3/ S DDFM=KR;
RANDOM INT month month2/ SUB=pat TYPE=UN;
```

```
                    Iteration History
Iteration    Evaluations    -2 Res Log Likelihood     Criterion
0                      1           2062.27052887
1                      2           1601.35658286    0.00904154
2                      1           1598.68139127    0.00345910
3                      1           1597.70596681    0.00073106
4                      1           1597.51448892    0.00004682
5                      1           1597.50321956    0.00000027
6                      1           1597.50315684    0.00000000

                Convergence criteria met.
            Covariance Parameter Estimates
        Cov Parm         Subject             Estimate
        UN(1,1)              pat               0.5960
        UN(2,1)              pat              -0.04470
        UN(2,2)              pat              0.02356
        UN(3,1)              pat             0.001705
        UN(3,2)              pat              -0.00071
        UN(3,3)              pat             0.000022
        Residual                                0.5258
```

`UN(1,1)`, `UN(2,2)` and `UN(3,3)` are the variance component estimates for the patient, patient·time and patient·time^2 random coefficients. `UN(2,1)`, `UN(3,1)` and `UN(3,2)` are the covariances between the random coefficients. Note that the relative sizes of the patient·time and patient·time^2 variance components cannot be compared directly with other variance components because they involve time.

Fit Statistics

−2 Res Log Likelihood	1597.5
AIC (smaller is better)	1611.5
AICC (smaller is better)	1611.7
BIC (smaller is better)	1624.0

Null Model Likelihood Ratio Test

DF	Chi−Square	Pr > ChiSq
6	464.77	<.0001

Solution for Fixed Effects

Effect	type	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		3.7385	0.2506	50	14.92	<.0001
age		−0.04852	0.03947	43.3	−1.23	0.2255
type	AL	−0.06019	0.2553	43	−0.24	0.8147
type	ST	0	.	.	51.	.8147
month		−0.1182	0.03635	39.6	−3.25	0.0023
month2		0.006543	0.002553	123	2.56	0.0116
month3		−0.00011	0.000064	79.5	−1.69	0.0959

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	43.3	1.51	0.2255
type	1	43	0.06	0.8147
month	1	39.6	10.58	0.0023
month2	1	123	6.57	0.0116
month3	1	79.5	2.84	0.0959

SAS code for other models tested (without output)

```
PROC MIXED; CLASS type pat;
MODEL virus=age type month/ S DDFM=KR;
RANDOM int month/ SUB=pat TYPE=UN;
PROC MIXED; CLASS type pat;
MODEL virus=age type month month2/ S DDFM=KR;
RANDOM int month month2/ SUB=pat TYPE=UN;
```

```
PROC MIXED; CLASS type pat; MODEL virus=age type month month2
   month3 type*month type*month2 type*month3/S DDFM=KR;
RANDOM int month month2/ SUB=pat TYPE=UN;

PROC MIXED; CLASS type pat;
MODEL virus=age type month month2 month3 age*month age*month2
age*month3/S DDFM=KR;
RANDOM int month month2/ SUB=pat TYPE=UN;
```

Model checking using Model 3

```
PROC MIXED; CLASS type pat;
MODEL virus=age type month month2 month3/ S DDFM=KR
   OUTP=resid OUTPM=work.predm;
RANDOM int month month2/ SUB=pat TYPE=UN SOLUTION;
ID pat month;
ODS OUTPUT SOLUTIONR=solut;

PROC PRINT NOOBS DATA=resid; VAR pat month resid pred;
TITLE 'RESIDUALS AND PREDICTED VALUES';

PROC PLOT DATA=resid; PLOT resid*pred;
TITLE 'RESIDUALS AGAINST THEIR PREDICTED VALUES';
PROC PLOT DATA=resid; PLOT resid*month;
TITLE 'RESIDUALS AGAINST TIME (MONTHS)';
PROC RANK DATA=resid OUT=norm NORMAL=TUKEY;
   VAR resid; RANKS s_resid;
PROC PLOT DATA=norm; PLOT resid*s_resid;
TITLE 'RESIDUALS - NORMAL PLOT';

DATA solut; SET solut;
patx=pat*1; * obtain numeric patient variable;
DROP pat;

DATA p_resid(KEEP=pat p_resid) pt_resid(KEEP=pat pt_resid)
   pt2_resi(KEEP=pat pt2_resi); SET solut;
   pat=patx;
IF effect='Intercept' THEN DO;
   p_resid=Estimate;
   OUTPUT p_resid;
```

```
END;
ELSE IF effect='month2' THEN DO;
   pt2_resi=Estimate;
   OUTPUT pt2_resi;
END;
ELSE DO;
   pt_resid=Estimate;
   OUTPUT pt_resid;
END;
PROC SORT DATA=predm; BY pat;
PROC MEANS NOPRINT DATA=predm; BY pat;
VAR pred; OUTPUT OUT=predm MEAN=p_pred N=freq;

DATA a; MERGE p_resid pt_resid pt2_resi predm; BY pat;
PROC PRINT NOOBS; VAR pat p_resid pt_resid pt2_resi
   p_pred freq;
TITLE 'RANDOM COEFFICIENTS AND PREDICTED VALUES
   FOR EACH PATIENT';
PROC PLOT; PLOT p_resid*pt_resid;
TITLE 'PATIENT COEFFICIENTS VS PATIENT.TIME COEFFICIENTS';
PROC PLOT; PLOT p_resid*pt2_resi;
TITLE 'PATIENT COEFFICIENTS VS PATIENT.TIME2 COEFFICIENTS';
PROC PLOT; PLOT pt_resid*pt2_resi;
TITLE 'PATIENT.TIME COEFFICIENTS VS PATIENT.TIME2
   COEFFICIENTS';
```

This code may not at first sight be straightforward to understand. The steps used are summarised as follows:

1. Fit Model 3.
2. Use option `OUTP` to output the residuals (and predicted values given by $\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}}$ which are not required here) to dataset `resid`. Use `OUTPM` to output the predicted values given by $\mathbf{X}\hat{\boldsymbol{\alpha}}$ to dataset `predm`. Use `ODS OUTPUT` statement to output the random effects estimates to dataset solut. The `ID` statement causes the `pat` and `month` variables to be included in the datasets `resid` and `predm`.
3. Produce a print and plots of residuals.
4. Create datasets `p_resid`, `pt_resid` and `pt2_resid` containing random coefficients estimates for the patient and patient·time and patient·time^2 effects, respectively.
5. Obtain predicted means for each patient (based on the predicted values $\mathbf{X}\hat{\boldsymbol{\alpha}}$) and merge these with the datasets of random coefficients.
6. Produce print and plots of the random coefficients.

RESIDUALS AND PREDICTED VALUES

pat	month	Resid	Pred
3	0.00	1.55766	3.93234
3	0.75	−0.33901	3.80413
3	2.00	−0.25526	3.61185
3	4.00	−0.42664	3.35687
3	7.00	−0.45030	3.08596
3	9.00	−0.13501	2.97222
3	10.00	−0.47612	2.93348
3	11.00	−0.01450	2.90597
3	12.00	−0.04408	2.88904
3	13.00	−0.80454	2.88206
3	16.00	−0.69730	2.91435
3	21.00	0.61393	3.10700
3	22.00	1.33944	3.16056
3	23.00	0.99236	3.21764
3	23.50	0.80270	3.24730
3	24.00	0.52241	3.27759
3	24.25	−0.09292	3.29292
3	26.00	−0.34351	3.40351
3	26.75	−0.53195	3.45195
3	27.00	−0.07820	3.46820
3	28.00	−1.05319	3.53319
5	0.00	0.30924	2.82254
5	0.50	−0.29235	2.78072
ETC			

RANDOM COEFFICIENTS AND PREDICTED VALUES FOR EACH PATIENT

pat	p_resid	pt_resid	pt2_resi	p_pred	freq
3	0.35106	−0.05926	0.002288204	3.07590	21
5	−0.71021	0.03226	−.001866166	3.05075	22
13	0.87930	0.04089	0.000270946	3.06776	32
61	−0.21678	0.11842	−.004333146	3.07713	22
63	0.72495	0.03671	−.000784259	2.82919	20
65	0.70394	−0.09316	0.003276862	2.82020	10
67	0.38619	−0.30506	0.009039204	2.93704	9
69	−0.27759	−0.00126	0.000636630	2.72964	26
71	0.64374	0.16498	−.004001423	2.90710	25
73	−0.13393	−0.05631	0.001523998	3.39389	7
77	1.17548	−0.27558	0.008724335	3.46298	2
79	0.49003	0.04483	−.000441731	3.04050	24

All the residual plots appear in the main text.

Note that Figure 6.6(a) and Figure 6.6(b) could be achieved in SAS version 9.3 as part of a residual panel by inserting `PLOTS=RESIDUAL` in the `PROC MIXED` statement, but the other graphs cannot be obtained in a simple way.

6.7 Sample size estimation

Sample sizes for repeated measures studies are often calculated as if a simple between-patient trial with no repeated measurements was planned. However, it is possible to take into account the correlation that occurs between the repeated observations in the sample size estimate. This will lead to a smaller sample size than that calculated for a simple between-patient study. It therefore seems desirable ethically and on the grounds of cost that correlations within patients are taken into account. Obviously, the covariance pattern of the data will not be known in advance, but the assumption of a constant correlation between patients (compound symmetry pattern) is likely to be adequate. When no previous estimate of the within-patient correlation is available, a conservative prediction of the correlation could be used (i.e. a higher correlation than anticipated).

6.7.1 Normal data

To obtain a formula for sample size estimation, we require the variance of the mean of measurements on individual patients. The variance of the sum of the observations on any single patient, i, is

$$\begin{aligned}\text{var}\left(\sum_j y_{ij}\right) &= m\text{var}(y_{ij}) + m(m-1)\text{cov}(y_{ij}, y_{ik}) \\ &= m\sigma^2 + m(m-1)\rho\sigma^2 \\ &= m\sigma^2[1 + (m-1)\rho].\end{aligned}$$

This gives the variance of each patient mean as

$$\text{var}(\overline{y}_i) = \sigma^2[1 + (m-1)\rho]/m,$$

where

m = number of repeated measures,
σ^2 = between-patient variation (sum of variance parameters when compound symmetry pattern fitted using `PROC MIXED`),
ρ = correlation between observations on same patient (compound symmetry variance parameter divided by sum of variance parameters in `PROC MIXED`).

Using the usual sample size estimation equation,

$$\Delta = (z_{1-\alpha/2} + z_{\beta}) \times \mathrm{SE}(t_i - t_j),$$

we obtain the number of patients required per group as

$$n = 2(z_{1-\alpha/2} + z_{\beta})^2 \sigma^2 [1 + (m-1)\rho]/m\Delta^2,$$

where

α = significance level,
β = power,
Δ = difference to be detected,
t_i = ith treatment effect.

If a small trial is planned, for example with less than about 10 patients per group, a more accurate sample size could be obtained by substituting t statistics for the z statistics (with DF equal to the number of patients minus the number of treatment groups).

Example

The analysis of the repeated DBP measurements in the hypertension trial using a compound symmetry covariance pattern model gave a residual variance of $\sigma^2 = 76$, and the repeated measures had correlation $\rho = 0.53$ (see Section 6.3). We calculate the sample size required for a future study involving four post-treatment visits required to detect a difference in DBP of 5 mm Hg at the 5% significance level with 80% power. The number of patients required per treatment group is

$$\begin{aligned} n &= 2(1.96 + 0.84)^2 \times 76 \times (1 + 3 \times 0.53)/(4 \times 25) \\ &= 31. \end{aligned}$$

Had no account been taken of the repeated measurements, then n would have been

$$\begin{aligned} n &= 2(z_{1-\alpha/2} + z_{\beta})^2 \sigma^2/\Delta^2 \\ &= 2(1.96 + 0.84)^2 \times 76/25 \\ &= 48. \end{aligned}$$

If there is flexibility in the number of repeated measurements, then it might be worth calculating sample sizes for varying numbers of repeated measurements. For example, if 10 repeated measures were used, then

$$\begin{aligned} n &= 2(1.96 + 0.84)^2 \times 76 \times (1 + 9 \times 0.53)/(10 \times 25) \\ &= 28. \end{aligned}$$

However, this reduction in the number of patients required would be unlikely to justify the use of six additional repeated measures.

6.7.2 Binary data

When the variable of primary interest is binary, the sample size formula can be adapted to incorporate the binomial variance and may be written as

$$n = \frac{(Z_{1-\alpha} + Z_{\beta})^2(p_1(1-p_1) + p_2(1-p_2))(1 + (m-1)\rho)}{m\Delta^2},$$

where p_1 and p_2 are the two group proportions, and ρ is the expected correlation between the repeated observations on individuals.

Alternatively, if each observation is a number of events out of a total rather than a binary observation, the formula may be written as

$$n = \frac{(Z_{1-\alpha} + Z_{\beta})^2(p_1(1-p_1) + p_2(1-p_2))(\phi + N(m-1)\rho)}{mN\Delta^2},$$

where there is assumed to be a fixed denominator (N) for each observation and ϕ is a dispersion parameter specifying any extra-binomial variation between the observations (see Section 3.1.4).

Example

We will now assume (unrealistically) that the incidence of the adverse event, cold feet, is the primary endpoint. A doubling in the proportion of cold feet is to be detected from 0.1 to 0.2, and thus the required difference is 0.1. The study should have sufficient power to detect a difference at the 5% significance level with 80% power. An estimate of the correlation between observations on the same patient is obtained by analysing our example data (Section 3.4) using `PROC GENMOD` with a `REPEATED` statement and a compound symmetry covariance structure. This model provided a correlation of 0.48. The number of patients required per treatment group is

$$n = \frac{(1.96 + 0.84)^2(0.1 \times 0.9 + 0.2 \times 0.8)(1 + 3 \times 0.48)}{4 \times 0.1^2}$$
$$= 119.6 = 120 \text{ (after rounding up).}$$

6.7.3 Categorical data

Sample size estimation is always difficult when the variable of interest is categorical. If there are more than about five categories, then the formula for continuous data is likely to provide a reasonable approximation. In other situations, the best approach might be to partition the categories and use the formula for binary data.

7

Cross-over trials

7.1 Introduction

In earlier chapters, we have considered parallel group designs, where each subject is randomised to receive one of a number of alternative treatments. In contrast, in cross-over trials, subjects are randomised to receive different sequences of treatments, with the outcome being assessed for each treatment period. As before, we have a choice in analysis between fixed effects models and random effects models. In this context, we describe the treatment effect as being *crossed* with a random effect (subjects).

The vast majority of cross-over trials that are carried out in practice have the same basic design. Every subject receives each of the treatments being evaluated, for a standard period of time, with the outcome variables being assessed in the same way in each period of treatment. The simplest and most commonly encountered such design employs just two treatments and is often referred to as a 2×2 cross-over trial or as an AB/BA design. The use of this design with normally distributed data will be covered in some depth in Section 7.3. The use of more than two treatments with patients receiving every treatment is known as a higher order complete block design and is covered in Section 7.4. More complicated designs are considered in Sections 7.5 and 7.6. In Section 7.7, we will show how covariance pattern models can be employed in the analysis of cross-over trials. The following two sections (7.8 and 7.9) will give examples of the analysis of binary data and categorical data in the setting of cross-over trials. Data following Poisson distributions are not directly covered but follow the same generalised linear mixed model approach used for binary data. Section 7.10 will consider the use of information from random effects models in the planning of future studies. The chapter finishes with a discussion of some general points in relation to the analysis of cross-over trials (Section 7.11).

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

7.2 Advantages of mixed models in cross-over trials

Random effects models can be expected to give more precise estimates of treatment effects in situations where it is possible to recover extra information on treatments from the between-patient error stratum. The most common situation in which this occurs is where there are missing data, irrespective of the particular cross-over design used. It also occurs for the unbalanced designs considered in Sections 7.5 and 7.6. In the balanced situations, which we meet with complete block designs, the results of a fixed effects analysis and a random effects analysis will generally be identical in the absence of missing data.

A very different application of mixed models to cross-over trials arises from the covariance pattern approach. By regarding the results in successive treatment periods as a form of repeated measures data, we can examine various ways to model the covariance between repeated observations on the same patients. This can lead to greater flexibility in the interpretation of the data than with conventional analyses, and we examine examples using this approach in Section 7.7.

7.3 The AB/BA cross-over trial

This design employs two treatments (A and B) and two treatment periods. Patients are randomised to receive either the AB sequence of treatments or the BA sequence. We met a simplified hypothetical example of such a trial in Section 1.2. At that time, for simplicity of presentation, we assumed that there was no effect of the period in which treatments were received. However, such an effect is always possible, and we recommend that such an 'order' effect should be included in the analysis. Our initial example was also restricted to single observations in each treatment period. In practice, the randomisation to the AB or BA sequence is often preceded by a run-in period. This approach has the advantage that patients showing poor compliance can be removed prior to randomisation, and the stability of the patient's condition can be assessed. With or without this run-in period, baseline levels for the outcome variables are usually recorded prior to randomisation. Following the first treatment period, there is often a 'washout' period prior to the commencement of the second treatment, and a second 'baseline' observation may be made. Details of design considerations, and analytical methods for a fixed effects analysis, are given in Senn (2002).

If all patients complete the trial without any missing values being generated for the outcome variables, the results of the fixed effects analysis and an analysis in which patient effects are regarded as random will usually be identical. This arises because of balance over random effects in the design, as discussed in Section 1.6. However, note that an exception occurs when the estimate of the patient variance component is negative and is set to zero. The standard errors of the treatment differences will then be lower with the random effects model. It is common for

missing values to occur, usually because of premature patient withdrawal from the trial. In the fixed effects analysis of such a trial, observations from subjects with a missing value are not used because all of the information in the single remaining observation would be needed to estimate the patient effect. When missing values do occur, and a random effects analysis is performed, the data from subjects with a single period of observation are utilised in the analysis in conjunction with the complete observations to improve the efficiency of treatment comparisons relative to the fixed effects analysis. This benefit was illustrated in Section 1.2 using the earlier example but with two observations deleted.

More generally, we will now consider a cross-over trial to compare treatments A and B, with N patients divided equally between the AB and the BA sequence, following Brown and Kempton (1994). We will also assume that a proportion, p, of patients only provide data for the first treatment period. On the assumption that these dropouts are also equally divided between the two treatment sequences, we will investigate the effect of p and the variance components on the relative efficiency of the random effects and fixed effects models in estimating treatment differences. To do this, we will look first at the variance of the estimate of treatment differences, $\text{var}_{\text{W}}(\text{A} - \text{B})$, obtained from within-patient comparisons. This will give the variance appropriate to the fixed effects analysis. We will then look at the corresponding term from the between-patient comparison, using those patients who only have an observation in the first treatment period. We will then pool these two estimates and obtain the variance of the pooled estimate. In this situation, this will correspond to the results of fitting a random effects model, and we will compare the variances of the fixed effects and random effects model treatment estimates.

Within-patient comparisons From our definitions, there will be $N(1-p)$ patients with complete data. If the residual variance is σ_{r}^2, then

$$\text{var}_{\text{W}}(\text{A} - \text{B}) = \frac{2\sigma_{\text{r}}^2}{N(1-p)}.$$

Between-patient comparisons Each treatment sequence will yield $Np/2$ patients with observations in the first period only. The variance of individual observations will be the sum of the residual variance σ_{r}^2 and the between-patient component σ_{p}^2. Hence

$$\begin{aligned}\text{var}_{\text{B}}\,(\text{A} - \text{B}) &= \frac{2(\sigma_{\text{r}}^2 + \sigma_{\text{p}}^2)}{Np/2} \\ &= 4\sigma_{\text{r}}^2(1+\gamma)/Np,\end{aligned}$$

where

$$\gamma = \sigma_{\text{p}}^2/\sigma_{\text{r}}^2.$$

Pooled comparisons If we obtain a weighted average of the treatment effect from the within-patient and between-patient estimates, using weights inversely

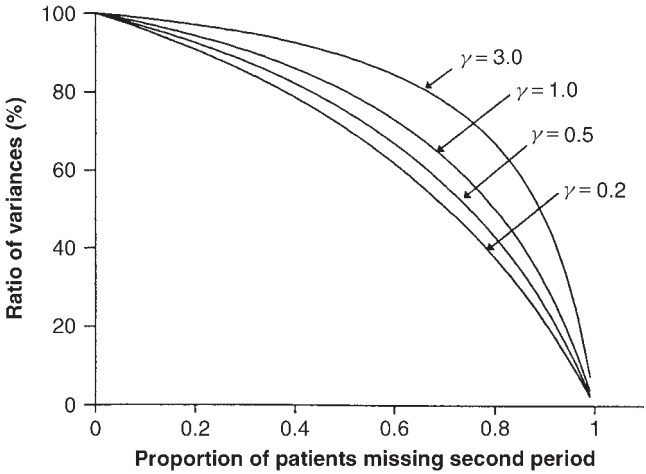


Figure 7.1 Ratio of variances of treatment differences, with and without recovery of between-patient information.

proportional to the variances, then we have the standard result that

$$\text{var}_\text{p}(\text{A} - \text{B}) = 1/(1/\text{var}_\text{W} + 1/\text{var}_\text{B}).$$

Thus,

$$1/\text{var}_\text{p}(\text{A} - \text{B}) = \frac{N(1-p)}{2\sigma_\text{r}^2} + \frac{Np}{4\sigma_\text{r}^2(1+\gamma)}$$

and

$$\text{var}_\text{p}(\text{A} - \text{B}) = \frac{4\sigma_\text{r}^2(1+\gamma)}{Np + 2N(1-p)(1+\gamma)}.$$

Relative efficiency The ratio of the variance of the treatment estimate using a fixed effects (within-patient) approach, to that using a random effects model (pooled), is plotted against p, the proportion of missing observations in the second period, for a range of values of γ, in Figure 7.1. From this figure, we can see that the recovery of between-patient information is most beneficial when γ is small, that is when the between-patient variance component is small. If the proportion of missing values is small, the benefit from analysis with a mixed model will be correspondingly small, although we can expect some reduction in the variance of the treatment estimate.

7.3.1 Example: AB/BA cross-over design

We illustrate the AB/BA cross-over design with results from an unpublished study comparing two diuretics in the treatment of mild to moderate heart failure.

After initial screening for suitability, there was a period of not less than 1 day and not more than 7 days, where diuretic treatment was withheld. Immediately prior to randomisation to either the AB sequence of treatment or the BA sequence, baseline observations were taken. Each treatment period lasted for 5 days, with an immediate transfer to the second treatment after the first treatment period was completed. As a washout was not employed between treatments, observations made in the first 2 days of each treatment period were not utilised in the analysis of the trial. The primary outcome measures were the frequency of micturition and the subjective assessment of urgency. As neither of these is suitable for illustrating the analysis of normally distributed data, we will instead use a secondary effectiveness variable, namely oedema status, together with diastolic blood pressure (DBP). Oedema status is formed by the sum of the left and right ankle diameters. The DBP was calculated from the mean of three readings. Both of these variables are measured prior to randomisation and at the end of each treatment period.

In total, 101 patients were recruited for the study, but seven withdrew prior to randomisation. Of the remaining 94 patients, only two failed to complete both treatment periods. Therefore, in order to illustrate the alternative methods of analysis, we have systematically removed approximately one in five of the observations from the second period. The structure of the data as analysed is shown in Table 7.1.

For each of our outcome variables, four analyses have been performed. In all of them, a treatment effect and a period effect were included as fixed. In two of the models, the baseline level was also included in the model as a covariate. Whether or not the baseline is included in the model, separate models are considered with

Table 7.1 Data structure for a cross-over trial comparing two diuretics in patients with heart failure.

			Baseline values		Post-treatment values	
Patient	**Treatment**	**Period**	**Oedema**	**DBP**	**Oedema**	**DBP**
1	B	1	45	60	45	55
1	A	2	45	60	45	60
2	A	1	51	50	48	60
2	B	2	51	50	48	65
3	A	1	53	70	50	70
3	B	2	53	70	52	80
4	B	1	49	68	47	60
4	A	2	49	68	47	60
5	A	1	46	65	45	60
6	A	1	61	95	60	95
6	B	2	61	95	59	97

Table 7.2 Analysis of cross-over trial of diuretics in heart failure.

Model	Fixed effects	Random effects
1	Treatment, period, patient	–
2	Treatment, period	Patient
3	Treatment, period, patient, baseline	–
4	Treatment, period, baseline	Patient

	Treatment effect: A–B (SE)	
Model	Oedema	DBP
1	0.304 (0.120)	0.812 (0.775)
2	0.301 (0.120)	0.926 (0.765)
3	0.304 (0.120)	0.812 (0.775)
4	0.309 (0.118)	1.013 (0.748)

	Variance components (SE)			
Model	Patient	Residual	Patient	Residual
1	–	0.530	–	22.19
2	66.825	0.530	76.77	22.25
3	–	0.530	–	22.19
4	3.763	0.526	25.60	21.91

the patient effect being fitted either as random or as fixed. The results of the models are summarised in Table 7.2.

Examination of the variance component terms shows that for all models the patient term is larger than the residual term. This indicates that there may have been substantial benefits from employing a cross-over design rather than a parallel group design. We note that this is particularly striking for oedema status. Note also the effect of including the baseline as a covariate in the analysis. This has the effect of reducing the size of the patient variance component term in Model 4. The implications of this are that the benefits of the cross-over are somewhat reduced when a (highly correlated) baseline covariate is available and, conversely, that the use of a mixed model is likely to be most helpful in these circumstances if there are missing values.

We see this in the estimates of the treatment standard errors. Comparison of Models 1 and 2 for the oedema status shows that the standard errors are identical (to the number of digits reported), indicating that the between-subject variation is so large that recovery of between-subject information is ineffective. With inclusion of the baseline level as a covariate, we see that Model 3 gives the same result as Model 1. This result is well known, showing that a single baseline has no effect on a

fixed effects analysis. It does, however, produce a small reduction in the treatment standard error when a mixed model is fitted, showing that some between-subject information has been utilised.

The results for DBP show the recovery of between-subject information more clearly because of the relatively smaller between-subject variation. We see a detectable reduction in the treatment standard error, even when baselines are not used, and with the inclusion of baselines, a reduction of about 4% in the standard error is seen with the mixed models approach. This gain is modest but worthwhile.

The greatest advantage of the mixed models approach will unfortunately be gained in situations where a cross-over trial shows little benefit over a parallel group study, that is where the between-subject variance component is small relative to the residual variance component.

Such a situation occurs in a trial reported by Jones and Kenward (1989). In this two-period, cross-over trial, an oral mouthwash was compared with a placebo mouthwash. There were two 6-week treatment periods, with a 3-week washout period separating them. The outcome variable reported was the average plaque score per tooth, with each tooth being assessed on an integer scale from zero to three. Results were presented for the 34 patients with data from both treatment periods. Interestingly, these data arose from a trial in which 41 patients were randomised, and 38 completed the trial. For the purposes of this illustration, we have deleted the second observation from five randomly selected patients from the 34 with complete data.

Two models were fitted to the data using `PROC MIXED`. In both, a treatment effect and a period effect were included as fixed. In one, the patient effect was fitted as random, and in the other, it was fitted as fixed. The results are shown in Table 7.3.

Examination of the variance component terms shows that the patient term is appreciably smaller than the residual term. This indicates that the benefit of employing a cross-over design rather than a parallel group design may be small. The estimate of the period effect (not shown) is small, but in accord with our recommendation in the previous section, we retain it in the model. The main interest, of course, lies in the estimates of the treatment difference and the associated standard errors. Both analyses demonstrate a clear advantage to using

Table 7.3 Analysis of oral mouthwash trial.

	Fixed patients	**Random patients**
Variance components		
Patients	–	0.029 (0.018)
Residual	0.069	0.066 (0.017)
Treatment difference (SE)	0.25 (0.069)	0.244 (0.065)

the active mouthwash. For our purposes in comparing the results of the two analytical strategies, it is the standard errors that interest us because it is purely a matter of chance which method gives the larger point estimate of the treatment effect. We see that the use of the random effects model has reduced the standard error of the estimate of the treatment difference by about 6%.

SAS code and output

The SAS code to generate analyses for oedema status (Table 7.2) is shown for Models 3 and 4. Models 1 and 2 differ only in the exclusion of the baseline value from the model.

```
PROC MIXED NOCLPRINT; CLASS treat period patient;
TITLE 'FIXED EFFECT ANALYSIS WITH BASELINE';
MODEL oed=treat period patient oedbase;
LSMEANS treat /DIFF PDIFF;

PROC MIXED NOCLPRINT; CLASS treat period patient;
TITLE 'RANDOM EFFECTS ANALYSIS WITH BASELINE';
MODEL oed=treat period oedbase / DDFM=KENWARDROGER;
RANDOM patient;
LSMEANS treat /DIFF PDIFF;
```

The output reproduced as follows is from Model 4:

```
                    Iteration History
Iteration       Evaluations         -2 Res Log        Criterion
                                    Likelihood
        0                 1       732.30378010
        1                 3       620.47659711       0.00104614
        2                 2       620.32698029       0.00001998
        3                 1       620.32375614       0.00000001

              Convergence criteria met
                Covariance Parameter
                     Estimates
            Cov Parm                  Estimate
            patient                     3.7632
            Residual                    0.5260
                     Fit Statistics
            -2 Res Log Likelihood        620.3
            AIC (smaller is better)      624.3
            AICC (smaller is better)     624.4
            BIC (smaller is better)      629.4
```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	75.3	6.79	0.0110
period	1	74.9	4.02	0.0487
oedbase	1	94.1	1433.62	<.0001

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	55.3621	0.2170	108	255.09	<.0001
treat	B	55.0536	0.2168	107	253.90	<.0001

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	B	0.3085	0.1184	75.3	2.61	0.0110

7.4 Higher order complete block designs

In these designs, there are as many treatment periods as there are treatments to be compared, and each patient receives every treatment. If there are no missing data, then a conventional least squares analysis fitting treatment, period and patient effects is fully efficient. Whenever there are missing data, some of the within-patient treatment comparisons are unavailable for every patient. Therefore, additional between-patient information can be utilised.

7.4.1 Inclusion of carry-over effects

In any cross-over trial, there is the possibility of carry-over effects. That is, the results in second or subsequent treatment periods may be influenced by treatment administered in earlier periods. In the simple two-period, cross-over trial considered previously, there is no possibility of estimating carry-over. In all of the remaining designs considered, carry-over effects can be estimated, and in our examples, we will consider results from models that include carry-over. However, we do this for completeness rather than in the belief that this is good practice, and we return to this point in Section 7.11.

7.4.2 Example: four-period, four-treatment cross-over trial

We consider the four-period, four-treatment cross-over trial described by Jones and Kenward (1989). Three drugs, A, C and D and a placebo B were compared to assess their effect on cardiac output, measured by the left ventricular ejection time (LVET). Each treatment was given for one week, with a one-week washout period between treatments. Observations were made at the end of each treatment period. Fourteen patients were used in the trial, yielding 56 observations. To demonstrate the use of the mixed models approach, we have arbitrarily set 13 of the 56 observations to be missing. The results of four analyses are presented in Table 7.4, from the combinations of inclusion or exclusion of carry-over effects, and handling the patient effects as fixed or random.

All pairwise comparisons of the carry-over effects were non-significant and will not be considered further or the details presented. We see that between-patient variation is moderate, being about 50% higher than the residual variance

Table 7.4 Estimates of variance components and treatment effects. Standard errors of estimates appear in brackets.

	Fixed patients	**Random patients**
Ignoring carry-over		
Variance components		
Patients	–	2721
Residual	1667	1657
Treatment differences		
A – B	77.4 (20.1)	72.5 (19.9)
A – C	36.8 (17.3)	32.8 (17.1)
A – D	77.3 (19.5)	74.6 (19.4)
B – C	−40.6 (19.9)	−39.7 (19.5)
B – D	−0.1 (21.2)	2.1 (21.0)
C – D	40.4 (18.7)	41.8 (18.5)
Including carry-over		
Variance components		
Patients	–	2750
Residual	1840	1831
Treatment differences		
A – B	76.0 (22.6)	69.1 (22.2)
A – C	40.5 (20.4)	32.1 (20.0)
A – D	84.9 (22.5)	79.0 (22.3)
B – C	−35.4 (23.0)	−37.0 (22.4)
B – D	8.9 (25.5)	9.9 (25.0)
C – D	44.3 (23.0)	46.9 (22.5)

component. The random effects analysis without carry-over effects produces an average 1% reduction in the standard errors of the paired treatment comparisons compared with the fixed effects model, a modest but worthwhile gain. Comparing these estimates with the analyses in which carry-over was also fitted, two points are clear. Firsly, the standard errors of the mean treatment differences are larger when carry-over terms are present, irrespective of whether a fixed effects model is fitted or whether the patient term is regarded as random. Secondly, the reduction in the standard error of the mean treatment difference by fitting patient effects as random is larger when carry-over terms are also fitted. This latter result is general, and we will see more dramatic differences in later examples.

SAS code and output

```
/*create dummy variables for carryover effects*/
DATA new; carry=treat; SET mydata; RUN;
DATA new; SET new;
IF period eq 1 THEN carry=4;
* setting a carryover value for the first period.
  it is arbitrary which treatment is selected.
  the choice will only influence the absolute values
  of the fixed effect estimates and not the difference
  between them;

PROC MIXED; CLASS treat period patient;
TITLE 'Fixed Effect Analysis Without Carryover';
MODEL lvet=treat patient period;
LSMEANS treat/DIFF PDIFF;

PROC MIXED; CLASS treat period patient;
TITLE 'Random Effects Analysis Without Carryover';
MODEL lvet=treat period/DDFM=KENWARDROGER;
RANDOM patient;
LSMEANS treat/DIFF PDIFF;

PROC MIXED; CLASS treat period patient carry;
TITLE 'Fixed Effect Analysis Including Carryover';
MODEL lvet=treat patient period carry;
LSMEANS treat carry/DIFF PDIFF;

PROC MIXED; CLASS treat period patient carry;
TITLE 'Random Effects Analysis Including Carryover';
MODEL lvet=treat period carry/DDFM=KENWARDROGER;
RANDOM patient;
LSMEANS treat carry/DIFF PDIFF;
```

The following output is that generated by the last `PROC MIXED` procedure.

```
        Random Effects Analysis Including Carryover
                    The Mixed Procedure
                     Model Information
Data Set                            WORK.NEW
Dependent Variable                  lvet
Covariance Structure                Variance Components
Estimation Method                   REML
Residual Variance Method            Profile
Fixed Effects SE Method             Prasad-Rao-Jeske-
                                    Kackar-Harville
Degrees of Freedom Method           Kenward-Roger

                   Class Level Information
Class         Levels       Values
treat              4       1 2 3 4
period             4       1 2 3 4
patient           14       1 2 3 4 5 6 7 8 9 10 11 12 13
                           14
carry              4       1 2 3 4

                     Dimensions
                Covariance Parameters   2
                Columns in X           13
                Columns in Z           14
                Subjects                1
                Max Obs Per Subject    43

                  Number of Observations
              Number of Observations Read    43
              Number of Observations Used    43
              Number of Observations Not Used 0

                     Iteration History
Iteration    Evaluations    -2 Res Log Likelihood     Criterion
0                    1            390.90717846
1                    2            381.33170184    0.00102427
2                    1            381.14250906    0.00008011
3                    1            381.12896819    0.00000062
4                    1            381.12886806    0.00000000
                Convergence criteria met
```

Covariance Parameter Estimates

Cov Parm	Estimate
patient	2749.71
Residual	1831.47

Fit Statistics

-2 Res Log Likelihood	381.1
AIC (smaller is better)	385.1
AICC (smaller is better)	385.5
BIC (smaller is better)	386.4

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	3	22.3	5.59	0.0052
period	3	21.8	3.90	0.0226
carry	3	23.4	0.15	0.9316

Least Squares Means

Effect	treat	carry	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	1		400.10	20.8116	27.5	19.23	<.0001
treat	2		331.02	23.0808	30.4	14.34	<.0001
treat	3		367.99	20.5018	26.8	17.95	<.0001
treat	4		321.12	21.3633	29	15.03	<.0001
carry		1	356.94	23.1999	31.3	15.39	<.0001
carry		2	364.74	24.3531	31.6	14.98	<.0001
carry		3	350.07	27.5438	33	12.71	<.0001
carry		4	348.48	21.4910	27.5	16.22	<.0001

Differences of Least Squares Means

Effect	treat	carry	_treat	_carry	Estimate	Standard Error	DF	t Value
treat	1		2		69.0876	22.1976	22.5	3.11
treat	1		3		32.1188	19.9731	22.3	1.61
treat	1		4		78.9793	22.2691	21.5	3.55
treat	2		3		-36.9688	22.4442	22.6	-1.65
treat	2		4		9.8917	25.0367	22.3	0.40
treat	3		4		46.8605	22.4790	22.8	2.08

```
carry        1        2    -7.8017    25.7917    23.6    -0.30
carry        1        3     6.8688    30.4508    24.4     0.23
carry        1        4     8.4543    25.8129    21.7     0.33
carry        2        3    14.6705    28.9287    23.2     0.51
carry        2        4    16.2560    29.1453    23.8     0.56
carry        3        4     1.5855    33.3982    25.4     0.05
```

```
                Differences of Least Squares Means
    Effect    treat    carry    _treat    _carry    Pr > |t|
    treat     1                 2                   0.0050
    treat     1                 3                   0.1219
    treat     1                 4                   0.0019
    treat     2                 3                   0.1133
    treat     2                 4                   0.6965
    treat     3                 4                   0.0485
    carry               1                 2         0.7649
    carry               1                 3         0.8234
    carry               1                 4         0.7464
    carry               2                 3         0.6168
    carry               2                 4         0.5822
    carry               3                 4         0.9625
```

7.5 Incomplete block designs

The previous example demonstrated a situation where we have a design that is intended to be balanced but becomes unbalanced owing to missing observations. In contrast, we now look at incomplete block designs, where the design in itself is unbalanced. They are used in situations where, for practical reasons, the maximum possible number of treatment periods in a cross-over trial is less than the number of treatments to be evaluated, so complete balance is impossible. It is an area we will look at again in Section 8.18. The principal reason for this constraint on the number of treatment periods will usually be the duration for which any patient is in the trial. Some treatments require to be assessed over a period of several weeks, in order for there to be sufficient time for a 'steady-state' response to be reached, and so the length of individual treatment periods in the trial can be considerable. In these circumstances, it is not feasible to have multiple treatment periods because of the following reasons.

- The chance that a patient will withdraw before completing the trial protocol increases with the required time in the trial.
- In the programme of testing of a new treatment, excessively long studies will delay drug registration.

- Ethical considerations require that trial participation should not place an excessive burden on the patient.

It is readily seen that fitting a model with fixed patient effects to such a design will be inefficient because it does not allow us to use between-patient information in our treatment comparisons. To demonstrate this, consider a two-period cross-over trial with three treatments, A, B and C. Direct information on the comparison of treatments A and B is given from the within-patient differences in patients receiving both of these treatments. However, taking the random effects approach to modelling patient effects, we can see additionally that the distribution of the sum of the responses in patients receiving treatments A and C, and in those receiving B and C, also yields comparative information about treatments A and B. Thus, the random effects approach allows recovery of between-subject information. Although these designs cannot be completely balanced, they can be partially balanced by ensuring that all possible treatment sequences are used with equal frequency. As long ago as 1940, Yates described a method for recovering between-block information (the patient is the block in cross-over designs) for balanced incomplete block designs, but the benefits of applying this in cross-over studies have not been widely recognised until recently.

7.5.1 Example: Three treatment two-period cross-over trial

Mead (1988) gives results of a two-period cross-over trial to compare three analgesic drugs labelled A, B and C. The trial involved 43 patients in total, and the numbers receiving each treatment combination were as follows: AB 7; BA 5; AC 7; CA 8; BC 8; CB 8. This design is sometimes referred to as 'Koch's design'. The effectiveness of each treatment was assessed by the numbers of hours of pain relief provided. The design did not include a washout period between treatments, and so there was a strong possibility of carry-over effects. The model fitted was as follows:

$$\text{Response} = \text{overall mean} + \text{patient effect} + \text{period effect} + \text{treatment effect} + \text{carry-over effect} + \text{random error}.$$

Period, treatment and carry-over effects were taken as fixed. Analyses were carried out with patient effects specified first as fixed in a conventional least squares analysis and then as random in a mixed models analysis.

The estimated variance components and treatments effects are shown in Table 7.5, first omitting and then including carry-over effects. Comparison of the two analyses omitting carry-over effects shows that the average standard error for treatment differences from the combined analysis with recovery of between-patient information is on average 12% less than for the within-patient analysis. The comparison A − B, which has the largest standard error, shows the greatest increase in precision, so that the combined analysis also leads

Table 7.5 Estimates of variance components and treatment effects for a cross-over trial comparing three analgesic drugs. Standard errors of estimates appear in brackets.

	Fixed patients	**Random patients**
Ignoring carry-over		
Variance components		
Patients	–	1.3
Units within patients	10.8	10.7
Treatment effects		
A – B	3.4 (1.05)	3.5 (0.91)
A – C	2.0 (0.99)	1.9 (0.88)
B – C	−1.4 (0.98)	−1.6 (0.87)
Including carry-over		
Variance components		
Patients	–	1.3
Units within patients	11.3	10.9
Treatment effects		
A – B	3.7 (2.01)	3.8 (0.99)
A – C	2.6 (2.14)	1.9 (0.99)
B – C	−1.1 (2.08)	−1.9 (0.99)
Carry-over effects		
A – B	0.4 (3.43)	1.0 (1.45)
A – C	1.1 (3.66)	0.2 (1.41)
B – C	0.7 (3.72)	−0.8 (1.46)

to treatment estimates with a smaller range of standard errors. For this trial, the between-patient component of variance is relatively small, and recovery of between-patient information is clearly worthwhile.

Although the estimated carry-over effects are generally small relative to their standard errors, the results from this model could be preferred on the basis of the trial design and the absolute magnitude of the estimated carry-over effects. With this model, we see that the magnitudes of the standard errors of the treatment effects using a mixed model are only half of those obtained from the fixed effects model. The standard errors of the carry-over effects show an even greater separation. Using the mixed models approach, the penalty in fitting carry-over terms is an increase of just over 10% in the standard errors of the treatment differences.

For this trial, the between-patient variance component is small compared with the within-patient component ($\gamma = 0.12$), suggesting that there is little advantage in using a cross-over trial for testing these analgesics, even if carry-over effects can be avoided. Indeed, the predicted average standard error for treatment comparisons for a parallel group trial with the same number of patient sessions

per treatment is 0.92, compared with 0.89 and 0.99 for the cross-over analysis ignoring and including carry-over effects, respectively.

SAS code and output

This is similar to that of the earlier examples and is omitted.

7.6 Optimal designs

There has been substantial research into cross-over designs in which estimates of treatment effects and carry-over effects are both of interest. As indicated earlier in this chapter, we would question whether such an approach is likely to be desirable in clinical trials, though it may prove useful in applications in agriculture. However, for various combinations of numbers of treatments, and numbers of treatment periods, the so-called optimal designs have been derived. They satisfy the property of giving uniformly most powerful unbiased estimates of treatment and carry-over effects.

One particular optimal design has been used in practice and arguably has a stronger justification for its use than the other optimal designs. This is Balaam's design for the situation of two treatments and two treatment periods. Of course, the most common design for this situation is not Balaam's design but the simple AB/BA design. However, its critics would argue that a weakness is its inability to estimate carry-over effects and simultaneously use data from the second period in estimating treatment effects. Balaam's design resolves the problem by employing all four possible treatment sequences – AA, BB, AB and BA.

7.6.1 Example: Balaam's design

This is a well-known example initially described by Hunter *et al.* (1970). The aim was to determine the effect of amantadine (treatment A) on subjects with Parkinsonism. The trial was placebo controlled (treatment B). After a run-in period of 1 week during which baseline information was recorded, there were two 4-weekly treatment periods, without a washout period. Weekly scores (0–4) were recorded for each of 11 physical signs, and the data presented in Table 7.6 give the weekly average total scores in each treatment period. Seventeen patients were randomised, and the data have no missing values.

Table 7.7 presents the results of analyses with and without inclusion of a carry-over term and with patient effects fitted as fixed or random.

An immediate point to note from the two mixed models is the very high patient variance component compared with the residual variance component (Table 7.8). This immediately suggests that little gain in efficiency will accrue from between-patient information. This is confirmed by comparison of the

Table 7.6 Average scores for amantadine trial.

Group	Subject	Baseline	Period 1	Period 2
1AA	1	14	12.50	14.00
	2	27	24.25	22.50
	3	19	17.25	16.25
	4	30	28.25	29.75
	Mean	**22.50**	**20.56**	**20.63**
2BB	1	21	20.00	19.51
	2	11	10.50	10.00
	3	20	19.50	20.75
	4	25	22.50	23.50
	Mean	**19.25**	**18.13**	**18.44**
3AB	1	9	8.75	8.75
	2	12	10.50	9.75
	3	17	15.00	18.50
	4	21	21.00	21.50
	Mean	**14.75**	**13.81**	**14.63**
4BA	1	23	22.00	18.00
	2	15	15.00	13.00
	3	13	14.00	13.75
	4	24	22.75	21.50
	5	18	17.75	16.75
	Mean	**18.60**	**18.30**	**16.60**

Table 7.7 Estimates of variance components and treatment effects. Standard errors of estimates appear in brackets.

	Fixed patients	**Random patients**
Ignoring carry-over		
Variance components		
Patients	–	30.3
Residual (within patients)	1.05	1.1
Treatment difference	1.29 (0.49)	1.24 (0.48)
Including carry-over		
Variance components		
Patients	–	30.5
Residual (within patients)	1.12	1.1
Treatment difference	1.42 (0.73)	1.28 (0.70)
Carry-over difference	0.25 (1.06)	0.10 (0.92)

treatment standard errors, where even in the model in which carry-over is fitted, the reduction is only 4%. In most situations, where the between-patient variation is less extreme, the existence of the AA and BB treatment groups would lead us to expect greater benefits from the mixed models approach.

In this study, there is no evidence of any carry-over effect, and most statisticians would choose to report the model that excludes carry-over. However, having chosen a design for its optimal properties in estimating both treatment and carry-over effects, there is a strong case for reporting the fuller model.

The presentation in this example has been restricted to the analysis of the results in the two treatment periods, and the fact that baseline observations were also recorded has been ignored. Jones and Kenward (1989) present additional analyses utilising this baseline data, and, in particular, use interactions with the baseline to test for an effect of the baseline level on the treatment effect, period effect and carry-over effect. Although none of these terms was statistically significant at the 10% level of significance, they found indications that the treatment differences were higher with greater baseline levels. These authors also handled carry-over in a more involved way than we have employed in our analyses. They allowed for the possibility that carry-over would be different in those on the AA or BB sequence from those on the AB or BA sequence, but this term in the analysis of variance was clearly non-significant.

The conclusions from the trial will, in this instance, be qualitatively similar whichever of the previously described analytical methods is used, as long as carry-over is ignored in estimating treatment effects. Amantadine produces a reduction in the physical signs of Parkinson's disease, which is statistically significant at the 5% level. Note, however, that inclusion of carry-over terms in the model produces a substantial increase in the standard error of the treatment effect, leading to non-significance of the treatment effect.

SAS code and output

The SAS code and the structure of the output are almost identical to that presented at the end of Section 7.4, except that two treatments are used instead of four. The key results are tabulated in Table 7.7.

7.7 Covariance pattern models

In the examples considered so far in this chapter, the mixed models approach has fitted the patient effects as random. As we have seen earlier, this implies that the observations within one patient are all assumed to have the same correlation and variance. However, it could be argued that in trials with three or more periods, the correlation may vary with different pairs of periods. In particular, periods that were closer together might be expected to show higher correlations. In this section, we explore the situation where the covariance patterns used are 'structured'.

7.7.1 Structured by period

This is perhaps the most obvious way to structure the residual covariance matrix. We have already mentioned the possibility that periods close together in time might have a higher correlation than those far apart in time. In addition, it is possible that the residual variance may itself change over successive periods. For example, in early periods of the trial, while the protocol is unfamiliar to patients, the observations may be more variable than in later periods. We have already seen that SAS offers a wide choice of covariance patterns, and so a strategy for investigating alternatives is preferable to a blunderbuss approach of examining the full range available. A comparison of the compound symmetry structure (equivalent to simply fitting patient effects as random) and the general covariance matrix will usually be helpful in determining whether use of a more complicated covariance structure is likely to be useful. We will explore this further in the forthcoming example.

7.7.2 Structured by treatment

Although structuring by period is the most obvious way of introducing structure, this can also be applied to treatments. In parallel group trials, we have already met situations where we might wish to fit separate variances for each of the treatment groups. There is an exact analogy in the cross-over situation, where the variances may differ for some of the treatments. In addition, there may be good reason to suspect that the results from certain pairs of treatments may be more highly correlated than others if they have a similar mode of action. Thus, a more complicated structure for treatments than the simple compound symmetry may be highly plausible. This type of approach has been found to be particularly useful in the analysis of bioequivalence trials where treatment differences in reproducibility are highly relevant; examples of its application are available on the FDA website (www.fda.gov). We also present an example in Section 8.15.

7.7.3 Example: four-way cross-over trial

We will consider the four-way cross-over trial analysed in Section 7.4, which compared the effects of three drugs A, C and D and a placebo B on blood flow. This time, no values are set to missing, and therefore the study is more balanced (across random effects). Each treatment period lasted a week and was followed by a washout period also lasting a week. An analysis fitting a random effects model (i.e. a compound symmetry pattern) is compared with analyses fitting general covariance matrices. The covariance matrix is first structured according to periods and then according to treatments. Treatment and period effects are

Table 7.8 Comparison of covariance pattern models for the four-way cross-over trial, with inclusion of carry-over effects.

Covariance pattern	Period/ treatment	Variances ($\times 10^2$)	Correlations	−2 log likelihood (no. of parameters)	Estimated treatment differences (model-based SE)	Estimated carry-over differences (model-based SE)
Compound symmetry	1–4	36	$\begin{pmatrix} 1 & & & \\ 0.47 & 1 & & \\ 0.47 & 0.47 & 1 & \\ 0.47 & 0.47 & 0.47 & 1 \end{pmatrix}$	519.2 (2)	A – B 38.0(17.4) A – C 13.4(17.9) A – D 68.7(17.5) B – C −24.6(17.9) B – D 30.7(18.3) C – D 55.3(18.7)	−27.5(20.6) −23.4(21.2) −8.9(21.1) 4.1(21.6) 18.6(22.5) 14.5(22.5)
General (across periods)	1 2 3 4	69 19 39 12	$\begin{pmatrix} 1 & & & \\ 0.67 & 1 & & \\ 0.41 & 0.68 & 1 & \\ 0.53 & 0.48 & 0.52 & 1 \end{pmatrix}$	502.2 (10)	A – B 37.1(17.6) A – C 23.6(19.4) A – D 64.1(16.6) B – C −13.5(17.5) B – D 27.0(18.8) C – D 40.5(17.9)	−20.0(18.4) −35.1(19.6) −5.6(18.2) −15.0(18.6) 14.4(20.8) 29.5(20.2)
General (across treatments)	A B C D	17 64 34 25	$\begin{pmatrix} 1 & & & \\ 0.64 & 1 & & \\ 0.83 & 0.82 & 1 & \\ 0.10 & -0.10 & 0.16 & 1 \end{pmatrix}$	498.0 (10)	A – B 43.8(23.5) A – C 8.3(19.2) A – D 66.4(18.9) B – C −35.5(14.4) B – D 22.5(18.1) C – D 58.1(13.2)	−23.5(18.7) −54.0(22.6) −4.3(16.5) −30.5(21.5) 19.3(21.4) 49.7(24.7)

fitted as fixed effects in each analysis. In the first instance, we also include a carry-over effect as an additional fixed effect in the analysis.

Examination of the second model in Table 7.8 shows that data are more variable in the first period than in other periods. This is a common occurrence in clinical trials and is one reason why a run-in period is often used. The correlations between periods indicate a tendency for more widely separated periods to have lower correlations than those close together. The third model shows a higher variation for the placebo treatment (B) than for the three active treatments. Interestingly, the correlations involving treatment D are substantially lower than the correlation between treatments A, B and C.

Both general covariance models show significant improvements over the compound symmetry model when tested by likelihood ratio tests ($\chi^2_8 = 17.0$, $p = 0.03$ and $\chi^2_8 = 21.2$, $p = 0.007$) and might therefore be preferred.

A comparison of the three models with respect to the estimates of treatment differences shows that the estimates vary substantially when treatment C is involved. The same comment applies when the carry-over effects are compared.

Tests of the null hypothesis of equal treatment effects are all highly significant whichever model is used. For the compound symmetry model, the p-value is 0.003 compared with 0.005 for each of structuring by period and structuring by treatment.

The tests for equality of carry-over effects are more dependent on the choice of model, though none is statistically significant at conventional levels. The three models yield p-values of 0.54, 0.35 and 0.19.

All three models would therefore produce similar overall conclusions with respect to the presence of significant treatment differences without evidence of carry-over. Every model estimates the greatest levels of LVET with treatment A, followed by treatments C, B and D. The conclusions in respect of particular pairs of treatments would depend, however, on the model selected. On the basis of the likelihoods, the model employing structuring by treatment might be preferred, leading to a conclusion of similar higher levels of LVET for treatments A and C, with treatments B and D at similar lower levels. One would also conclude that results with treatment D were only weakly correlated with results on other treatments, which in turn are quite highly correlated.

In view of the absence of any significant carry-over in any of the models, it would seem reasonable to go on to consider models in which carry-over is omitted. We would advocate this even if the above model had demonstrated significant carry-over for reasons that are elaborated upon in Section 7.11. We note for the moment that this trial used reasonable washout periods, making physical carry-over of drugs unlikely, and that even if carry-over does occur, the simple carry-over model used may be inappropriate.

The use of the same three covariance structures as before, but without carry-over in the model, is summarised in Table 7.9. Many of the comments made on the previous models still apply. The patterns of covariances are similar, and structuring by period or by treatment produces a significant increase in the

Table 7.9 Comparison of covariance pattern models for the four-way cross-over trial, without carry-over effects.

Error model	Period/ treatment	Variances ($\times 10^2$)	Correlations	− 2 log likelihood (no. of parameters)	Estimated treatment differences (model SE)
Compound symmetry	1–4	36	$\begin{pmatrix} 1 & & & \\ 0.49 & 1 & & \\ 0.49 & 0.49 & & \\ 0.49 & 0.49 & 0.49 & 1 \end{pmatrix}$	544.6 (2)	A – B 43.1(16.1) A – C 23.1(16.3) A – D 70.4(16.3) B – C −20.0(16.3) B – D 27.3(16.3) C – D 47.3(16.3)
General (across periods)	1 2 3 4	74 22 37 13	$\begin{pmatrix} 1 & & & \\ 0.75 & 1 & & \\ 0.44 & 0.61 & 1 & \\ 0.43 & 0.43 & 0.58 & 1 \end{pmatrix}$	528.2 (10)	A – B 45.9(15.6) A – C 47.9(15.5) A – D 70.5(15.2) B – C 2.0(15.5) B – D 24.6(15.3) C – D 22.6(14.8)
General (across treatments)	A B C D	21 61 34 28	$\begin{pmatrix} 1 & & & \\ 0.56 & 1 & & \\ 0.82 & 0.76 & 1 & \\ 0.20 & 0.13 & 0.39 & 1 \end{pmatrix}$	527.4(10)	A – B 43.1(22.0) A – C 24.4(15.8) A – D 69.0(17.4) B – C −18.7(12.3) B – D 25.8(17.6) C – D 44.6(10.6)

likelihoods compared with the compound symmetry pattern. All of the models give similar estimates of the treatment differences, which do not involve treatment C. The estimates involving treatment C are similar for the compound symmetry model and for the structured-by-treatment model but are substantially different with structuring by period. In view of the highest likelihood being obtained when structuring by treatment and the consistency of the estimates when compared with the compound symmetry model, this model might be tentatively accepted.

As noted in the corresponding model when carry-over was included, the correlations involving treatment D are relatively low. This treatment is also the one producing the lowest measurements. This may therefore increase further the plausibility of this model, suggesting as it does a different but less effective form of action.

All of the above results have been obtained using the Kenward–Roger option to correct for the fixed effects standard error bias. In many situations, this bias has been small in comparison with the use of the Satterthwaite option, but in this example, it is occasionally large. In the models in which we structure by periods, the standard errors of the treatment differences are over 30% larger using the Kenward–Roger option. In some other models, the influence is marginal. Further details can be obtained by comparison with the corresponding table in the first edition of this book for which the Kenward–Roger option was unavailable. We believe that use of the Kenward–Roger option is the correct approach, and there is indirect support for this view by consideration of the standard errors across the different models. Their variation is relatively small, and this compares with the substantial variation seen with other approaches. In the first edition, we also reported the 'empirical' standard errors (see Section 2.4.3) and raised questions of their validity in this context. For all but one of the 18 treatment comparisons in Table 7.9, the empirical standard errors are less than the Kenward–Roger standard error, and sometimes the inconsistency is gross. For example, the empirical standard error of C − D in the structured-by-period model is only 4.6 compared with 11.2 for the model-based standard error and 14.8 for the Kenward–Roger standard error. We therefore see no place for the use of unadjusted empirical standard errors with normally distributed, repeated measures data in relatively small datasets. The role of empirical standard errors with non-normal data and the adequacy of the corrections to the empirical standard errors available in PROC GLIMMIX are still an under-researched area.

Returning to the consideration of our findings in Tables 7.8 and 7.9, we believe that there is a case for basing our inferences on the model without carry-over, with the covariances being structured by treatment. Choosing the model from which to present findings is not always straightforward, however. Decisions concerning the inclusion of carry-over terms may be based primarily on how the trial was designed and, in particular, on the adequacy of washout periods. The choice between different covariance pattern models may be influenced by consideration of the likelihoods. However, statistical tests need not be the only factor determining model choice. The validity of the assumptions relating to a model is also important.

For example, if we believe that periods and treatments are unlikely to have varying correlations based on past experience or if the trial is too small to give precise covariance estimates, then a compound symmetry structure possibly may be the one of choice.

It will always be a cause for concern when different models give qualitatively different conclusions, and this is always a greater danger when the data are relatively sparse. The different conclusions that were reached with different models and different approaches to the calculation of standard errors that we reported previously have, however, been resolved with the correction of the fixed effects standard error bias using the Kenward–Roger method.

This example makes the point that treating a multi-period, cross-over trial as repeated measures data with a covariance pattern that is structured by treatment provides an additional approach to analysis, which can be informative. For analyses that are conducted in the pharmaceutical industry for drug registration purposes, the requirement to specify the analysis plan in the trial protocol may be restrictive. This is likely to cause the compound symmetry model to be the one of choice for a primary analysis. It should be acceptable, though, to specify a secondary analysis that is structured by treatment, so that the interrelationships of responses to different treatments can be explored.

SAS code and output

The following code uses the same variable names specified at the end of Section 7.4. The code is given as follows initially for the compound symmetry model, with inclusion of carry-over effects.

```
PROC MIXED; CLASS treat period patient carry;
TITLE 'COMPOUND SYMMETRY';
MODEL lvet=treat period carry/ DDFM=KENWARDROGER;
REPEATED period/ SUBJECT=patient TYPE=CS RCORR;
LSMEANS treat carry/DIFF PDIFF;
```

The code for the other covariance pattern models is identical except for the REPEATED statement.

```
TITLE 'STRUCTURED BY PERIOD';
REPEATED period/ SUBJECT=patient TYPE=UN RCORR;

TITLE 'STRUCTURED BY TREATMENT';
REPEATED treat/ SUBJECT=patient TYPE=UN RCORR;
```

The changes to remove carry-over effects are obvious. The following output is for the example where structuring by treatment is employed.

STRUCTURED BY TREATMENT

The Mixed Procedure

Model Information

Data Set	WORK.NEW
Dependent Variable	lvet
Covariance Structure	Unstructured
Subject Effect	patient
Estimation Method	REML
Residual Variance Method	None
Fixed Effects SE Method	Prasad-Rao-Jeske-Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

Class Level Information

Class	Levels	Values
treat	4	1 2 3 4
period	4	1 2 3 4
patient	14	1 2 3 4 5 6 7 8 9 10 11 12 13 14
carry	4	1 2 3 4

Dimensions

Covariance Parameters	10
Columns in X	13
Columns in Z	0
Subjects	14
Max Obs Per Subject	4

Number of Observations

Number of Observations Read	56
Number of Observations Used	56
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Likelihood	Criterion
0	1	531.19328518	
1	2	508.00614950	0.02010791
2	1	502.46576070	0.01043261
3	1	499.65259977	0.00468615
4	1	498.42652859	0.00154375
5	1	498.03992412	0.00028985
6	1	497.97162927	0.00002064
7	1	497.96709658	0.00000023
8	1	497.96704876	0.00000000

Convergence criteria met

Estimated R Correlation Matrix for patient 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6367	0.8333	0.09861
2	0.6367	1.0000	0.8184	-0.09949
3	0.8333	0.8184	1.0000	0.1628
4	0.09861	-0.09949	0.1628	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	patient	1681.02
UN(2,1)	patient	322.45
UN(2,2)	patient	6360.25
UN(3,1)	patient	390.80
UN(3,2)	patient	3891.90
UN(3,3)	patient	3429.52
UN(4,1)	patient	-205.26
UN(4,2)	patient	2555.15
UN(4,3)	patient	2411.65
UN(4,4)	patient	2531.98

Fit Statistics

-2 Res Log Likelihood	498.0
AIC (smaller is better)	518.0
AICC (smaller is better)	524.3
BIC (smaller is better)	524.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
9	33.23	0.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	3	10.3	7.75	0.0054
period	3	11	8.91	0.0028
carry	3	14.1	1.82	0.1890

Least Squares Means

Effect	treat	carry	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	1		388.46	11.5910	13.4	33.51	<.0001
treat	2		344.64	21.7665	12.8	15.83	<.0001
treat	3		380.15	17.9227	16.2	21.21	<.0001
treat	4		322.10	13.9013	10.8	23.17	<.0001
carry		1	338.38	16.3402	21.4	20.71	<.0001
carry		2	361.92	18.4008	23.7	19.67	<.0001
carry		3	392.39	21.0386	33.6	18.65	<.0001
carry		4	342.66	15.5753	19.2	22.00	<.0001

Differences of Least Squares Means

Effect	treat	carry	_treat	_carry	Estimate	Standard Error	DF	t Value
treat	1		2		43.8265	23.5349	12.4	1.86
treat	1		3		8.3143	19.2337	12.1	0.43
treat	1		4		66.3686	18.8818	9.8	3.51
treat	2		3		-35.5123	14.4029	13.3	-2.47
treat	2		4		22.5421	18.1408	14.1	1.24
treat	3		4		58.0543	13.2261	15.3	4.39
carry		1		2	-23.5450	18.7010	17	-1.26
carry		1		3	-54.0162	22.6449	17.6	-2.39
carry		1		4	-4.2811	16.4810	7.98	-0.26
carry		2		3	-30.4712	21.5069	17	-1.42
carry		2		4	19.2639	21.4383	11.6	0.90
carry		3		4	49.7351	24.6643	30.2	2.02

Differences of Least Squares Means

Effect	treat	carry	_treat	_carry	Pr > |t|
treat	1		2		0.0864
treat	1		3		0.6731
treat	1		4		0.0058
treat	2		3		0.0280
treat	2		4		0.2343
treat	3		4		0.0005
carry		1		2	0.2250
carry		1		3	0.0286
carry		1		4	0.8016
carry		2		3	0.1747
carry		2		4	0.3871
carry		3		4	0.0527

7.8 Analysis of binary data

The earlier sections in this chapter have considered trials where the response variable was assumed to be normally distributed. Although these types of data are almost certainly the most common, binary data are by no means unusual. We have seen in Chapter 3 how the mixed models approach can be extended from normally distributed data to binary data using generalised linear mixed models. The approach described in that chapter can be used for any of the designs considered earlier in this chapter, but we will restrict ourselves to examining the simplest, and probably most common, cross-over design: the AB/BA design.

We utilise an example presented by Jones and Kenward (1989), with deletion of some observations in the second treatment period. The sample comes from safety data from a trial on cerebrovascular insufficiency. The response variable was whether an electrocardiogram was assessed by a cardiologist to be normal (1) or abnormal (0). The modified data are presented in Table 7.10.

In a fixed effects analysis, those subjects with a missing observation do not contribute to the evaluation of treatment effects. Testing the null hypothesis of no treatment difference can be performed most powerfully using Prescott's test (Prescott, 1981). This requires reorganisation of the data in Table 7.10 to the form in Table 7.11 where the rows can be thought of as representing a 'change score' from period 1 to period 2, with values −1, 0 or 1.

Values of this change score are then compared between the treatment sequence groups. This is undertaken most appropriately using an exact trend test (which is

Table 7.10 Data from an AB/BA trial on cerebrovascular deficiency. Outcomes 0 and 1 correspond to abnormal and normal ECG readings with deleted observations denoted by •.

	Outcomes						
Sequence	**(0, 0)**	**(0, 1)**	**(0, •)**	**(1, 0)**	**(1, 1)**	**(1, •)**	**Total**
AB	11	1	2	6	27	3	50
BA	12	5	2	5	23	3	50
Total	23	6	4	11	50	6	100

Table 7.11 Data from Table 7.10 reorganised in form for the application of Prescott's test.

	Change score			
Sequence	**− 1**	**0**	**1**	**Total**
AB	1	38	6	45
BA	5	35	5	45
Total	6	73	11	90

equivalent to a permutation t test). Within SAS, this can be achieved using an `EXACT` option within `PROC FREQ`. It yields a p-value of 0.33 (the corresponding asymptotic test gives $p = 0.22$).

A problem with this approach is that it is based purely on significance testing. It does not yield an estimate of the magnitude of the treatment effect. The use of a mixed model allows us to both recover the between-patient information and obtain meaningful estimates of the magnitude of the treatment effect.

As discussed in Sections 3.2.3 and 3.3.2, problems caused by uniform random effects categories are likely to arise if patients are fitted as random, since there are only two observations per patient. This can be avoided by using instead a covariance pattern model with a compound symmetry structure. The SAS code and most relevant parts of the output are presented at the end of this section.

The correlation parameter estimate is 0.59 and indicates that observations on the same patient are quite strongly correlated.

The estimate of the period effect (0.19) is similar to its standard error (0.20) and is therefore clearly non-significant. It is not negligible, however, and we repeat our recommendation that the period effect should always be included in the model.

The treatment estimate of -0.32 with 95% confidence limits from -0.71 to 0.07 corresponds to the estimate on the logistic scale. By exponentiating these figures, we obtain a point estimate and 95% confidence intervals for the odds ratio (i.e. a ratio of proportion normal to proportion abnormal in the two treatment groups). This gives an estimated odds ratio of 0.73, with 95% confidence limits of 0.49–1.07.

Note that this confidence interval was obtained using the `MODELSE` option in `PROC GENMOD` to give the model-based variance estimator. The default given is the empirical variance estimator. If empirical variance estimators are used, the 95% confidence interval is slightly narrower, but only the third decimal place is affected (these results obtained from Version 9 of `SAS` differ slightly from those in the first edition of this book, based on Version 6. The previous computational method can be used by specifying `V6CORR` as an option in the `REPEATED` statement.)

Although we should not be unduly influenced by single examples, we note that the significance level obtained for the treatment effect in this analysis was $p = 0.10$ using the empirical variance estimator and $p = 0.11$ with the model-based variance estimator compared with $p = 0.33$ for Prescott's test. The ability of the model to recover information from patients with incomplete data, together with its capacity to provide meaningful estimates of treatment effects, should make it the preferred option. This latter feature means that it may also be preferred when data are complete. When data are complete, treatment estimates can also be obtained via a bivariate logistic model proposed by Jones and Kenward (1989). They differ, however, from those provided by the mixed models approach. The odds ratio that they produce is conditional on different responses being observed in the two treatment periods. Thus, they report an odds ratio of 0.0385 on the complete data from which our example was derived. Senn (2002) has also

suggested a method based on applying ordinal logistic regression techniques to the data as configured in Table 7.11. This method yields yet another different estimate of a treatment effect in the form of an odds ratio. Our initial examination of this method suggests that it is less powerful than the alternatives considered, and we do not consider it further. The estimate from the mixed model provides, we believe, a natural, comprehensible estimate and is our recommendation for all cases where confidence interval estimates are required for the magnitude of the treatment effect. Note, however, that it is based on asymptotic theory, and the values from small studies must be regarded as approximate.

SAS code and output

```
PROC GENMOD;
CLASS treat period patient;
MODEL outcome/one=period treat/ ERROR=B;
REPEATED SUBJECT=patient/ WITHIN=period TYPE=CS MODELSE CORRW;
```

```
                    The GENMOD Procedure
                      Model Information
              Data Set                          WORK.A
              Distribution                    Binomial
              Link Function                      Logit
              Response Variable (Events)      outcome
              Response Variable (Trials)          one
              Number of Observations Read         200

              Number of Observations Used         190
              Number of Events                    125
              Number of Trials                    190
              Missing Values                       10

                    Class Level Information
Class     Levels  Values
treat        2    A B
period       2    1 2
patient    100    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
                  21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
                  38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
                  55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
                  72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87
                  ...
```

Parameter Information

Parameter	Effect	treat	period
Prm1	Intercept		
Prm2	period		1
Prm3	period		2
Prm4	treat	A	
Prm5	treat	B	

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	187	242.0829	1.2946
Scaled Deviance	187	242.0829	1.2946
Pearson Chi-Square	187	189.9881	1.0160
Scaled Pearson X2	187	189.9881	1.0160
Log Likelihood		-121.0414	

Algorithm converged.

Analysis Of Initial Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	0.8128	0.2752	0.2734	1.3522	8.72	0.0031
period	1	1	0.1146	0.3077	-0.4884	0.7177	0.14	0.7095
period	2	0	0.0000	0.0000	0.0000	0.0000	.	.
treat	A	1	-0.4233	0.3082	-1.0273	0.1808	1.89	0.1696
treat	B	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

As in the example considered in Section 6.4, the above output will always be produced, but it is irrelevant.

GEE Model Information

Correlation Structure	Exchangeable
Within-Subject Effect	period (2 levels)
Subject Effect	patient (100 levels)
Number of Clusters	100
Clusters With Missing Values	10
Correlation Matrix Dimension	2
Maximum Cluster Size	2
Minimum Cluster Size	1

Algorithm converged.

Working Correlation Matrix

	Col1	Col2
Row1	1.0000	0.5851
Row2	0.5851	1.0000

Exchangeable Working Correlation

Correlation 0.5850903245

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		0.6842	0.2484	0.1973	1.1712	2.75	0.0059
period	1	0.1901	0.1976	-0.1972	0.5774	0.96	0.3360
period	2	0.0000	0.0000	0.0000	0.0000	.	.
treat	A	-0.3213	0.1979	-0.7091	0.0665	-1.62	0.1044
treat	B	0.0000	0.0000	0.0000	0.0000	.	.

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		0.6842	0.2418	0.2103	1.1581	2.83	0.0047
period	1	0.1901	0.2001	-0.2020	0.5822	0.95	0.3419
period	2	0.0000	0.0000	0.0000	0.0000	.	.
treat	A	-0.3213	0.2004	-0.7140	0.0715	-1.60	0.1089
treat	B	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1.0013	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

7.9 Analysis of categorical data

Apart from the case of binary data, response variables that are purely categorical, without an underlying scale, are extremely rare. We will therefore only consider data on ordinal scales in this section. Variables classified as none, mild, moderate and severe will arise in a variety of contexts.

To illustrate techniques, we will again take an example from Jones and Kenward (1989) and delete five observations from the second treatment period. The example is a placebo-controlled trial of a treatment for primary dysmenorrhoea.

Table 7.12 Data from an AB/BA trial on a treatment for primary dysmenorrhoea (A: placebo; B: high-dose analgesic; deleted observations in the second period denoted by •).

Sequence	(1, 1)	(1, 2)	(1, 3)	(1, •)	(2, 1)	(2, 2)	(2, 3)	(2, •)	(3, 1)	(3, •)	Total
AB	2	3	5	1	1	1	2	1	0	0	16
BA	3	2	0	0	1	0	1	1	4	2	14
Total	5	5	5	1	2	1	3	2	4	2	30

	Change score					
Sequence	**− 2**	**− 1**	**0**	**1**	**2**	**Total**
AB	5	5	3	1	0	14
BA	0	3	3	1	4	11
Total	5	8	6	2	4	25

Thirty patients entered the trial, and in each treatment period, the amount of relief obtained was recorded as none or minimal (1), moderate (2) and complete (3). The data as we analyse them are summarised in Table 7.12.

Taking a fixed effects approach, a test of significance is most readily obtained using methods based on the analysis of an appropriate contingency table. A simple but inefficient way of producing such a contingency table would be to categorise the changes in the outcome variable from the first treatment period to the second as 'worse', 'no change' and 'better' and to tabulate this variable against the treatment sequence. The significance of the treatment effect could then be determined from this 3×2 contingency table, as in Prescott's test. However, this configuration does not use the information that observations of 'none' and 'complete' in the two treatment periods represent a larger difference than between 'none' and 'moderate' or 'moderate' and 'complete'. If we arbitrarily assign numbers of 1, 2 and 3 to the outcome categories, we can generate a 5×2 contingency table based on the change scores. The 'obvious' approach is to then apply a permutation t test (test for trend) to this table to assess the significance of the treatment effect. Application to the change scores presented in Table 7.12 gives $p = 0.005$.

In applying this test, it should be appreciated that the scores of -2, -1, 0, $+1$ and $+2$ are arbitrary. They should not be taken to imply that the difference between 'none' and 'complete' is twice as large as the difference between 'none' and 'moderate' nor that the difference between 'none' and 'moderate' is the same as that between 'moderate' and 'complete'. For this reason, some statisticians may wish to replace the change scores with ranks and apply an (exact) Wilcoxon rank sum test. The choice will rarely make any practical difference, but it is clearly good practice to make this choice prior to analysis, rather than reporting the more favourable result. Note that the situation becomes more complicated when there are more than three categories for the outcome variable. Analysis could still

be based on the change scores, but there would be an implicit strong assumption about the meaning of the intervals between the categories. Without such strong assumptions, many of the categories of change would be indistinguishable from each other, and a simplified 5×2 contingency table would result. Such an example is presented by Senn (2002).

The mixed models approach with random patient effects and fixed period and treatment effects, based on carrying out ordinal logistic regression, is now available through PROC GLIMMIX. The patient variance component is, surprisingly for a cross-over trial, estimated to be negative and therefore set to zero. The coefficient for the treatment effect, on the logistic scale, is 1.92, with a standard error of 0.57 ($p = 0.003$). On exponentiation, the estimate of the odds ratio is 6.8, with 95% confidence limits of 2.1 and 22.1. The interpretation of the odds ratio in this situation is that the estimated odds of being in a favourable outcome category when treated with the analgesic compared with placebo is 6.8, whether favourable is defined as complete relief or moderate/complete relief.

SAS code and output

```
PROC GLIMMIX; CLASS patient period treat;
MODEL outc=period treat/DIST=MULT LINK=CLOGIT SOLUTION OR;
RANDOM patient;
```

Note that, in this example, if the option DDFM = KR is used in the model statement, the denominator degrees of freedom erroneously appear as 1.

```
   Number of Observations Read    60
   Number of Observations Used    55

          Response Profile
Ordered                      Total
Value          outc        Frequency
1                0             26
2                1             15
3                2             14

The GLIMMIX procedure is modeling the probabilities
of levels of outc having lower Ordered Values in the
Response Profile table.

          Dimensions
          G-side Cov. Parameters      1
          Columns in X                6
          Columns in Z               30
          Subjects (Blocks in V)      1
          Max Obs per Subject        55
```

Optimization Information

Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	1	371.01856686	2.00000000	2.187345
1	0	0	386.34874615	0.07614525	1.541061
2	0	0	386.2767642	0.00860819	1.571373
3	0	0	386.30470584	0.00062015	1.571598
4	0	0	386.30608754	0.00005055	1.571655
5	0	0	386.30626243	0.00000363	1.571655
6	0	0	386.30627135	0.00000029	1.571655
7	0	0	386.30627234	0.00000002	1.571655
8	0	0	386.3062724	0.00000000	1.571655

Convergence criterion (PCONV=1.11022E-8) satisfied.
Estimated G matrix is not positive definite.

NOTE: The covariance matrix is the null matrix.

Fit Statistics

-2 Res Log Pseudo-Likelihood	386.31

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
patient	0	.

Solutions for Fixed Effects

Effect	outc	period	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	0			-1.3774	0.5041	29	-2.73	0.0106
Intercept	1			0.09043	0.4577	29	0.20	0.8448
period		1		0.4195	0.5359	22	0.78	0.4421
period		2		0	.	.	.	.
treat			0	1.9162	0.5679	22	3.37	0.0027
treat			1	0	.	.	.	.

Odds Ratio Estimates

Effect	period	treat	_period	_treat	Estimate	DF	95% Confidence Limits	
period	1		2		1.52	22	0.501	4.62
treat		0		1	6.80	22	2.093	22.06

```
        Type III Tests of Fixed Effects
           Num   Den
Effect      DF    DF    F Value      Pr > F
period       1    22       0.61      0.4421
treat        1    22      11.39      0.0027
```

7.10 Use of results from random effects models in trial design

Once a cross-over trial has been analysed and reported, there is a subsequent question to consider. What has the present study taught us about the trial design to be used in future studies of this condition? The factors we will wish to take into consideration are the sizes of the residual and patient variance components and the dropout rates at various stages during cross-over. If between-subject variability is large compared with residual variation, a cross-over design may be vastly more efficient than a corresponding parallel group study. However, the analysis of a cross-over trial requires more assumptions than a parallel group design, and if between-subject variability is relatively small, a parallel group study may be the design of choice. If a multi-period cross-over trial experiences a substantial dropout in the later phases of the cross-over or if the required duration of each treatment is long, then an incomplete block design may be considered.

Of course, the most basic way in which we can use data from one cross-over trial to plan a succeeding trial is to use the estimate of the residual variance in standard sample size formulae (see, for example, Senn, 2002). If there has been a sequence of similar trials, then a weighted average of the estimates of the residual variance would provide a more robust figure.

The more interesting question is whether a cross-over design should be used at all. This can usually be assessed satisfactorily by comparing the standard error of the treatment differences in the cross-over trial with the standard errors that would be expected from a comparable parallel group study. We reported such information for the trial described in Section 7.5, and we now consider another example.

7.10.1 Example

We consider the results from the oral mouthwash trial, summarised in Table 7.3. The estimate of the patient variance component is 0.029 and that of the residual variance is 0.066. Thus, the estimated residual variance in a parallel group trial is 0.095, the sum of these two variance components. The expected standard error of the treatment difference in a parallel group trial with 34 patients per group

(giving the comparable number of treatment periods to the cross-over trial) is

$$\sqrt{\frac{0.095}{34} + \frac{0.095}{34}} = 0.075.$$

This compares with the standard error of 0.065 in the data analysed. To achieve a similar standard error would require 45 patients in each arm of a parallel group study. There is therefore a trade-off to be made between the advantages of each design. The advantages of the parallel group design are as follow:

- 6 weeks in the trial per patient compared with 15 weeks;
- simpler administration;
- simpler analysis; and
- absence of assumptions concerning carry-over.

The advantage of the cross-over design is its greater efficiency:

- 34 patients (68 treatment periods) versus 90 patients.

A deciding factor between the alternatives in examples such as this will often be the availability of an adequate number of patients for the parallel group study.

7.11 General points

This chapter has demonstrated that random effects models can have advantages over fixed effects models in the context of cross-over trials. In balanced situations, with normally distributed data, the results of both analyses will generally be identical. In unbalanced situations, however, the random effects models will lead to smaller standard errors of the estimates of treatment differences. If the degree of imbalance is slight (e.g. few missing observations in a balanced design) and/or if the patient variance component is large compared with the residual variance component, this reduction in the size of the standard error will be modest. It should be remembered, though, that some clinical trials are very expensive to conduct, and even a small gain in statistical efficiency may be equivalent to the recruitment of one or two additional costly patients.

It is tempting, therefore, to recommend routine use of the random effects model. There are situations, however, where the methods may not be sufficiently robust. This is of particular concern when we are dealing with non-normal data. The methods are based on asymptotic theory, and we are not aware of sufficient research to quantify the biases that may occur with small samples. The capability of the random effects model to summarise treatment effects on binary data in the form of odds ratios is an attractive feature, but at present, it is perhaps prudent to exercise caution in its use if sample sizes are fairly small.

In the case of normally distributed data, there is somewhat greater experience, and the methods may be used with more confidence. In the past, a major source

of concern was the bias in the standard errors of the fixed effects. This arises from imprecision in the estimates of the variance components, which is inevitably greater in relatively small studies. For example, in a cross-over design in Phase I or Phase II trials where the number of subjects is low, the variance components may be estimated imprecisely. The 'standard' estimates of the fixed effects standard errors are, however, based on the assumption that the variance components are known. Whenever a fixed effect is estimated from two (or more) error strata, it is known that the standard errors are biased downwards to some extent. This will occur in cross-over designs that are unbalanced by design (Section 7.5) or because of missing values (Sections 7.3 and 7.4) and in which mixed models are fitted. The implementation of Kenward and Roger's method for correcting the fixed effects standard error bias by inflation of $\mathtt{var}(\hat{\boldsymbol{\alpha}})$, together with the use of Satterthwaite degrees of freedom, has meant that mixed models can now be applied to normally distributed data with much greater confidence, even when the sample sizes are fairly small. We recommend that within `SAS` the model option `DDFM = KENWARDROGER` (or equivalently `DDFM = KR`) should be used routinely.

Several of the examples presented in this chapter have demonstrated that the benefits of using a random effects model are much more pronounced in models where carry-over is being estimated. We have already alluded briefly to the fact that the use of such models may be inadvisable. Senn (2002) has expressed eloquently the arguments against the inclusion of simple carry-over terms, and we find them compelling.

Senn summarises the case against adjusting for carry-over as follows:

- The simple carry-over model has been developed without reference to pharmacological or biological models.
- It does not provide a useful approximation to reality.
- It leads to more complicated estimation procedures that are more difficult to describe and understand.
- Usually, the adjusted estimators have higher variance than the unadjusted ones.
- Although the adjusted estimators will be unbiased if simple carry-over applies, in practice, if carry-over occurs, then they will be biased, and it is perfectly possible that this bias will be larger than it is for unadjusted estimators.
- The most serious objection, however, is that the use of such approaches encourages the erroneous belief that the validity of estimates obtained from cross-over trials does not depend on adequate washout having taken place.

We end this section by highlighting the potential for the use of covariance pattern models in the analysis of cross-over trials. The nature of the cross-over with repeated observations on the same subject leads naturally to the consideration of a 'standard' repeated measures approach, with the covariance pattern structured by the visits. The use of this approach to structure by treatment is perhaps not immediately obvious, but in multi-treatment trials, this may

well have much to offer, with subsets of similar treatments producing greater correlations than those with very different modes of action. Experimentation with such plausible covariance structures can provide a greater insight into the data. Importantly, the treatment effect estimates and their standard errors will also be more appropriate if the best covariance structure is modelled. This type of modelling is becoming more important in the analysis of multi-period cross-over trials to assess bioequivalence, and this is considered further in Section 8.15.

8

Other applications of mixed models

In this chapter, the use of mixed models in a variety of situations is considered. In Chapters 5, 6 and 7, we covered three different types of data structure: hierarchical, repeated measures and crossed. Designs with a combination of these features can also arise, and some of these are considered in Sections 8.1–8.4. In Section 8.5, the matched case–control study data are considered, and in Section 8.6, a covariance pattern model is used to allow treatment groups to have different variances in a simple between-patient study. The examples in Sections 8.7–8.14 have arisen from consultancy work and have a variety of structures. In Section 8.15, we look at bioequivalence studies with replicate cross-over designs. The use of mixed models in the analysis of cluster randomized trials is considered in Section 8.16. Section 8.17 looks at the analysis of bilateral data. The chapter finishes in Section 8.18 by looking at the design of incomplete block studies.

8.1 Trials with repeated measurements within visits

Sometimes, repeated measurements occur within visits in cross-over or repeated measures trials. For example, bioequivalence trials often record several blood or urine measurements at each visit within a cross-over design. Studies in cardiology sometimes involve exercise tests where repeated measurements are made throughout the test at each visit. Cross-over trials in asthma may involve a series of lung function measurements made after a 'challenge' designed to provoke an asthma attack.

When the data are complete at each visit, a simple approach would be to calculate summary statistics for each visit (e.g. area under the curve, maximum value

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

or time to maximum value) and to analyse these derived variables using methods suggested for ordinary repeated measures or cross-over data (see Chapters 6 and 7). This approach has the advantage of simplicity and gives a straightforward interpretation. It cannot, however, test the treatment·reps interaction (reps are repeated measurements within visits) or always overcome problems caused by missing data.

When there are missing data, the use of summary statistics may not be satisfactory, and a mixed model is often more appropriate. If visits and reps occur at fixed time intervals, a covariance pattern model can be used to structure the covariances by visits and reps. Alternatively, if the visits and/or reps occur at irregular intervals or if it is of interest to model the relationship of the response with time, then a random coefficients model can be used instead.

8.1.1 Covariance pattern models

There are several ways in which the covariance of the data can be modelled when measurements are taken across both visits and reps. In this section, we will present five of the more plausible options. In models for ordinary repeated measures trials (considered in Chapter 6), the overall variance matrix, $\mathbf{V}$, had a block diagonal form, with zero correlations between observations on different patients:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_5} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_6} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_7} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_9} \end{pmatrix},$$

where the $\mathbf{V_i}$ are blocks of covariances for observations on the ith patient. We will again use this form for $\mathbf{V}$, but now there are more ways in which the $\mathbf{V_i}$ can be structured. We will illustrate a variety of possible structures, assuming a dataset with three visits and three reps per visit (nine observations per patient), which leads to each $\mathbf{V_i}$ being a 9×9 submatrix.

Constant covariances A very simple structure for $\mathbf{V_i}$ would assume a constant correlation between all observations on the same patient regardless of the visit or

rep number such that

$$\mathbf{V}_i = \begin{array}{c} \begin{matrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \end{matrix} \\ \begin{pmatrix} \sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta & \sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta & \theta & \sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta & \theta & \theta & \sigma^2 & \theta & \theta & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta & \sigma^2 & \theta & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta & \theta & \sigma^2 & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta & \theta & \theta & \sigma^2 & \theta & \theta \\ \theta & \theta & \theta & \theta & \theta & \theta & \theta & \sigma^2 & \theta \\ \theta & \theta & \theta & \theta & \theta & \theta & \theta & \theta & \sigma^2 \end{pmatrix} \end{array} \begin{array}{cc} & \textbf{Visit} \\ \textbf{Rep} & \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 1 & 2 \\ 2 & 2 \\ 3 & 2 \\ 1 & 3 \\ 2 & 3 \\ 3 & 3 \end{array} \qquad \text{(A)}$$

where

θ = covariance between observations on same patient,
σ^2 = residual variance.

Extra covariance for observations at the same visit The above pattern is perhaps oversimplistic, as it takes no account of the possibility that observations taken at the same visit are more highly correlated than those taken at different visits. A simple way to account for this would be to parameterise $\mathbf{V}_i$ with a different covariance for observations on the same visit,

$$\mathbf{V}_i = \begin{pmatrix} \sigma^2 & \theta_v & \theta_v & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta_v & \sigma^2 & \theta_v & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta_v & \theta_v & \sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta & \theta & \theta & \sigma^2 & \theta_v & \theta_v & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta_v & \sigma^2 & \theta_v & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta_v & \theta_v & \sigma^2 & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta & \theta & \theta & \sigma^2 & \theta_v & \theta_v \\ \theta & \theta & \theta & \theta & \theta & \theta & \theta_v & \sigma^2 & \theta_v \\ \theta & \theta & \theta & \theta & \theta & \theta & \theta_v & \theta_v & \sigma^2 \end{pmatrix}, \qquad \text{(B)}$$

where

θ = covariance between observations on different visits,
θ_v = covariance between observations at the same visit,
σ^2 = residual variance.

A covariance pattern structured by visits Alternatively, it is possible that the correlation between observations is different for each pair of visits leading to

$$\mathbf{V_i} = \begin{pmatrix} \sigma_1^2 & \theta_{11} & \theta_{11} & \theta_{12} & \theta_{12} & \theta_{12} & \theta_{13} & \theta_{13} & \theta_{13} \\ \theta_{11} & \sigma_1^2 & \theta_{11} & \theta_{12} & \theta_{12} & \theta_{12} & \theta_{13} & \theta_{13} & \theta_{13} \\ \theta_{11} & \theta_{11} & \sigma_1^2 & \theta_{12} & \theta_{12} & \theta_{12} & \theta_{13} & \theta_{13} & \theta_{13} \\ \theta_{12} & \theta_{12} & \theta_{12} & \theta_{22} & \theta_{22} & \sigma_2^2 & \theta_{23} & \theta_{23} & \theta_{23} \\ \theta_{12} & \theta_{12} & \theta_{12} & \theta_{22} & \sigma_2^2 & \theta_{22} & \theta_{23} & \theta_{23} & \theta_{23} \\ \theta_{12} & \theta_{12} & \theta_{12} & \theta_{23} & \theta_{22} & \sigma_2^2 & \theta_{23} & \theta_{23} & \theta_{23} \\ \theta_{13} & \theta_{13} & \theta_{13} & \theta_{23} & \theta_{23} & \theta_{23} & \sigma_3^2 & \theta_{33} & \theta_{33} \\ \theta_{13} & \theta_{13} & \theta_{13} & \theta_{23} & \theta_{23} & \theta_{23} & \theta_{33} & \sigma_3^2 & \theta_{33} \\ \theta_{13} & \theta_{13} & \theta_{13} & \theta_{23} & \theta_{23} & \theta_{23} & \theta_{33} & \theta_{33} & \sigma_3^2 \end{pmatrix}, \tag{C1}$$

where

θ_{ij} = covariance between observations at visits i and j,

σ_i^2 = residual variance at visit i (this may be parameterised as $\theta_{ii} + \sigma^2$).

This matrix, in fact, has a general covariance pattern structured by visits.

Another alternative would be to use a different covariance pattern for the correlations between visits. For example, by using a Toeplitz pattern, the correlation between observations will depend on the separation of the visits and has the form

$$\mathbf{V_i} = \begin{pmatrix} \sigma^2 & \theta_1 & \theta_1 & \theta_2 & \theta_2 & \theta_2 & \theta_3 & \theta_3 & \theta_3 \\ \theta_1 & \sigma^2 & \theta_1 & \theta_2 & \theta_2 & \theta_2 & \theta_3 & \theta_3 & \theta_3 \\ \theta_1 & \theta_1 & \sigma^2 & \theta_2 & \theta_2 & \theta_2 & \theta_3 & \theta_3 & \theta_3 \\ \theta_2 & \theta_2 & \theta_2 & \sigma^2 & \theta_1 & \theta_1 & \theta_2 & \theta_2 & \theta_2 \\ \theta_2 & \theta_2 & \theta_2 & \theta_1 & \sigma^2 & \theta_1 & \theta_2 & \theta_2 & \theta_2 \\ \theta_2 & \theta_2 & \theta_2 & \theta_1 & \theta_1 & \sigma^2 & \theta_2 & \theta_2 & \theta_2 \\ \theta_3 & \theta_3 & \theta_3 & \theta_2 & \theta_2 & \theta_2 & \sigma^2 & \theta_1 & \theta_1 \\ \theta_3 & \theta_3 & \theta_3 & \theta_2 & \theta_2 & \theta_2 & \theta_1 & \sigma^2 & \theta_1 \\ \theta_3 & \theta_3 & \theta_3 & \theta_2 & \theta_2 & \theta_2 & \theta_1 & \theta_1 & \sigma^2 \end{pmatrix}, \tag{C2}$$

where

θ_i = covariance between observations separated by $i-1$ visits,

σ^2 = residual variance at visit.

A covariance pattern structured by reps It is also possible that correlation between observations at the same visit differs depending on the rep number. A structure assuming constant correlation between observations on different

visits (as in (B)) but a different correlation for each pair of reps at the same visit is

$$\mathbf{V_i} = \begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta_{12} & \sigma_2^2 & \theta_{23} & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta_{13} & \theta_{23} & \sigma_3^2 & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta & \theta & \theta & \sigma_1^2 & \theta_{12} & \theta_{13} & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta_{12} & \sigma_2^2 & \theta_{23} & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta_{13} & \theta_{23} & \sigma_3^2 & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta & \theta & \theta & \sigma_1^2 & \theta_{12} & \theta_{13} \\ \theta & \theta & \theta & \theta & \theta & \theta & \theta_{12} & \sigma_2^2 & \theta_{23} \\ \theta & \theta & \theta & \theta & \theta & \theta & \theta_{13} & \theta_{23} & \sigma_3^2 \end{pmatrix}, \qquad \text{(D1)}$$

where

θ = covariance between observations on different visits,
θ_{ij} = covariance between observations on reps i and j (at the same visit),
σ_i^2 = residual variance at rep i (this may be parameterised as $\theta_{ii} + \sigma^2$).

This matrix, in fact, has a general covariance pattern structured by reps (within visits).

Alternatively, a different pattern from the general pattern could be considered. For example, a first-order autoregressive pattern allowing the correlation between observations to decrease exponentially depending on the separation of the reps would have the form

$$\mathbf{V_i} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \theta & \theta & \theta & \theta & \theta & \theta \\ \theta & \theta & \theta & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \theta & \theta & \theta \\ \theta & \theta & \theta & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \theta & \theta & \theta \\ \theta & \theta & \theta & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \theta & \theta & \theta \\ \theta & \theta & \theta & \theta & \theta & \theta & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \theta & \theta & \theta & \theta & \theta & \theta & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \theta & \theta & \theta & \theta & \theta & \theta & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}, \qquad \text{(D2)}$$

where

θ = covariance between observations on different visits,
$\rho^{|i-j|}$ = correlation between observations on reps i and j,
σ^2 = residual variance.

In both of these models, observations at the same visit could be estimated to be less correlated than those at different visits. This will usually be implausible and can be avoided if we add θ to the covariance of observations at the same visit but

with different rep. As we will see in Section 8.1.2, it is this more plausible version of the model that can be implemented in SAS.

Extra covariance for observations on the same reps It is also possible that there is additional correlation between observations on the same reps. Adding this feature to structure (B) previously shown we obtain the structure

$$\mathbf{V}_i = \begin{pmatrix} \sigma^2 & \theta_v & \theta_v & \theta_r & \theta & \theta & \theta_r & \theta & \theta \\ \theta_v & \sigma^2 & \theta_v & \theta & \theta_r & \theta & \theta & \theta_r & \theta \\ \theta_v & \theta_v & \sigma^2 & \theta & \theta & \theta_r & \theta & \theta & \theta_r \\ \theta_r & \theta & \theta & \sigma^2 & \theta_v & \theta_v & \theta_r & \theta & \theta \\ \theta & \theta_r & \theta & \theta_v & \sigma^2 & \theta_v & \theta & \theta_r & \theta \\ \theta & \theta & \theta_r & \theta_v & \theta_v & \sigma^2 & \theta & \theta & \theta_r \\ \theta_r & \theta & \theta & \theta_r & \theta & \theta & \sigma^2 & \theta_v & \theta_v \\ \theta & \theta_r & \theta & \theta & \theta_r & \theta & \theta_v & \sigma^2 & \theta_v \\ \theta & \theta & \theta_r & \theta & \theta & \theta_r & \theta_v & \theta_v & \sigma^2 \end{pmatrix}, \tag{E1}$$

where

θ = covariance between observations at different visits and reps,
θ_v = covariance between observations at the same visit but different rep,
θ_r = covariance between observations at the same rep but different visits,
σ^2 = residual variance.

We can also make this structure slightly more complex by assuming that each visit has a different variance and each pair of visits has a different covariance:

$$\mathbf{V}_i = \begin{pmatrix} \sigma^2_{11} & \theta_v & \theta_v & \theta_{12} & \theta & \theta & \theta_{13} & \theta & \theta \\ \theta_v & \sigma^2_{11} & \theta_v & \theta & \theta_{12} & \theta & \theta & \theta_{13} & \theta \\ \theta_v & \theta_v & \sigma^2_{11} & \theta & \theta & \theta_{12} & \theta & \theta & \theta_{13} \\ \theta_{12} & \theta & \theta & \sigma^2_{22} & \theta_v & \theta_v & \theta_{23} & \theta & \theta \\ \theta & \theta_{12} & \theta & \theta_v & \sigma^2_{22} & \theta_v & \theta & \theta_{23} & \theta \\ \theta & \theta & \theta_{12} & \theta_v & \theta_v & \sigma^2_{22} & \theta & \theta & \theta_{23} \\ \theta_{13} & \theta & \theta & \theta_{23} & \theta & \theta & \sigma^2_{33} & \theta_v & \theta_v \\ \theta & \theta_{13} & \theta & \theta & \theta_{23} & \theta & \theta_v & \sigma^2_{33} & \theta_v \\ \theta & \theta & \theta_{13} & \theta & \theta & \theta_{23} & \theta_v & \theta_v & \sigma^2_{33} \end{pmatrix}, \tag{E2}$$

where

θ = covariance between observations at different visits and reps,
θ_v = covariance between observations at the same visit but at different reps,
θ_{ij} = covariance between observations at the same rep at visits i and j,
σ^2_{ii} = residual variance at visit i.

Choosing the covariance pattern

It is clear from the options given previously that there is a huge amount of flexibility in choosing how to define the covariance structure of the data. If the covariance structure in itself is of particular interest, it may be worthwhile to experiment with several different structures, building up a model by starting with a very simple structure such as (A), and then using likelihood ratio tests (see Section 6.2.2) to determine whether more complex structures lead to significant improvements. However, if interest lies only with, say, comparing two treatment groups or their interaction with visit or rep, estimates of means and standard errors from a fairly simple structure such as (B), which allows for correlation between observations on the same visit, are likely to differ little from those obtained using a more complex structure.

We note that the range of covariance structures available may be limited by the software, and care is sometimes needed to ensure that covariance parameters are not confounded.

8.1.2 Example

The data were from a three-period cross-over trial taken from Jones and Kenward (1989) to compare the effects of three treatments on systolic blood pressure. Treatments A and B are 20 and 40 mg doses of an active drug, and treatment C is a placebo. There were 12 patients, and 10 measurements were made at each visit. These were taken at 30 and 15 min before treatment and at 15, 30, 45, 60, 75, 90, 120 and 240 min after treatment. We assume that the objective is to assess whether there are any post-treatment differences between treatments and to test whether these differences are constant over the reps. Since the data are complete, calculating summary statistics (e.g. AUC, minimum or maximum value, time to maximum value) and analysing them as cross-over data would form a simple strategy, although the treatment·rep interaction could not then be tested. However, in this section, for illustration, we will analyse the raw data using covariance pattern models. Measurements at 15 min pre-treatment will be taken as baseline values.

Although the main interest in this example lies with the comparison of treatments, for illustration, we will use likelihood ratio tests to determine an appropriate covariance structure for the data. The models that we examined and the resulting values of $-2\log(L)$ are shown in Table 8.1. The covariance structures relate to the example structures defined previously. Initially, Model 1 was used to test treatment·visit, treatment·rep and carry-over effects to determine whether they could be omitted from the models. None of these effects was significant; however, treatment·rep effects were retained in the models so that mean treatment·rep profiles could be produced. The fixed effects included in the models were therefore baseline (15 min pre-treatment), treatment, visit, rep and treatment·rep effects.

Table 8.1 Values of −2 log(L) for all models.

Model	Covariance structure	$-2\log(L)$	Parameters
1	B	1879.79	3
2	E1	1876.84	4
3	C1	1878.09	7
4	D2	1866.98	3
5	D2, by treatment	1861.68	9

Models 1 and 2 are compared to establish whether there is extra correlation between observations on the same rep. The likelihood ratio statistic is $\chi^2_1 = 1879.79 - 1876.84 = 2.95 (p = 0.09)$. Therefore, we do not have strong statistical evidence that there is any additional correlation across the reps, and structures taking account of this are not considered further.

Model 3, with a general structure across visits, is compared with Model 1 to determine whether different correlations are justified for each pair of visits. On comparison with Model 1, we obtain $\chi^2_4 = 1.70$ ($p = 0.79$), which indicates no significant improvement over Model 1 that uses the same correlation between all visits.

Model 4 tests whether the correlation between observations within the same visit decreases as reps become more widely separated. The structure fitted is similar to (D2) but with the modification to ensure that covariances at the same visit are not less than those across different visits. The same number of parameters is used as in Model 2, and therefore the log likelihoods can be compared directly. Model 4 has a higher likelihood, and we therefore conclude that the first-order autoregressive structure is the more appropriate and that correlation does decrease with the separation of the reps.

Model 5 is similar to Model 4 but fits a separate covariance matrix for each treatment. The likelihood ratio statistic on comparison with Model 4 is $\chi^2_6 = 5.30$ ($p = 0.51$). This is non-significant, and so there is no evidence that the covariances differ between treatments, and we can choose to base our conclusions on results from Model 4.

Results from Model 4

The exact form of structure (D2) could not be fitted using `PROC MIXED` because the `REPEATED` statement can only be used to fit a single covariance pattern. It was necessary to use a `RANDOM` statement to give the constant covariance between observations on the same patient. The covariance parameter estimates obtained were

$$\text{Covariance between observations on same patient} = 46.100$$

$$\text{Autoregressive correlation coefficient} = 0.323$$

$$\text{Residual} = 55.600.$$

Thus, the variance matrix block for observations on the same patient, $\mathbf{V}_i$, has the form

$$\mathbf{V}_i = \begin{pmatrix} 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \\ 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 & 46 \end{pmatrix}$$

$$+ 56 \begin{pmatrix} 1 & 0.32 & 0.32^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.32 & 1 & 0.32 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.32^2 & 0.32 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.32 & 0.32^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.32 & 1 & 0.32 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.32^2 & 0.32 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.32 & 0.32^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.32 & 1 & 0.32 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.32^2 & 0.32 & 1 \end{pmatrix}.$$

Note that the diagonal blocks are in fact 8×8, with autoregressive correlations ranging between 0.32 and 0.32^8. However, owing to space limitations, their general form is illustrated in this section using 3×3 blocks.

The results for the treatment effects were as follows:

Effect	**Difference (SE)**	***t* test DF**	***p* - value**
A – B	−3.18 (1.46)	261	0.03
A – C	3.05 (1.61)	234	0.06
B – C	6.24 (1.52)	261	0.0001

Thus, the 40 mg treatment (B) produces significantly higher systolic blood pressure than the 20 mg treatment (A) and the placebo (C). The treatment·rep interaction was non-significant ($p = 0.35$) and, therefore, we can be reasonably confident in reporting treatment effects over all reps. The standard errors are 'model based' and are thus calculated from the covariance pattern parameters estimated by the model. However, if we had assumed a simple pattern (e.g. Model A or B) without testing more complex patterns, an alternative would be to use the 'empirical' estimates, with a correction for the standard error bias, which would have taken more account of the observed covariance of the data.

SAS code and output

Variables
`pat` = patient,
`visit` = visit,
`treat` = treatment (A = 20 mg dose, B = 40 mg dose, C = placebo),
`rep` = rep,
`sbp` = post-treatment systolic blood pressure (mm Hg),
`pre` = pre-treatment systolic blood pressure (mm Hg).

The following statements were used in all the models to specify the fixed effects:

```
PROC MIXED; CLASS pat visit treat rep;
MODEL sbp=pre treat visit rep treat*rep/ DDFM=KR;
```

The following `RANDOM` and `REPEATED` statements were added to specify the covariance pattern for each model. Fitting effects as random leads to compound symmetry covariance structures within each effect specified. For example, in Model 1, a compound symmetry structure is fitted within patients and patient·visits.

```
Model 1: RANDOM pat pat*visit;
Model 2: RANDOM pat pat*visit pat*rep;
Model 3: RANDOM visit/ SUB=pat TYPE=UN;
Model 4: RANDOM pat;
         REPEATED rep/ SUB=pat*visit TYPE=AR(1) R RCORR;
Model 5: RANDOM pat/ GROUP=treat;
           REPEATED rep/SUB=pat*visit TYPE=AR(1) R RCORR
           GROUP=treat;
```

We note that `TYPE` options are also available in `SAS` to fit covariance patterns structured by two effects (e.g. by visit and rep). However, the range of structures available is limited. The first-level structure (e.g. across visits) is always unstructured (`UN`), and the second-level structure (e.g. across reps) can be compound symmetry, first-order autoregressive or unstructured. A model with a similar covariance pattern to Model 3 could be fitted by replacing the `RANDOM` statement with

```
REPEATED visit rep/SUB=pat TYPE=UN@CS;
```

However, the resulting pattern is not identical because the compound symmetry pattern is fitted as a correlation term, rather than a covariance term.

Since the last edition of this text, the improved adjustment based on Kenward and Roger's 2009 publication, `DDFM=KR(LINEAR)`, has become available in SAS/STAT 12.1. In general, we would recommend its use when the AR(1) covariance pattern is used. However, in this example, there are no missing data, and use of the option leads to the same standard errors as the `DDFM=KR` option shown previously.

Model 4 code and output The full code and output is given as follows for Model 4 on which the conclusions were based.

```
PROC MIXED; CLASS pat visit treat rep;
MODEL sbp=pre treat visit rep treat*rep/ DDFM=KR;
RANDOM pat;
REPEATED rep/ SUB=pat*visit TYPE=AR(1) R RCORR;
LSMEANS treat/ PDIFF DIFF;
```

Model Information

Data Set	WORK.A
Dependent Variable	sbp
Covariance Structures	Variance Components, Autoregressive
Subject Effect	pat*visit
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Prasad-Rao-Jeske-Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

Class Level Information

Class	Levels	Values
pat	12	1 2 3 4 5 6 7 8 9 10 11 12
visit	3	1 2 3
treat	3	A B C
rep	8	2 3 4 5 6 7 8 9

Dimensions

Covariance Parameters	3
Columns in X	40
Columns in Z	12
Subjects	1
Max Obs Per Subject	288

Number of Observations

Number of Observations Read	288
Number of Observations Used	288
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	−2 Res Log Like	Criterion
0	1	2004.94119030	
1	2	1866.97552103	0.00000136
2	1	1866.97456592	0.00000000

Convergence criteria met.

Covariances in the **R** matrix are calculated as 55.6×0.323^d, where d = separation of reps.

Estimated R Matrix for pat*visit 1 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	55.6430	17.9844	5.8127	1.8787	0.6072	0.1963	0.06343	0.02050
2	17.9844	55.6430	17.9844	5.8127	1.8787	0.6072	0.1963	0.06343
3	5.8127	17.9844	55.6430	17.9844	5.8127	1.8787	0.6072	0.1963
4	1.8787	5.8127	17.9844	55.6430	17.9844	5.8127	1.8787	0.6072
5	0.6072	1.8787	5.8127	17.9844	55.6430	17.9844	5.8127	1.8787
6	0.1963	0.6072	1.8787	5.8127	17.9844	55.6430	17.9844	5.8127
7	0.06343	0.1963	0.6072	1.8787	5.8127	17.9844	55.6430	17.9844
8	0.02050	0.06343	0.1963	0.6072	1.8787	5.8127	17.9844	55.6430

Estimated R Correlation Matrix for pat*visit 1 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.3232	0.1045	0.03376	0.01091	0.003527	0.001140	0.000368
2	0.3232	1.0000	0.3232	0.1045	0.03376	0.01091	0.003527	0.001140
3	0.1045	0.3232	1.0000	0.3232	0.1045	0.03376	0.01091	0.003527
4	0.03376	0.1045	0.3232	1.0000	0.3232	0.1045	0.03376	0.01091
5	0.01091	0.03376	0.1045	0.3232	1.0000	0.3232	0.1045	0.03376
6	0.003527	0.01091	0.03376	0.1045	0.3232	1.0000	0.3232	0.1045
7	0.001140	0.003527	0.01091	0.03376	0.1045	0.3232	1.0000	0.3232
8	0.000368	0.001140	0.003527	0.01091	0.03376	0.1045	0.3232	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Pat		46.0598
AR(1)	pat*visit	0.3232
Residual		55.6430

Fit Statistics

−2 Res Log Likelihood	1867.0
AIC (smaller is better)	1873.0
AICC (smaller is better)	1873.1
BIC (smaller is better)	1874.4

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
pre	1	63.8	11.08	0.0015
treat	2	54.6	8.63	0.0006
visit	2	59.1	3.39	0.0402
rep	7	209	3.28	0.0025
treat*rep	14	219	1.10	0.3612

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	105.57	2.2285	13.5	47.37	<.0001
treat	B	108.75	2.2049	13.2	49.32	<.0001
treat	C	102.51	2.2440	13.8	45.68	<.0001

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	B	−3.1820	1.4457	52.3	−2.20	0.0322
treat	A	C	3.0540	1.6161	57.7	1.89	0.0638
treat	B	C	6.2360	1.5157	54.8	4.11	0.0001

8.1.3 Random coefficients models

Random coefficients models can also be utilised for analysing data with repeated measurements made within visits. This type of model would be more appropriate if greatest interest is centred on explaining the relationship of a measurement with time. However, time is now measured across both visits and reps, and it is therefore possible to consider fitting slopes across both of these time scales. Alternatively, slopes can be fitted across just one time scale, that is, against either visit time or rep time. In many applications, one of these simpler models will be of greatest relevance. We will now describe three possible models. Model 1 fits slopes against visit time only, Models 2(a) and 2(b) fit slopes against rep time only, and Model 3 fits slopes against both visit time and rep time. The SAS code required to fit each model will be supplied following Section 8.1.4.

Model 1 – modelling response against visit time

In some examples, only the relationship of the response with visit time may be of interest. For example, this might be the case for a trial where patients make varying numbers of visits to a hospital clinic at unevenly spaced intervals during the course of a treatment, with only two replicates of measurements per visit. In this situation, a slope against visit time can be fitted, and reps can be taken as categorical fixed effects. To allow the slopes to vary randomly between the patients, patient and patient·tvisit (tvisit = visit time) are fitted as random coefficients. The patient coefficients will represent the intercepts, and the patient·tvisit coefficients the slopes of separate regression lines for each patient. The average slopes will be determined by the fixed effects tvisit and treatment·tvisit. The random coefficients cause the regression lines (each made up of an intercept and slope) to be compared between treatments against an appropriate background of between-patient variability. It is unlikely that the

estimation of slopes against visit will be of interest in a cross-over trial; so this model will be of greatest use for analysing repeated measures designs.

Random coefficients	Fixed effects
Patient	Baseline
Patient·tvisit	Treatment
	Tvisit
	Rep
	Treatment·tvisit
	Treatment·rep

Note that tvisit represents the actual times of the visits (e.g. in weeks) rather than the visit number. However, in clinical trials with evenly spaced visits, use of the visit number will often suffice as the visit time.

Model 2 – modelling response against rep time

In some applications, the slope against rep time is of greatest interest. To allow the slopes to vary randomly between the patients, patient and patient·trep (trep = rep time) are fitted as random coefficients. In addition, to allow the slopes to vary randomly within the patients, patient·visit and patient·visit·trep (trep = rep time) are also fitted as random coefficients. Thus, a separate regression line is fitted for each patient at each visit. In repeated measures trials, slopes will be appropriately compared between treatments against a background of between-patient variability, and in cross-over trials, against a background of within-patient variability. Slope effects can also be compared between visits, and this is always against a background of within-patient variability.

Random coefficients	Fixed effects
Patient	Baseline
Patient·trep	Treatment
Patient·visit	Visit
Patient·visit·trep	Trep
	Treatment·visit
	Visit·trep
	Treatment·trep

Visit effects are still included in the model but as categorical fixed effects. In a cross-over study, fixed carry-over effects can also be included if required, though we advise caution in doing this (see Section 7.11). Note that the trep represents the actual times of the reps (e.g. in minutes) rather than the rep number (denoted by rep in Model 1).

Model 3 – modelling response against rep time and visit time

This model can be used when both the slopes against visit time and rep time are of interest. Separate slopes are fitted across visit time and rep time. To allow slopes to be compared between treatments against an appropriate background of between-patient variability, patient, patient·trep and patient·tvisit are fitted as random coefficients. As with Model 2, this model is appropriate for comparing slopes between treatments in repeated measures designs only. We do not suggest a corresponding model for cross-over trials because it is unlikely that the estimation of slopes against visit will be sensible in a cross-over trial.

Random coefficients	Fixed effects
Patient	Baseline
Patient·tvisit	Treatment
Patient·trep	Tvisit
	Trep
	Treatment·tvisit
	Treatment·trep

Non-linear models

We have in the previous section considered only linear relationships with time. However, the models can also be adapted to fit non-linear relationships if required, as discussed in Section 6.5.

Choice of model

The models we have introduced each fit different fixed effects and so cannot be compared using likelihood ratio tests. However, a statistical comparison of the models is not really relevant in this case, since the choice of model should depend on which model best answers the required questions. For this reason, statistical comparisons between random coefficients models and covariance pattern models are also not helpful, since the two approaches are designed to answer different questions.

8.1.4 Example: random coefficients models

We again consider the three-period cross-over trial introduced in Section 8.1.2. A random coefficients model might be chosen instead of a covariance pattern model if interest were principally focused on modelling the relationship of systolic blood pressure across either visits or reps. In this trial, it is likely that comparing the rep slopes between treatments will be of greatest interest, and therefore only Model 2 will be considered. This model examines whether the rates of change of systolic blood pressure during each phase of the cross-over differ between

Table 8.2 Covariance parameter estimates.

	Covariance parameters
Covariances for patient and patient·trep coefficients	$\begin{pmatrix} 63.6 & \\ -8.7 & 2.96 \end{pmatrix}$
Covariances for patient·visit and patient·visit·trep coefficients	$\begin{pmatrix} 182.5 & \\ -33.8 & 6.48 \end{pmatrix}$
Residual	42.9

Note: Here, trep = log (rep time).

treatments. It was found that taking the log of rep time gave a more linear relationship with systolic blood pressure. Thus, log(rep time) is used in place of rep time in the model.

The covariance parameter estimates are shown in Table 8.2. The positive diagonal terms in the two covariance matrices indicate that there is additional random variation in the regression lines both between patients and within patients over the set of visits (i.e. the regression lines differ to a greater extent than would be expected as a result of the residual variation). The negative covariance term is not surprising because of the expected negative correlation between estimates of slopes and intercepts in regression analysis. It is hard to assess the relative sizes of the covariance parameters, since the patient·trep and patient·visit·trep coefficients involve continuous effects of differing sizes. However, a comparison of the patient, patient·visit and residual variances indicates that most of the variation in the data is occurring at the patient·visit level.

The results for the treatment and treatment·rep effects are shown in Table 8.3. The rep slope estimates are positive for each treatment showing an increase in blood pressure following the administration of treatment. The overall test of the treatment·slope interaction was not significant ($p = 0.22$), neither were any of the pairwise comparisons of rep slopes between treatments. Note that identical standard errors are obtained in this case for each of the rep slopes because the data were complete.

Table 8.3 Rep slope estimates (standard errors).

Rep slopes		
A	1.41	(1.22)
B	4.11	(1.22)
C	2.04	(1.22)
Treatment differences in rep slope		
A – B	−2.70	(1.58)
A – C	−0.63	(1.58)
B – C	2.07	(1.58)

SAS code and output

Variables
pat = patient,
sbp = post-treatment systolic blood pressure (mm Hg),
pre = pre-treatment systolic blood pressure (mm Hg),
treat = treatment (A = 20 mg dose, B = 40 mg dose, C = placebo),
visit = visit number,
rep = rep,
trep = log(time since dose in minutes).

The SAS code and output are shown for Model 2, which was used to analyse this example. Following this, the code required for the other models described in Section 8.1.3. (Models 1 and 3) is shown.

```
PROC MIXED; CLASS pat treat visit;
TITLE 'MODEL 2B. SLOPES ACROSS REP TIME - RANDOM SLOPES
   FOR PATIENTS AND PATIENT.VISITS';
MODEL sbp=pre treat visit trep treat*visit treat*trep/
DDFM=KR;
RANDOM int trep / SUB=pat TYPE=UN;
RANDOM int trep / SUB=pat*visit TYPE=UN;
LSMEANS treat/ DIFF PDIFF;
ESTIMATE 'A, REP SLOPE' trep 1 treat*trep 1 0 0;
ESTIMATE 'B, REP SLOPE' trep 1 treat*trep 0 1 0;
ESTIMATE 'C, REP SLOPE' trep 1 treat*trep 0 0 1;
ESTIMATE 'A-B, REP SLOPE' treat*trep 1 -1 0;
ESTIMATE 'A-C, REP SLOPE' treat*trep 1 0 -1;
ESTIMATE 'B-C, REP SLOPE' treat*trep 0 1 -1;
```

The ESTIMATE statements are required to estimate the slope effects by treatment and their differences in this case. This is because the LSMEANS statement can only be used with CLASS variables.

```
                The Mixed Procedure
                 Model Information
Data Set                      WORK.A
Dependent Variable            sbp
Covariance Structure          Unstructured
Subject Effects               pat, pat*visit
Estimation Method             REML
Residual Variance Method      Profile
Fixed Effects SE Method       Prasad-Rao-Jeske-
                              Kackar-Harville
Degrees of Freedom Method     Kenward-Roger
```

Class Level Information

Class	Levels	Values
pat	12	1 2 3 4 5 6 7 8 9 10 11 12
treat	3	A B C
visit	3	1 2 3

Dimensions

Covariance Parameters	7
Columns in X	21
Columns in Z Per Subject	8
Subjects	12
Max Obs Per Subject	24

Number of Observations

Number of Observations Read	288
Number of Observations Used	288
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	−2 Res Log Like	Criterion
0	1	2068.81350873	
1	2	1940.47935288	0.00001352
2	1	1940.46924844	0.00000006
3	1	1940.46920458	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	pat	63.6303
UN(2,1)	pat	−8.6675
UN(2,2)	pat	2.9646
UN(1,1)	pat*visit	182.46
UN(2,1)	pat*visit	−33.8450
UN(2,2)	pat*visit	6.4842
Residual		42.9001

`UN(1,1)` and `UN(2,2)` are the variance component estimates for the patient and patient·trep random coefficients, and `UN(2,1)` is the covariance between the two random coefficients. Likewise, the second `UN(1,1)` and `UN(2,2)` are the variance component estimates for the patient·visit and patient·visit·trep random coefficients, and `UN(2,1)` is their covariance.

Fit Statistics

−2 Res Log Likelihood	1940.5
AIC (smaller is better)	1954.5
AICC (smaller is better)	1954.9
BIC (smaller is better)	1957.9

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	128.34	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Pre	1	25.9	7.69	0.0102
Treat	2	21.1	0.58	0.5680
visit	2	16.4	2.65	0.1004
trep	1	11	9.61	0.0101
treat*visit	4	18.4	0.48	0.7520
trep*treat	2	22	1.61	0.2234

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
A, REP SLOPE	1.4114	1.2207	31.3	1.16	0.2564
B, REP SLOPE	4.1118	1.2207	31.3	3.37	0.0020
C, REP SLOPE	2.0411	1.2207	31.3	1.67	0.1045
A-B, REP SLOPE	−2.7004	1.5768	22	−1.71	0.1008
A-C, REP SLOPE	−0.6297	1.5768	22	−0.40	0.6935
B-C, REP SLOPE	2.0707	1.5768	22	1.31	0.2026

The first three estimates give the rep slopes corresponding to each of the treatment groups. The last three estimates are of the differences between the slopes for each treatment and are the slope differences shown in Table 8.3.

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	105.42	2.3081	14.5	45.67	<.0001
treat	B	108.71	2.2658	14.1	47.98	<.0001
treat	C	102.70	2.3357	14.8	43.97	<.0001

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	B	−3.2906	1.7999	14.9	−1.83	0.0875
treat	A	C	2.7164	2.0506	17.9	1.32	0.2020
treat	B	C	6.0070	1.9035	16.2	3.16	0.0060

SAS code for Model 1

```
PROC MIXED; CLASS pat time treat;
TITLE 'MODEL 1. SLOPES ACROSS VISIT TIME ONLY';
MODEL sbp=pre treat tvisit time treat*tvisit treat*time/
DDFM=KR;
RANDOM INT tvisit / SUB=pat TYPE=UN;
LSMEANS treat/ DIFF PDIFF;
ESTIMATE 'A, VISIT SLOPE' tvisit 1 treat*tvisit 1 0 0;
ESTIMATE 'B, VISIT SLOPE' tvisit 1 treat*tvisit 0 1 0;
ESTIMATE 'C, VISIT SLOPE' tvisit 1 treat*tvisit 0 0 1;
ESTIMATE 'A-B, VISIT SLOPE' treat*tvisit 1 -1 0;
ESTIMATE 'A-C, VISIT SLOPE' treat*tvisit 1 0 -1;
ESTIMATE 'B-C, VISIT SLOPE' treat*tvisit 0 1 -1;
```

SAS code for Model 3

```
PROC MIXED; CLASS pat treat;
TITLE 'MODEL 3. SLOPES ACROSS VISIT TIME AND REP TIME';
MODEL sbp=pre treat tvisit trep treat*tvisit treat*trep
/DDFM=KR;
RANDOM INT tvisit trep / SUB=pat TYPE=UN;
LSMEANS treat/ DIFF PDIFF;
ESTIMATE 'A, VISIT SLOPE' tvisit 1 treat*tvisit 1 0 0;
ESTIMATE 'B, VISIT SLOPE' tvisit 1 treat*tvisit 0 1 0;
ESTIMATE 'C, VISIT SLOPE' tvisit 1 treat*tvisit 0 0 1;
ESTIMATE 'A, REP SLOPE' trep 1 treat*trep 1 0 0;
ESTIMATE 'B, REP SLOPE' trep 1 treat*trep 0 1 0;
ESTIMATE 'C, REP SLOPE' trep 1 treat*trep 0 0 1;
ESTIMATE 'A-B, VISIT SLOPE' treat*tvisit 1 -1 0;
ESTIMATE 'A-C, VISIT SLOPE' treat*tvisit 1 0 -1;
ESTIMATE 'B-C, VISIT SLOPE' treat*tvisit 0 1 -1;
ESTIMATE 'A-B, REP SLOPE' treat*trep 1 -1 0;
ESTIMATE 'A-C, REP SLOPE' treat*trep 1 0 -1;
ESTIMATE 'B-C, REP SLOPE' treat*trep 0 1 -1;
```

8.2 Multi-centre trials with repeated measurements

It is not uncommon for clinical trials to record measurements over several visits and also to recruit patients from several centres. The hypertension study introduced in Section 1.3 in fact has this structure. A mixed model can then be used to allow treatment effects to vary randomly across centres and also to fit a covariance pattern for the repeated measurements. As in ordinary multi-centre studies, treatments can be allowed to vary randomly across centres by fitting centre and centre·treatment effects as random, and inference is then wider and can be applied to the population of centres (see Chapter 5). The treatment standard errors will be increased compared with models omitting centre·treatment effects or fitting them as fixed, whenever the centre·treatment variance component is positive. Time and treatment·time effects can also be allowed to vary randomly between centres by fitting centre·time and centre·treatment·time effects as random. Alternatively, if a 'local' interpretation is required, interactions with centre effects can be omitted or taken as fixed. If the interaction terms involving centre are omitted, retaining centre effects as random in the model will still allow any additional information on treatments available from the centre stratum to be recovered. Covariance patterns for the repeated observations occurring on the same patients can be constructed in the same way as described for ordinary repeated measures trials and compared using likelihood ratio tests (see Section 6.2.2).

8.2.1 Example: multi-centre hypertension trial

The multi-centre hypertension trial introduced in Section 1.3 had visits at 2, 4, 6 and 8 weeks post-treatment. We will initially analyse the primary endpoint, DBP, using the following mixed model:

Random effects	**Fixed effects**
Centre	Treatment
Centre·treatment	Time
Centre·time	Treatment·time
Centre·treatment·time	Baseline
Patient	

Note that a compound symmetry structure for the repeated measurements is obtained in this case by fitting patients as random.

The resulting variance component estimates were as follows:

Effect	Variance component
Centre	4.4
Centre·treat	0.5
Visit·centre	1.0
Visit·centre·treat	0.5
Patient	35.0
Residual	34.3

These values indicate that most of the variation occurs at the patient and residual level. The very small variance components for treatment·centre, visit·centre and visit·centre·treatment effects will influence the results very little and will be removed from the analysis in order to simplify the model. The small centre·treatment variance component indicates that treatments are varying hardly at all across the centres. Thus, results can still be related to the population of centres with some confidence, even when centre·treatment effects are removed. However, retaining the centre effects as random allows additional information on treatments to be recovered from the centre error stratum. The fixed treatment·time interaction was not significant, and we will omit it from the model.

8.2.2 Covariance pattern models

We can also consider fitting covariance patterns to the repeated measurements, although in many situations, a constant compound symmetry covariance (achieved by fitting patients as random) will be sufficient. As in ordinary repeated measures trials, models using different covariance patterns can be compared using likelihood ratio tests (see Section 6.2.2). The six models discussed in the following section were fitted to the hypertension data. In each model, centre effects are fitted as random, and treatment, time and baseline effects are fitted as fixed. The other terms that were considered in the initial model but found to be negligible have been omitted. The covariance parameters estimated in each model are shown in Table 8.4 and values of $-2\log(L)$ in Table 8.5.

The comparison of the covariance pattern models is quite similar to that presented in Section 6.3, where the effects of centre were not considered. Model 2 has the same number of covariance parameters as Model 1, and a direct comparison of their likelihoods can be made. Model 1 has the highest likelihood, and therefore we have no evidence that correlations decay exponentially as visits become further apart. The likelihood ratio tests comparing Models 3–6 with Model 1 are all highly significant, indicating that all these models are preferable

Table 8.4 Variance parameters from covariance pattern models.

		Variance components		
Model	**Period**	**Centre**	**Residual**	**Correlation**
1. Compound symmetry	1 – 4	4.8	71.4	0.52
2. First-order autoregressive	1	5.1	70.8	$\begin{pmatrix} 1 & & & \\ 0.53 & 1 & & \\ 0.53^2 & 0.53 & 1 & \\ 0.53^3 & 0.53^2 & 0.53 & 1 \end{pmatrix}$
	2			
	3			
	4			
3. Compound symmetry with separate covariances for treatments	1 – 4 A	4.6	78.7	0.49
	1 – 4 B		63.5	0.35
	1 – 4 C		72.6	0.61
4. Toeplitz	1	4.8	71.3	$\begin{pmatrix} 1 & & & \\ 0.54 & 1 & & \\ 0.45 & 0.54 & 1 & \\ 0.42 & 0.45 & 0.54 & 1 \end{pmatrix}$
	2			
	3			
	4			
5. Unstructured	1	4.5	73.2	$\begin{pmatrix} 1 & & & \\ 0.49 & 1 & & \\ 0.44 & 0.57 & 1 & \\ 0.42 & 0.47 & 0.56 & 1 \end{pmatrix}$
	2		66.8	
	3		80.0	
	4		66.4	
6. Toeplitz with separate covariances for treatments	1 A	4.7	78.5	$\begin{pmatrix} 1 & & & \\ 0.54 & 1 & & \\ 0.45 & 0.54 & 1 & \\ 0.46 & 0.45 & 0.54 & 1 \end{pmatrix}$
	2 A			
	3 A			
	4 A			
	1 B		63.4	$\begin{pmatrix} 1 & & & \\ 0.38 & 1 & & \\ 0.29 & 0.38 & 1 & \\ 0.40 & 0.29 & 0.38 & 1 \end{pmatrix}$
	2 B			
	3 B			
	4 B			
	1 C		72.1	$\begin{pmatrix} 1 & & & \\ 0.68 & 1 & & \\ 0.60 & 0.68 & 1 & \\ 0.43 & 0.60 & 0.68 & 1 \end{pmatrix}$
	2 C			
	3 C			
	4 C			

to Model 1. Models 3 and 4 are not nested, and therefore cannot be compared using a likelihood ratio test. Model 4 is nested within Model 5, and the likelihood ratio test statistic is $6.78 \sim \chi^2_6$. This is not significant, and therefore Model 5 can be rejected. Models 3 and 4 are nested within Model 6, and the likelihood ratio statistics for comparisons with Model 6 are $22.47 \sim \chi^2_6$ and $27.92 \sim \chi^2_8$, which are both highly significant. Since Model 6 is a significant improvement over both Models 3 and 4, we might choose to base our conclusions on it. However, we note that in practice it may be difficult to justify experimentation with so many structures, particularly in regulatory trials where it is necessary to specify analysis methods in the protocol.

Table 8.5 Values of $-2\log(L)$ for covariance pattern models.

Model	$-2\log(L)$	Covariance parameters
1. Compound symmetry	7483.21	3
2. First-order autoregressive	7492.70	3
3. Compound symmetry separate treatment covariances	7454.48	7
4. Toeplitz	7459.93	5
5. Unstructured	7453.15	11
6. Toeplitz separate treatment covariances	7432.01	13

Table 8.6 Treatment effect estimates from Model 6 and corresponding estimates from Section 6.3 without centre effects.

Difference	Mean difference	SE
Model 6 including centre effects		
A – B	1.26	0.94
A – C	3.01	1.02
B – C	1.75	0.95
Model 6 excluding centre effects		
A – B	1.23	0.99
A – C	3.02	1.07
B – C	1.79	1.00

The treatment effects from Model 6 are given in Table 8.6 with model-based standard errors. They indicate that treatment C produces a significantly lower DBP than treatment A. Ideally, we would like additionally to calculate empirical standard errors to reassure ourselves further that the covariance pattern fitted by Model 6 is close to the observed covariance in the data. However, an error message occurred when these standard errors were requested using `PROC MIXED`. Model 6 in Section 6.3 used an identical covariance pattern to Model 6 in this section but did not include centre effects. This model gave similar treatment effect results but with larger standard errors, indicating that fitting centre effects as random had led to some recovery of extra information on the treatments.

`SAS` code and output

Variables

`pat` = patient,
`centre` = centre,
`visit` = visit,
`treat` = treatment (A,B,C),
`dbp` = post-treatment diastolic blood pressure (mmHg),
`dbp1` = pre-treatment diastolic blood pressure (mmHg).

The following first three statements were used in all the models. The RANDOM and REPEATED statements to set the covariance structure are then shown for each model. The full code and output are given only for Model 6 on which the conclusions were based.

```
PROC MIXED NOCLPRINT; CLASS visit centre pat treat;
MODEL dbp=dbp1 treat visit/ DDFM=KR;
LSMEANS treat/ DIFF PDIFF;

Model 1: RANDOM centre;
         REPEATED visit/SUB=pat TYPE=CS GROUP=treat;
Model 2: RANDOM centre;
         REPEATED visit/SUB=pat TYPE=AR(1);
Model 3: RANDOM centre;
         REPEATED visit/SUB=pat TYPE=CS GROUP=treat;
Model 4: RANDOM centre;
         REPEATED visit/SUB=pat TYPE=TOEP;
Model 5: RANDOM centre;
         REPEATED visit/SUB=pat TYPE=UN;
Model 6: RANDOM centre;
         REPEATED visit/SUB=pat TYPE=TOEP GROUP=treat;
```

Model 6

```
PROC MIXED NOCLPRINT; CLASS visit centre pat treat;
MODEL dbp=dbp1 treat visit centre/ DDFM=KR;
RANDOM centre;
REPEATED visit/ SUB=pat TYPE=TOEP GROUP=treat;
LSMEANS treat/ DIFF PDIFF;
```

Model Information

Data Set	WORK.A
Dependent Variable	dbp
Covariance Structures	Variance Components, Toeplitz
Subject Effect	pat
Group Effect	treat
Estimation Method	REML
Residual Variance Method	None
Fixed Effects SE Method	Prasad-Rao-Jeske-Kackar-Harville
Degrees of Freedom Method	Kenward-Roger

Dimensions

Covariance Parameters	13
Columns in X	9
Columns in Z	29
Subjects	1
Max Obs Per Subject	1092

Number of Observations

Number of Observations Read	1092
Number of Observations Used	1092
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	−2 Res Log Like	Criterion
0	1	7814.58209591	
1	3	7432.27303121	0.00008952
2	1	7432.01391988	0.00000111
3	1	7432.01088005	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
centre			4.7466
Variance	pat	treat A	78.4613
TOEP(2)	pat	treat A	41.6205
TOEP(3)	pat	treat A	35.0193
TOEP(4)	pat	treat A	35.5072
Variance	pat	treat B	63.4263
TOEP(2)	pat	treat B	24.2283
TOEP(3)	pat	treat B	18.0183
TOEP(4)	pat	treat B	24.5094
Variance	pat	treat C	72.1075
TOEP(2)	pat	treat C	48.6675
TOEP(3)	pat	treat C	42.5151
TOEP(4)	pat	treat C	30.6658

Fit Statistics

−2 Res Log Likelihood	7432.0
AIC (smaller is better)	7458.0
AICC (smaller is better)	7458.4
BIC (smaller is better)	7475.8

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	283	28.31	<.0001
treat	2	174	4.36	0.0143
visit	3	433	12.75	<.0001

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	93.3399	0.8633	71.8	108.12	<.0001
treat	B	92.0772	0.7822	51.1	117.71	<.0001
treat	C	90.3293	0.8743	71.9	103.32	<.0001

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|
treat	A	B	1.2626	0.9361	172	1.35	0.1792
treat	A	C	3.0106	1.0214	175	2.95	0.0036
treat	B	C	1.7479	0.9541	169	1.83	0.0687

8.3 Multi-centre cross-over trials

Multi-centre trials are most frequently used for demonstrating the effectiveness of a drug in its later stages of development and usually have between-patient designs. However, occasionally, multi-centre cross-over trials will be encountered. Although we do not show an example of this design, we suggest how a mixed model can be applied.

As in ordinary multi-centre trials, 'global' estimates can be obtained by allowing treatment effects to vary randomly across centres. This is achieved by fitting centre and centre·treatment effects as random. In addition, period effects can be allowed to vary randomly across the centres by fitting centre·period effects as random. A 'local' model (Model A) where results relate only to the centres sampled can be obtained by fitting interactions with centre effects as fixed. If the interaction terms involving centre can safely be removed from the model, then there can be an advantage in setting the centre effects to random to allow any additional information on treatments and periods to be recovered from the between-centre error stratum (Model B).

The fixed and random effects that are fitted in global and local models are listed in the following section. Fitting patients as random gives a constant correlation for observations on the same patient. Alternatively, a more complex covariance pattern can be fitted by using a covariance pattern model (see Section 7.7).

Global model

Random	Fixed
Centre	Treatment
Centre·treatment	Period
Centre·period	Baseline
Patient	

Local model (A)

Random	Fixed
Patient	Treatment
	Period
	Centre
	Centre·treatment
	Centre·period
	Baseline

Local model (B)

Random	Fixed
Centre	Treatment
Patient	Period
	Baseline

8.4 Hierarchical multi-centre trials and meta-analysis

Sometimes, centres within a multi-centre trial, or trials within a meta-analysis, may be grouped in some way: for example, by country or continent. Such trials can then be described as having a double hierarchical structure. Although we do not consider an example with this design, we suggest how a mixed model can be applied. The double hierarchical structure can be taken into account in the analysis by modelling both hierarchies and their interactions with treatment effects as random. For example, in a multi-centre trial with centres grouped by

country, a 'global' model taking into account possible random variation in the treatment effect between centres and countries would be as follows:

Random	**Fixed**
Centre	Treatment
Centre·treatment	Baseline
Country	
Country·treatment	

This model will allow treatment effect results to be related with more confidence to the potential population of centres and countries, and 'shrunken' estimates of treatment effects can be obtained for each country and centre. Note that it is not necessary in the mixed model to specify formally that countries are nested within centres, provided each centre is numbered individually. However, if centres are numbered separately within each country, specifying centre(country) in the software will usually avoid the need to create a variable with separate numbers for each centre (this is the case with `PROC MIXED`). In a meta-analysis with trials grouped by country, trial effects could be substituted for centre effects in the above model. Note though that some meta-analysts might prefer the trial effect to be fixed and only the trial·treatment effect to be random, as noted in Chapter 5. There are also other possibilities for double hierarchical structures in which this type of analysis can be used. For example, trials within a meta-analysis may each have multi-centre designs, and the model can be fitted with trial effects substituted for country effects.

8.5 Matched case–control studies

In a matched case–control study, a group of subjects who have a particular disease or outcome (cases) is compared with a group of subjects who do not have the disease or outcome (controls). Each case is matched to one or more controls using one or more factors that are known to be connected with the disease; for example, age and sex are often used. We will refer to sets of matched subjects as 'matched sets'. The primary objective in a case–control analysis is to determine which factors (not used in matching) differ between the case and control groups. However, in doing so, it is important to allow for the matched nature of the data. This can be achieved by taking candidate risk factors as outcome variables and by fitting matched sets as a random effect.

The design of the matched case–control study has similarities with that of the cross-over trial. In the cross-over trial, the treatment effects are 'crossed' with patient effects (i.e. each patient may receive several treatments). In the matched case–control study, group effects (i.e. whether case or control) are crossed with

matched set effects (each matched set may contain cases and controls). The effect of fitting matched sets as random in a case–control study is similar to that of fitting patients as random in a cross-over study. Results will be identical to an analysis fitting matched sets as a fixed effect when there are the same number of controls for every case and the matched set variance component is positive (when it is negative and set to zero, the mean group differences will be identical but their standard errors will differ). In an analysis fitting matched sets as fixed, information is completely lost on group effects in any matched sets that contain either only a case or only controls (although matched sets containing two or more controls, but no cases can contribute information if the model fits at least one effect in addition to the group effect). In addition, in a fixed effects analysis of binary data (which can be performed using conditional logistic regression), matched sets whose members all have identical outcomes (i.e. are uniform) do not contribute information to the analysis. This loss of information does not occur when matched sets are fitted as random because information is then 'recovered' from the matched sets error stratum.

Another fixed effects approach sometimes used is to fit the matching variables (e.g. sex and age) as covariates but otherwise to ignore the matching. However, this can sometimes cause a bias in the group estimates if the matching variables are associated with group (i.e. case or control) and there are an uneven number of controls per case, causing the groups to be unbalanced. Results may also be misleading if the relationship with quantitative matching variables such as age is non-linear.

8.5.1 Example

This study was carried out by the Scottish Cot Death Trust, which aimed to interview the parents of every baby with sudden infant death syndrome (SIDS) in Scotland between 1992 and 1995 (Brooke *et al.*, 1997). The parents of two matched control babies born immediately before and after each case at the same hospital were also interviewed. As with most interview studies, not all parents agreed to participate, and this caused some of the matched sets to be incomplete. A summary of the content of the matched sets is given as follows for the interviewed subjects. Only 65% of the matched sets had their full complement of subjects.

Matched set content	Number of matched sets (%)
ABB	108 (65)
AB	128 (17)
A	11 (7)
BB	12 (7)
B	17 (4)

Note: A = case, B = control.

8.5.2 Analysis of a quantitative variable

We consider analysing a social deprivation score (`depcat`) that is measured on a scale of one to seven and is derived from post codes using information given in the 1991 Census (Carstairs and Morris, 1991). The distribution of `depcat` by group is shown in Figure 8.1. Although the score is an ordered categorical variable, we analyse it using a normal mixed model, and then check that this is a reasonable approximation by examining residual plots.

Results from a random effects model are compared with those from three alternative fixed effects approaches. In Model 1, group (case or control) is fitted as fixed and matched set as random. Model 2 allows for the matching by fitting matched sets as fixed. Model 3 fits the matching variables, age and season (at death/interview), as fixed effects. Season is categorised as Summer (June–August), Winter (December–February) or Spring/Autumn (March–May, September–November). Model 4 simply treats the data as unmatched.

The results are shown in Table 8.7. From all the models, we would conclude that SIDS is associated with deprivation, since the cases on average have increased `depcat` scores. The residual estimates in Models 3 and 4 are greatly increased over those in Models 1 and 2, indicating that for analyses of `depcat`, it is important to allow for matching.

Model 1 fits matched sets as random, and the positive variance component shows that `depcat` scores are positively correlated within the matched sets. This is likely

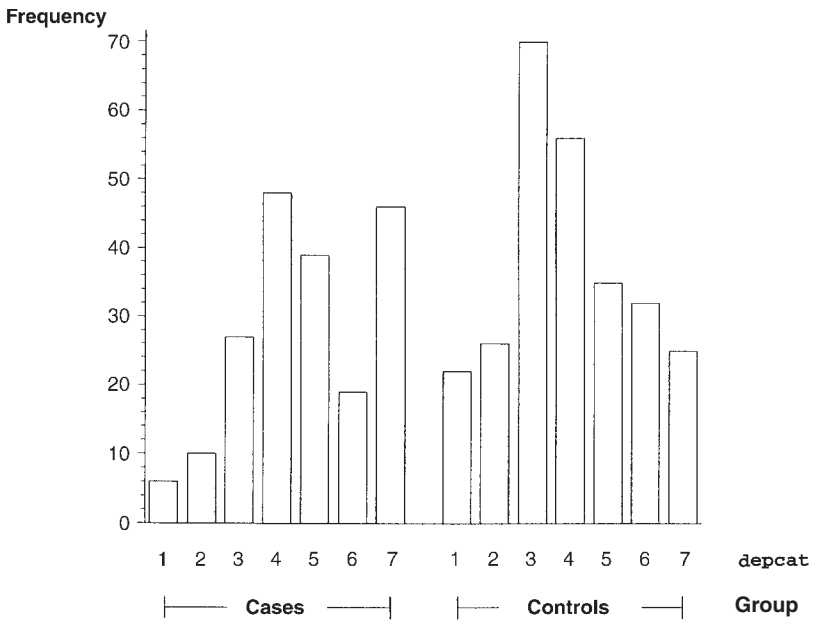


Figure 8.1 Histogram of deprivation category (`depcat`).

Table 8.7 Results from analyses of deprivation category (depcat).

Model	Fixed effects	Random effects	Method
1	Group	Matched set	REML
2	Group, matched set	–	OLS
3	Group, age, season	–	OLS
4	Group	–	OLS
Variance components			
Model	**Matched set**	**Residual**	
1	0.96	1.86	
2	–	1.87	
3	–	2.83	
4	–	2.82	
Group difference and SE (cases–controls)			
Model			
1	0.84 (0.13)		
2	0.88 (0.14)		
3	0.82 (0.16)		
4	0.82 (0.16)		

to be because matching is carried out within hospitals, which each have different catchment areas reflecting different levels of deprivation.

Model 2 takes account of the matching by fitting matched set as fixed. The group estimate differs from that in Model 1, and its standard error is larger. This is mainly because Model 2 does not use information from the matched set error stratum, and the 30 matched sets that contain either just cases or just controls are effectively lost from the analysis.

In Model 3, adjustment is made for the two matching variables, age and season of case birth. Neither of these is significant ($p = 0.76$ and 0.55, respectively), and the group estimates are the same as in Model 4, which ignores matching.

8.5.3 Check of model assumptions

Since `depcat` is an ordered categorical variable, we should check model assumptions to ensure that a normal mixed model is reasonable. The residuals from Model 1 are plotted against their predicted values in Figure 8.2 and appear evenly distributed. Note that the diagonal pattern is caused by the fact that the score is an ordered categorical variable. The agreement with normality is quite reasonable whether assessed by a histogram of the residuals or by a normal plot. The normal plot of the matched set effects in Figure 8.3 indicates no significant

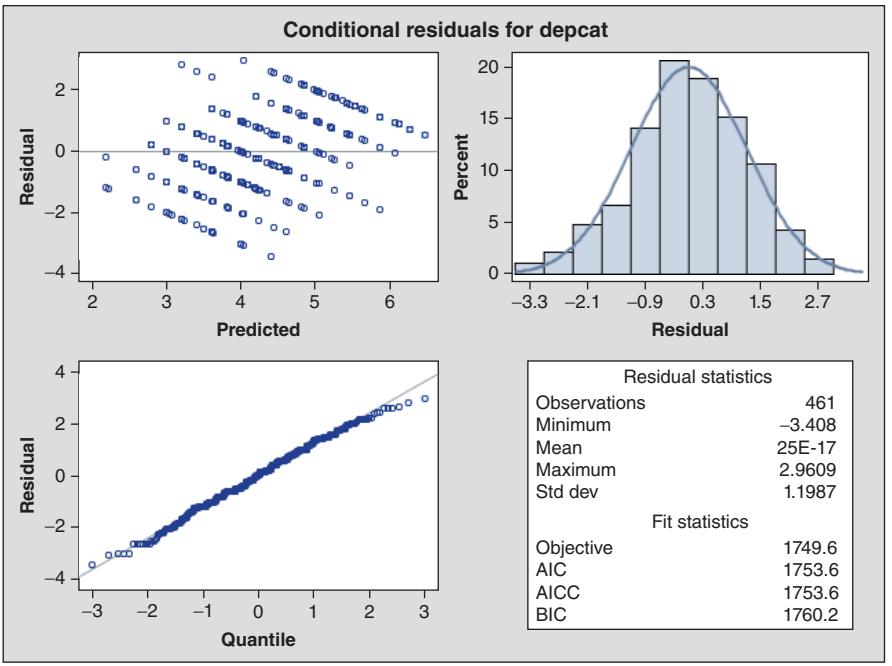


Figure 8.2 Panel of conditional residual plots.

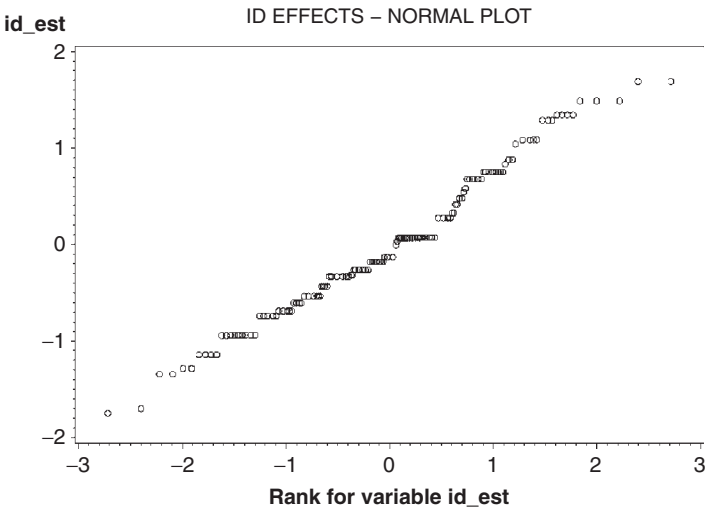


Figure 8.3 Normal plot of matched set effects.

deviation from normality or outlying matched sets. However, as cautioned in Section 4.6.4, the predicted random effects will not show up non-normality in all situations.

8.5.4 Analysis of binary Variables

In this section, we consider analysing two binary variables recorded in the cot death study – the sex of the infant and whether the infant slept in a cot. Now that the analysis variable is binary, several uniform categories (see Section 3.2.3) are likely to occur for the matched set effects (i.e. there will be several matched sets in which all subjects have the same response, e.g. are of the same sex). In the random effects model, this may cause bias in the matched sets variance component, and we will therefore reparameterise the model as a covariance pattern model with a compound symmetry covariance structure (see Sections 3.2.3 and 3.3.2). This approach is usually preferable to a conditional logistic regression model (see Section 3.1.8), where information is lost on all tied matched sets (i.e. any sets with all members answering 'no' or all members answering 'yes') and on any matched sets that contain either only a case or only controls.

A similar set of models is considered to those used for analysing `depcat` (see Table 8.8) except that Model 2 is fitted using conditional logistic regression

Table 8.8 Results from analyses of sex of the infant and sleeps in cot.

Model	Fixed effects	Random effects	Method
1	Group	Matched set	GLMM
2	Group, (matched set)		Conditional LR
3	Group, age, season		GLM
4	Group		GLM

Group effect on logit scale (case−controls) (SE) and odds ratio (cases ÷ controls) (95% confidence interval)

	Male		Sleeps in cot	
Model	On logit scale	OR	On logit scale	OR
1	0.78 (0.19)	2.18 (1.50, 3.17)	−0.67 (0.18)	0.51 (0.36, 0.73)
2	0.61 (0.21)	1.84 (1.22, 2.77)	−1.04 (0.29)	0.35 (0.20, 0.62)
3	0.77 (0.20)	2.15 (1.47, 3.15)	−1.01 (0.27)	0.37 (0.22, 0.62)
4	0.78 (0.19)	2.18 (1.50, 3.17)	−0.65 (0.22)	0.52 (0.34, 0.79)

Variance parameters for Model 1

Parameter	Male	Sleeps in cot
Within matched set correlation	0.00	0.29
Dispersion	1.00	1.00

(see Section 3.1.8) and conditions the data on the matched sets. We will consider the results obtained for sex and sleeping in a cot separately.

Sex

The within-matched-set correlation parameter was 0.00 in Model 1. This indicates, not surprisingly, that control babies were no more likely to be of the same sex as the case baby than of the opposite sex. The zero correlation parameter causes the results from Model 1 to be the same as Model 4 where matching is completely ignored. We should mention that since Model 1 was reparameterised as a covariance pattern model, a negative correlation estimate will not be set to zero by default. When a negative variance correlation occurs, it would be advisable to remove the effect of matched sets from the analysis by fitting Model 4 (effectively setting the correlation parameter to zero).

All models show statistically significant group effects, with a greater chance of SIDS in male infants. However, the group estimate in Model 2 is lower than in the other models and the standard error higher. This is because information from the matched set error stratum is lost from matched sets with all cases or all controls or with the same sex for all members. Age and season were not significantly related to sex in Model 3 ($p = 0.21$ and 0.76), and this leads to only slight differences in the results compared with Model 4.

Sleeps in cot

The within-matched-set correlation parameter is now positive, and this causes the results to differ between Models 1 and 4. It indicates that control infants were more likely to sleep in a cot if the case infant did and vice versa. This is not surprising since the infants were matched by age, and older babies are more likely to sleep in cots.

All of the models show a statistically significant group effect, indicating that the risk of SIDS is less in those infants who were sleeping in a cot compared with those who slept elsewhere (in carrycots, Moses baskets, prams, etc.). However, the group effect differs widely between the four models. This is mainly because sleep place is associated with age, and each model allows for this association in a different way. In Models 1 and 2, the exact matching is taken into account, and the results only differ because information from the matched set error stratum is completely omitted in Model 2 from the matched sets containing all cases or all controls or all with the same sleep place. The age effect was highly significant in Model 3 ($p = 0.0001$). However, this model fits age as a quantitative variable and therefore does not allow for any non-linear effects (i.e. it only allows the proportion of babies sleeping in a cot to increase linearly with age on the logistic scale). Model 4 did not fit age at all and is clearly inappropriate, given the influence that age has on sleep place. At first sight, it is therefore surprising that the standard error in Model 4 is smaller than that in Model 3. This has occurred because the pattern of 'dropouts'

from the study has resulted in cases and controls being unbalanced for age, with a resultant increase in the standard error in Model 3.

Analysis of both variables demonstrates the value of the mixed models approach. The standard errors of the group effect are minimised, as we might expect, because the data are being utilised more fully than is the case with any of the fixed effects approaches.

SAS code and output

Depcat analyses

Variables
group = group (A = cases, B = controls),
id = matched set,
depcat = deprivation score (1–7, 7 = most deprived),
age = age (weeks),
seas = 1 = Winter, 2 = Spring/Autumn, 3 = Summer.

Full code and output are given for Model 1. SAS code only is given for Models 2–4.

Model 1

```
PROC MIXED DATA=a NOCLPRINT PLOTS=RESIDUALPANEL;
     CLASS group id;
MODEL depcat= group/ DDFM=KR;
RANDOM ID;
ESTIMATE 'A-B' group 1 -1;
```

```
                    Model Information
       Data Set                        WORK.A
       Dependent Variable              depcat
       Covariance Structure            Variance Components
       Estimation Method               REML
       Residual Variance Method        Profile
       Fixed Effects SE Method         Prasad-Rao-Jeske-
                                       Kackar-Harville
       Degrees of Freedom Method       Kenward-Roger

                      Dimensions
              Covariance Parameters           2
              Columns in X                    3
              Columns in Z                  201
              Subjects                        1
              Max Obs Per Subject           477
```

Number of Observations

Number of Observations Read	477
Number of Observations Used	461
Number of Observations Not Used	16

Iteration History

Iteration	Evaluations	−2 Res Log Like	Criterion
0	1	1789.12644328	
1	2	1749.59861552	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
id	0.9576
Residual	1.8602

Fit Statistics

−2 Res Log Likelihood	1749.6
AIC (smaller is better)	1753.6
AICC (smaller is better)	1753.6
BIC (smaller is better)	1760.2

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	1	315	40.11	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
A-B	0.8446	0.1334	315	6.33	<.0001

Model 1 – model checking

```
ODS HTML FILE=<output file.html> GPATH=<directory>;
ODS GRAPHICS ON;
PROC MIXED DATA=a NOCLPRINT; CLASS group id;
MODEL depcat= group/ RESIDUAL DDFM=KR;
RANDOM id/ SOLUTION;
ESTIMATE 'A-B' group 1 -1;
ODS LISTING EXCLUDE SOLUTIONR;
ODS OUTPUT SOLUTIONR=solut; RUN;
ODS GRAPHICS OFF; ODS HTML CLOSE;
```

```
PROC RANK data=SOLUT OUT=norm NORMAL=TUKEY DATA=id_est;
  VAR estimate; RANKS rank;
PROC GPLOT DATA=norm; PLOT estimate*rank;
```

Figure 8.2 shows part of the output generated by the option `PLOTS = RESIDUALPANEL` in the code for Model 1. Similar output would be produced by the code in the model checking section. Figure 8.3 cannot be generated directly from `PROC MIXED` and is obtained from the code just shown.

Model 2

```
PROC GLM; CLASS group id;
MODEL depcat= group id;
ESTIMATE 'A-B' group 1 -1;
```

Model 3

```
PROC GLM; CLASS group id seas;
MODEL depcat= group age seas;
ESTIMATE 'A-B' group 1 -1;
```

Model 4

```
PROC GLM; CLASS group;
MODEL depcat= group;
ESTIMATE 'A-B' group 1 -1;
```

Binary analyses

`SAS` code is given in the following section for the analyses of 'sleeps in cot'. Identical code is used for the sex analyses with '`sex`' replacing '`sleepn1`' in the `MODEL` statement. Output is only given for Models 1 and 2 to illustrate the use of different procedures to fit each type of model. The output for Models 3 and 4 is similar to that from Model 1.

Variables

`grp` = group (1 = cases, 2 = controls),
`id` = matched set,
`one` = 1 for all observations,
`sleepn1` = sleeps in cot at night (1 = yes, 0 = no).

Model 1

```
PROC GLIMMIX; CLASS id grp;
MODEL sleepn1/one=grp / LINK=LOGIT DDFM=KR SOLUTION OR CL;
RANDOM INTERCEPT/ SUBJECT=id TYPE=CS RESIDUAL;
```

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	2	1785.8341804	0.16337371	9.27E-7
1	0	2	1771.2181425	0.00397342	4.03E-7
2	0	2	1771.494358	0.00003078	5.362E-9
3	0	0	1771.4950761	0.00000013	4.703E-6
4	0	0	1771.4950787	0.00000000	4.721E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics

−2 Res Log Pseudo-Likelihood	1771.50
Generalized Chi-Square	290.03
Gener. Chi-Square / DF	0.69

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
CS	id	0.3138	0.07081
	Residual	0.6922	0.06138

Solutions for Fixed Effects

Effect	grp	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept		−0.08622	0.1353	231.4	−0.64	0.5245	0.05
Grp	1	−0.6732	0.1825	279.8	−3.69	0.0003	0.05
Grp	2	0	.1825	.8	.69	.5245	.05

Effect	grp	Lower	Upper
Intercept		−0.3527	0.1803
grp	1	−1.0324	−0.3141
grp	2	.0324	.3141

Odds Ratio Estimates

Effect	grp	_grp	Estimate	DF	95% Confidence Limits	
grp	1	2	0.51	279.8	0.356	0.730

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
grp	1	279.8	13.61	0.0003

Model 2

A conditional logistic regression analysis can be performed by including a STRATA statement in the PROC LOGISTIC procedure. The EVENT option specifies the order for comparing 'sleepn1' effects and the CLODDS option causes odds ratios and their confidence intervals to be produced.

```
PROC LOGISTIC; CLASS id grp;
MODEL sleepn1(EVENT='1') = grp/ CLODDS=WALD;
STRATA id;
```

```
                    The LOGISTIC Procedure
                     Conditional Analysis
                       Model Information

    Data Set                            WORK.A
    Response Variable                   sleepn1
    Number of Response Levels           2
    Number of Strata                    166
    Number of Uninformative Strata      97
    Frequency Uninformative             226
    Model                               binary logit
    Optimization Technique              Newton-Raphson ridge

              Number of Observations Read  477
              Number of Observations Used  421

                      Response Profile
              Ordered                      Total
              Value       sleepn1      Frequency
              1              0            246
              2              1            175

Probability modeled is sleepn1=1.
NOTE: 56 observations were deleted due to missing values for
  the response, explanatory, or strata variables.

                   Class Level Information
                                        Design
              Class       Value       Variables
              grp           1              1
                            2             -1
```

Strata Summary

Response Pattern	sleepn1 = 0	sleepn1 = 1	Number of Strata	Frequency
1	0	1	8	8
2	1	0	11	11
3	0	2	16	32
4	1	1	12	24
5	2	0	11	22
6	0	3	13	39
7	1	2	27	81
8	2	1	30	90
9	3	0	38	114

Newton-Raphson Ridge Optimization
Without Parameter Scaling
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
AIC	141.877	129.273
SC	141.877	133.315
−2 Log L	141.877	127.273

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14.6047	1	0.0001
Score	14.0950	1	0.0002
Wald	13.1683	1	0.0003

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
grp	1	13.1683	0.0003

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
grp	1	1	−0.5216	0.1437	13.1683	0.0003

```
                Odds Ratio Estimates and Wald Confidence Intervals
Effect          Unit      Estimate     95% Confidence     Limits
grp 1 vs 2     1.0000       0.352          0.201          0.619
```

Note that the 'grp' effect in this case is half of that given in Table 8.8. This is due to the different parameterisation used in PROC LOGISTIC.

Model 3

```
PROC GLIMMIX;
MODEL sleepn1/one = grp seas age/LINK = LOGIT SOLUTION OR CL;
```

Model 4

```
PROC GLIMMIX;
MODEL sleepn1/one = grp/ LINK = LOGIT SOLUTION OR CL;
```

8.6 Different variances for treatment groups in a simple between-patient trial

In a simple between-patient trial, the treatment groups will sometimes have different variances. Allowing for this can produce more appropriate standard errors for treatment estimates, and also the variance values themselves may aid the understanding of the different treatment mechanisms. Often, the possibility of different treatment variances is not considered when choosing an analysis model. However, it can easily be allowed for in a mixed model by structuring the residual matrix to have a different variance for each treatment group. Observations will remain uncorrelated (provided no random effects are specified), and thus the model retains many of the features of a fixed effects analysis. Consider the example data used to illustrate the notation in Section 2.1.

Centre	Treatment	Baseline systolic BP	Post-treatment systolic BP
1	A	178	176
1	A	168	194
1	B	196	156
1	B	170	150
2	A	165	150
2	B	190	160
3	A	175	150
3	A	180	160
3	B	175	160

If different variances are allowed for the treatment groups, the variance matrix would have the form (assuming no random effects are fitted)

$$\mathbf{V} = \mathbf{R} = \begin{pmatrix} \sigma_A^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_A^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_B^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_A^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_B^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_A^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_A^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_B^2 \end{pmatrix}.$$

8.6.1 Example

We now consider the hypertension trial introduced in Section 1.3 as a simple between-patient trial. The final DBP measurement is analysed, and centre effects are ignored. A model with separate variances for each treatment group (Model 2) is compared with a standard model that assumes a constant variance over all treatments (Model 1). Both models fit baseline and treatment effects as fixed. The results obtained are given in Table 8.9.

Model 2 indicates that treatment A may be more variable than treatments B and C. A likelihood ratio test is used to test whether this model is significantly better

Table 8.9 Results from Models 1 – 3.

Model	Treatment	Residual variance	Centre variance component	Centre. treatment variance component	Log (*L*)
1	A – C	79.8	–	–	−1036.15
2	A	104.3	–	–	−1031.72
	B	56.5			
	C	76.9			
3	A	92.4	5.76	4.43	−1022.47
	B	44.5			
	C	66.7			

	Fixed effects (SEs)			
Model	**Baseline**	**A−B**	**A−C**	**B−C**
1	0.30 (0.11)	1.01 (1.29)	3.04 (1.28)	2.04 (1.31)
2	0.31 (0.11)	1.00 (1.29)	3.05 (1.36)	2.05 (1.19)
3	0.30 (0.11)	1.23 (1.42)	2.88 (1.49)	1.65 (1.35)

than Model 1. This gives $2 \times (1036.15 - 1031.72) = 8.86 \sim \chi^2_2$ (two DF are used because Model 2 uses two additional variance parameters) and shows that Model 2 is a significant improvement ($p = 0.012$). Thus, it is unlikely that the differences in the treatment variances have occurred by chance. The standard errors in Model 2 reflect the different treatment variances. The standard error has become larger for A − C and smaller for B − C.

Including random centre and centre·treatment effects

Different treatment variances can be modelled at the residual level even when random effects are included. We illustrate this by adding random centre and centre·treatment effects (Model 3). The standard errors are increased over those in Models 1 and 2 because treatment effects are assumed to vary randomly across the centres. As in Model 2, the standard errors reflect the different residual variances for treatments.

SAS code and output

Variables

`pat` = patient,
`centre` = centre,
`treat` = treatment (A = 20 mg dose, B = 40 mg dose, C = placebo),
`dbp` = post-treatment DBP (mmHg),
`dbp1` = pre-treatment DBP (mmHg).

Output is only given for Model 2, since it is the main model we are illustrating in this section.

Model 1

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat/ S;
LSMEANS treat/ DIFF PDIFF CL;
```

Model 2

```
PROC MIXED NOCLPRINT; CLASS centre treat pat;
MODEL dbp = dbp1 treat/ S DDFM=KR;
REPEATED /SUBJECT=pat GROUP=treat;
LSMEANS treat/ DIFF PDIFF CL;
```

The `REPEATED` statement causes the covariance matrix to be blocked by patients. Because there is only one observation per patient, there will be no correlation between the observations. The `GROUP` option causes a separate variance parameter to be used for each treatment group. Only the more relevant parts of the output are listed.

Dimensions

Covariance Parameters	3
Columns in X	5
Columns in Z	0
Subjects	288

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Residual	pat	treat A	104.31
Residual	pat	treat B	56.4517
Residual	pat	treat C	76.8822

Fit Statistics

−2 Res Log Likelihood	2063.4
AIC (smaller is better)	2069.4
AICC (smaller is better)	2069.5
BIC (smaller is better)	2080.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
2	8.87	0.0119

In this case, the likelihood ratio test compares this model to Model 1 where the same variance is fitted across all treatment groups. It demonstrates that fitting separate variances for treatment groups leads to a significant improvement.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	281	7.89	0.0053
treat	2	188	2.73	0.0676

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
treat	A	91.5727	1.0215	98.6	89.64	<.0001	0.05
treat	B	90.5767	0.7811	92.4	115.96	<.0001	0.05
treat	C	88.5274	0.9002	93.5	98.34	<.0001	0.05

Least Squares Means

Effect	Treat	Lower	Upper
treat	A	89.5456	93.5997
treat	B	89.0254	92.1280
treat	C	86.7398	90.3149

```
                Differences of Least Squares Means
                                    Standard
Effect treat _treat Estimate   Error   DF  t Value Pr > |t| Alpha
treat    A      B     0.9960   1.2869  181   0.77   0.4400  0.05
treat    A      C     3.0453   1.3611  191   2.24   0.0264  0.05
treat    B      C     2.0494   1.1934  183   1.72   0.0876  0.05

                Differences of Least Squares Means
       Effect     treat     _treat      Lower       Upper
       treat        A          B       −1.5432     3.5351
       treat        A          C        0.3606     5.7301
       treat        B          C       −0.3053     4.4041
```

Model 3

```
PROC MIXED; CLASS centre treat pat;
MODEL dbp = dbp1 treat/ S DDFM=KR;
RANDOM centre centre*treat;
REPEATED /SUBJECT=pat GROUP=treat;
LSMEANS treat/ DIFF PDIFF CL;
```

8.7 Estimating variance components in an animal physiology trial

The purpose of this experiment was to calculate variance components for breathing measurements in rabbits and to use them to design a future clinical trial to compare two treatments. The inspiration times for 100 breaths were measured on four rabbits on each of 4 days. This gave rise to four potential sources of variation rabbits, days, rabbit·day interaction and residual (between breaths). Note that the variability of individual breaths is given by the sum of the variance components, $\sigma_b^2 + \sigma_d^2 + \sigma_{bd}^2 + \sigma_r^2$. A random effects model was fitted to inspiration time (seconds) with each of these effects taken as random. The resulting variance components were as follows:

Source	Variance component
Rabbit (σ_b^2)	0.00231
Day (σ_d^2)	0.00247
Rabbit·day (σ_{bd}^2)	0.00442
Residual (σ_r^2)	0.00318

The positive rabbit component indicates that, not surprisingly, inspiration time varies between rabbits. The positive day and rabbit·day components show there is

variation between days and additional variation within rabbits on different days. Thus, the rabbits' breathing is sensitive to the trial environment on each day. Factors such as time of day, time since last meal, the experimenter, differences in trial set-up, could all have contributed to this variation. However, we should bear in mind that there were only three DF available for estimating the rabbit and day components (σ_b^2 and σ_d^2), and these estimates are therefore only approximate.

8.7.1 Sample size estimation for a future experiment

We seek to design a study to compare two treatments that are expected to influence breathing in rabbits. Two possibilities are a parallel group study, where each rabbit receives only one treatment or a cross-over design where each rabbit will receive each treatment. For ethical reasons, the number of rabbits should be kept as low as possible, as should the number of days for which each rabbit is studied. We set a maximum of 10 days in the study for each rabbit, but as many as 100 breaths could easily be sampled on each day.

Initially, we must consider the mean difference in inspiration time between the two treatments, Δ, which we wish the study to be able to detect. We must also set the significance level α and the power β. Then, for either design, the standard sample size estimation method can be used to estimate the required number of rabbits, number of days and the number of breaths per session. We require

$$\Delta = (Z_{(1-\alpha)/2} + Z_\beta) \times \mathrm{SE}(t_1 - t_2).$$

A range of possible values for the number of rabbits per treatment (n_t), the number of days (n_d) and the number of breaths (n_b) can be calculated from which the preferred design can be chosen. If the relative costs of using rabbits, days and breaths were specified, we could determine the optimal design. In practice, though, it is usually easier to select the design by scrutinising the range of possible designs that satisfy the power requirements. In the context of this design, there is a clear hierarchy in the design priorities. The first priority is to minimise the number of rabbits, then the number of days per rabbit, and then the number of breaths per day.

In the first instance, we consider setting $\Delta = 0.10$, $\alpha = 0.05$ and $\beta = 0.1$.

Between-rabbit design

The variation of the mean treatment difference in inspiration duration is given by

$$\mathrm{var}(t_1 - t_2) = 2[\sigma_b^2/n_t + \sigma_d^2/n_d + \sigma_{bd}^2/(n_t \times n_d) + \sigma_r^2/(n_t \times n_d \times n_b)],$$

and from this we obtain

$$\Delta = (Z_{\alpha/2} + Z_{1-\beta}) \times \{2[\sigma_b^2/n_t + \sigma_d^2/n_d + \sigma_{bd}^2/(n_t \times n_d)/ + \sigma_r^2/(n_t \times n_d \times n_b)]\}^{1/2}.$$

The approach that we recommend to determine the final design is to calculate Δ for a range of values for n_t, n_d and n_b, accepting only those which yield $\Delta < 0.10$.

A SAS program to undertake this calculation is given at the end of this section. The smallest number of rabbits to give a viable design is 14 rabbits per treatment with 9 days of observation and 25 breaths per day. By increasing the number of rabbits to 18 per treatment, the number of days per rabbit can be reduced to eight.

Within-rabbit design

In this design, each treatment is received by each of the n_t rabbits for n_d days. Again, n_b breaths are measured on each day. The variation of the mean treatment difference in inspiration duration is now given by

$$\text{var}(t_1 - t_2) = 2 \times [\sigma_d^2/n_d + \sigma_{bd}^2/(n_t \times n_d) + \sigma_r^2/(n_t \times n_d \times n_b)],$$

$$\Delta = (Z_{\alpha/2} + Z_{1-\beta}) \times \{2 \times [\sigma_d^2/n_d + \sigma_{bd}^2/(n_t \times n_d) + \sigma_r^2/(n_t \times n_d \times n_b)]\}^{1/2}.$$

As before, Δ is calculated for a range of values for n_t, n_d and n_b. We find in this instance that none of the designs satisfies the study requirements. This occurs because the requirement to limit rabbits to 10 days of study means that a maximum of 5 days are allowed per treatment. Thus, however, large n_t is set, $\text{var}(t_1 - t_2) > 2\sigma_d^2/n_d$, and $\Delta > (Z_{\alpha/2} + Z_{1-a}) \times (2\sigma_d^2/n_d)^{1/2} = 0.102$.

Thus, with these design requirements, we could only undertake a between-rabbit design with 14 rabbits per treatment group. This could cause the design requirements to be rethought, and we illustrate the following method with Δ changed to 0.15.

The following table lists, for the between-rabbit design, combinations of the numbers of rabbits, days and breaths that satisfy $\Delta < 0.15$. The table is structured so that for any specified number of rabbits, the number of days is the minimum possible to give $\Delta < 0.15$, and then for that number of rabbits and days, the number of breaths is also the minimum to satisfy $\Delta < 0.15$.

Rabbits per group (n_t)	Days (n_d)	Breaths (n_b)	Δ
4	8	5	0.148 16
6	5	5	0.148 39
8	4	5	0.149 57
10	4	5	0.143 15
12	4	5	0.138 71
14	4	5	0.135 45
16	3	10	0.149 72
18	3	5	0.148 23
20	3	5	0.146 65
–	–	–	–
–	–	–	–
100	3	5	0.134 71

Since our priority in this study is to minimise the number of rabbits, we would prefer the design with four rabbits per group, studied for eight days each, with five breaths measured per day. If our priorities were different, we could alternatively consider six rabbits studied for five days, eight rabbits studied for four days, each with five breaths per day, or even 16 rabbits studied for three days with 10 breaths per day.

With the requirement that $\Delta < 0.15$, it is now also possible to use the within-rabbit design. The following table, constructed along the lines of the previous table, provides alternative designs.

Rabbits (n_t)	**Days (n_d)**	**Breaths (n_b)**	**Δ**
2	5	5	0.144 95
4	4	5	0.140 08
6	4	5	0.131 94
8	3	5	0.147 43
10	3	5	0.144 39
12	3	5	0.142 33
14	3	5	0.140 84
–	–	–	–
–	–	–	–
48	3	5	0.134 33
50	3	5	0.134 22

We could therefore opt for only two rabbits studied for five days per treatment with five breaths per day. Although this does satisfy the sample size requirement as specified, it is intuitively unappealing to conduct such a small study, and the design with four rabbits studied for four days per treatment might be preferred.

Note that in these tables, we have only considered intervals of two in the number of rabbits. We could, of course, fill in the gaps in the regions, which we are considering implementing. We should bear in mind, however, the limited accuracy of the estimates of the variance components on which our calculations are based. Apparently, precise sample size calculations could be misleading. Therefore, we recommend that the sample size calculations be viewed as establishing ballpark figures for the size of the study and helping in determining the most appropriate type of design.

We also recommend that some form of sensitivity analysis is performed before a trial design is finalised. This can have two dimensions. One of these is how the design changes with the choice of Δ, α and β. We have seen previously that a between-rabbit design would be essential with our first choice of figures, while a within-subject design would be preferable when Δ was larger. The second dimension is the sensitivity of the design to the values used for the estimates of variance components. Sometimes, quite small changes can modify the design

appreciably. The wisest design might not be the one that is optimal with the initial choice of parameters but the one that performs well over a wider range of possible parameter values.

SAS code and output

Variables

`rabbit` = rabbit number
`ins` = inspiration time (seconds)
`day` = day number

Analysis model

```
PROC MIXED; CLASS rabbit day;
MODEL ins=;
RANDOM day rabbit day*rabbit;
```

Class Level Information

Class	Levels	Values
rabbit	4	1 2 3 4
day	4	1 2 3 4

Iteration History

Iteration	Evaluations	−2 Res Log Like	Criterion
0	1	−2724.63410044	
1	2	−4378.01162932	0.00014476
2	1	−4378.64036487	0.00002574
3	1	−4378.74485209	0.00000138
4	1	−4378.75001721	0.00000001

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
day	0.002467
rabbit	0.002312
rabbit*day	0.004419
Residual	0.003185

Fit Statistics

−2 Res Log Likelihood	−4378.8
AIC (smaller is better)	−4370.8
AICC (smaller is better)	−4370.7
BIC (smaller is better)	−4373.2

Sample size estimation for between-rabbit trial

```
DATA a; SET a;
DO rabbit = 2 TO 100 BY 2;
DO day = 1 TO 10;
DO breath = 5 TO 100 BY 5;
OUTPUT;
END;
END;
END;

DATA a; SET a;
d = (2*10.51*(0.00231/rabbit + 0.00247/day +
    0.00442/(day*rabbit) +
    0.00318/(breath*day*rabbit)))**0.5;
*IF d<0.10; * for delta<0.10;
IF d<0.15;
PROC SORT; BY rabbit day breath;

DATA a; SET a; BY rabbit;
IF FIRST.rabbit;
PROC PRINT NOOBS; VAR rabbit day breath d;
```

Output appears in the main text.

Sample size estimation for within-rabbit trial

```
DATA a; SET a;
DO rabbit = 2 TO 50 by 2;
DO day = 1 TO 10;
DO breath = 5 TO 100 BY 5;
OUTPUT;
END;
END;
END;

DATA a; SET a;
d = (2*10.51*(0.00442/(day*rabbit)
   + 0.00247/day + 0.00318/(breath*day*rabbit)))**0.5;
IF d<0.15;
PROC SORT; BY rabbit day breath;

DATA a; SET a; BY rabbit;
IF first.rabbit;
PROC PRINT NOOBS; VAR rabbit day breath d;
```

Results appear in the main text.

8.8 Inter- and intra-observer variation in foetal scan measurements

Ultrasound scans are often used during pregnancy to predict gestation (age of foetus). However, predictions made using ultrasound can be unreliable, particularly in the later stages of pregnancy. An experiment to measure inter- and intra-observer variability in ultrasound measurements was carried out at an Edinburgh maternity hospital. Six radiologists participated in the experiment. Fifty-two women in the latter stages of pregnancy with a mean gestation of 29.9 (SD 3.2) weeks were each scanned by two of the radiologists selected at random. Both scans were carried out in the same session. Note that it would not have been ethical or, indeed, feasible to have used all six radiologists at each session.

A random effects model was fitted to the data, with radiologist effects taken as random and subject (women) effects as fixed. In this case, subjects are fitted as fixed because they each have a different gestation and therefore cannot be treated as a randomly distributed sample. The resulting variance components were as follows:

Radiologist	0.000,
Residual	0.287.

The zero radiologist component indicates that there was no systematic variability between the radiologists. This reassured the radiologists that although they often obtained different gestation predictions on the same women, none of them had a tendency to produce particularly high or low readings. The residual variance of 0.287 potentially incorporates several types of variability such as image variation, variation caused by foetus changing position, and radiologist measurement error. The residual value indicates that predicted gestations have a standard deviation of $\sqrt{0.287} = 0.54$ weeks and hence a 95% confidence interval of $\pm t_{47,0.975} \times 0.54 = \pm 2.01 \times 0.54 = \pm 1.09$ weeks (a t_{47} statistic is used because 47 is the residual DF). Note that this is the error encountered in measuring foetus size from which gestation is predicted. It does not incorporate the variability in foetal size that occurs naturally at given gestations. The standard error would be higher if this were taken into account.

SAS code and output

Variables

`obs` = observer ID,
`pat` = patient number,
`gest` = estimated gestation (weeks).

```
PROC MIXED; CLASS pat obs;
MODEL gest=pat/ DDFM=KR;
RANDOM obs;
```

```
                Class Level Information
Class   Levels  Values
pat       52    1 2 3 4 5 6 7 8 9 10 11 12 13
                14 15 16 17 18 19 20 21 22 23
                24 25 26 27 28 29 30 31 32 33
                34 35 36 37 38 39 40 41 42 43
                44 45 46 47 48 49 50 51 52
obs        6    e i j n r s

                  Iteration History
Iteration     Evaluations     -2 Res Log Like      Criterion
0                   1           118.98383527
1                   1           118.98383527       0.00000000

                  Convergence criteria met.
                    Covariance Parameter
                          Estimates
                Cov Parm            Estimate
                Obs                        0
                Residual              0.2886

                       Fit Statistics
                -2 Res Log Likelihood          119.0
                AIC (smaller is better)        121.0
                AICC (smaller is better)       121.1
                BIC (smaller is better)        120.8

                Type 3 Tests of Fixed Effects
                   Num     Den
Effect             DF      DF     F Value      Pr > F
Pat                51      52       71.92      <.0001
```

8.9 Components of variation and mean estimates in a cardiology experiment

In this experiment, the heart wall thickness of 11 healthy dogs was measured using ultrasound scans. Each scan consisted of 20 thickness measurements taken over a single heartbeat cycle. In this case, we consider the maximum thickness (millimetres) obtained over the cycle. This was not a carefully planned experiment, and a varying number of scans and observers were used for each dog. Each dog had between two and six scans, and each scan was assessed by between one and three observers (see Table 8.10). Although the data are not balanced,

Table 8.10 Number of scans and observers per scan for each dog.

Dog	Number of scans	Minimum number of observers	Maximum number of observers	Mean thickness
Aussie	4	1	2	3.52
Corrie	2	3	3	3.78
Dance	4	1	3	4.27
Gem	2	1	3	5.53
Gus	2	1	1	3.87
Isla	2	2	3	6.54
Jenny	2	3	3	4.51
Jos	2	3	3	4.64
Midge	4	1	1	2.94
Mist	3	1	3	3.78
Tara	2	3	3	3.67

a random effects model can still be used to estimate variance components and to calculate an appropriate estimate for the mean and standard error of heart wall thickness. Taking the raw mean thickness and its standard error would be inappropriate in this study, since greater weight would then be given to the dogs and observers who were used most frequently. A random effects model was fitted with dog, observer, dog·observer and scan effects taken as random. The variance component estimates were as follows:

Source	Variance component
Dog	0.294
Obs	0.083
Dog·obs	0.307
Scan	0.705
Residual	0.566

These indicate little systematic variation (bias) between observers. Not surprisingly, there is some additional variation occurring between the heart wall thicknesses of individual dogs. There is also some variation between the observers in their overall readings (i.e. over all scans) for each dog (dog·obs). However, most of the variation occurs at the scan and residual levels. Thus, there is variability between scans even after allowing for between-dog variability, and there is variability in readings made from the same scans by different observers (residual).

The overall mean maximum thickness estimate was 4.42 mm, with a standard error of 0.33. This compares with a raw mean of 4.25 mm, with standard error 0.18 (calculated taking the naive approach that all observations are independent). While the means are of a similar order, the raw mean standard error does not adequately take into account the variation from each source.

SAS code and output

Variables

dog = dog name,
obs = observer initials,
max = maximum heart wall thickness (mm),
scan = scan ID

```
PROC MIXED; CLASS dog obs scan;
MODEL max =/ SOLUTION DDFM=KR;
RANDOM dog obs dog*obs scan;
```

```
                 Class Level Information
        Class   Levels   Values
        dog       11     Aussie Corrie Dance Gem Gus
                         Isla Jenny Jos Midge Mist Tara
        obs        3     AA BB CC
        scan      29     dog 100 dog 101 dog 102 dog
                         105 dog 106 dog 107 dog 108
                         dog 110 dog 111 dog 112 dog
                         113 dog 114 dog 115 dog 67 dog
                         77 dog 78 dog 79 dog 80 dog 83
                         dog 84 dog 85 dog 86 dog 90
                         dog 92 dog 95 dog 96 dog 97
                         dog 98 dog 99
```

(dog in scan values is irrelevant).

```
                        Covariance Parameter
                             Estimates
                       Cov Parm     Estimate
                       Dog           0.2939
                       Obs           0.08281
                       dog*obs       0.3070
                       scan          0.7047
                       Residual      0.5656

                          Fit Statistics
                 -2 Res Log Likelihood     177.8
                 AIC (smaller is better)   187.8
                 AICC (smaller is better)  189.1
                 BIC (smaller is better)   189.8

                    Solution for Fixed Effects
                           Standard
Effect       Estimate       Error      DF    t Value     Pr > |t|
Intercept     4.4157       0.3298      2      13.39       0.0055
```

8.10 Cluster sample surveys

Sometimes, data occur within clusters; for example, patients may be treated at particular hospitals or by particular GPs. In a cluster sample survey, clusters are usually sampled at random. Either all items (e.g. patients) within a cluster are then observed or, alternatively, a random sample of items is taken from each cluster (a two-stage cluster sample). Random variation between clusters can be allowed by fitting cluster effects as random in the analysis. The 'global' results obtained can then be related with some confidence to the population of clusters. Note that inference will be stronger than in multi-centre analyses where centres are rarely sampled randomly. The model can also be used to produce shrunken cluster estimates, which help to prevent unrealistic estimates occurring as a result of chance variation when cluster sizes are small.

8.10.1 Example: cluster sample survey

This was a cluster sample survey undertaken to determine the prevalence of a disease in animals, taken from Thrusfield (1995) (neither disease nor animal species is specified in this reference). There were a total of 865 farms in the population of interest, and 14 were sampled at random. All animals at these farms were assessed for the presence of the disease. The disease frequencies and prevalences are listed as follows:

Farm	Animals	Diseased animals	Prevalence
1	272	17	0.063
2	87	15	0.172
3	322	71	0.220
4	176	17	0.097
5	94	9	0.096
6	387	23	0.059
7	279	78	0.280
8	194	59	0.304
9	65	37	0.569
10	110	34	0.309
11	266	23	0.087
12	397	57	0.144
13	152	19	0.125
14	231	17	0.074
Total	3 032	476	0.157

A random effects model is fitted to the data using pseudo-likelihood, with farm effects taken as random. The data are analysed in binomial form. Since there will be a separate farm effect for each observation, the dispersion parameter is fixed at one to prevent the farm variance component from becoming incorporated into the dispersion parameter. There are no farms with a prevalence of zero. Therefore, there will be no uniform farm effects, and we would not expect any bias in the farm variance component (see Section 3.2.3).

The farm variance component was 0.727 and indicates that, not surprisingly, disease rates vary by more than chance variation between the farms. The random effects model gave the logit (SE) for overall disease as −1.684 (0.236), which leads to an overall prevalence rate of 15.7%, with 95% confidence limits of 10.0–23.6% (calculated by $-1.684 \pm t_{13,0.975} \times 0.236 = -1.684 \pm 2.160 \times 0.236$ and converting to rates by $(1 + \exp(-\text{logit}))^{-1}$).

This compares with the 'local' raw prevalence rate of 15.7(14.4,17.0)%, which has a much narrower confidence interval, since it does not take the between-farm variation into account. (Note also that since there is no estimation of a variance component, the asymptotic normality of the estimate is used to calculate the 95% confidence limits from estimate $\pm 1.96 \times$ SE.) The latter estimate and standard error would only be applicable if the animals were (erroneously) assumed to be a random sample of all animals in the population. In contrast, the 'global' estimate obtained from the random effects model can legitimately be related to the potential population of farms.

SAS code and output

Variables

`farm` = farm ID,
`dis` = number of animals with the disease,
`n` = number of animals at farm.

```
PROC GLIMMIX; CLASS farm;
MODEL dis/n= /SOLUTION LINK=LOGIT DDFM=KR OR CL;
RANDOM farm;
```

```
                Fit Statistics
-2 Res Log Pseudo-Likelihood  36.29
Generalized Chi-Square        13.03
Gener. Chi-Square / DF         1.00

      Covariance Parameter
            Estimates
Cov                  Standard
Parm   Estimate       Error
Farm    0.7267       0.3055
```

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept	−1.6843	0.2356	12.9	−7.15	<.0001	0.05

Solutions for Fixed Effects

Effect	Lower	Upper
Intercept	−2.1937	−1.1750

8.11 Small area mortality estimates

There is a need to assess health needs of particular populations to allocate medical resources effectively. We consider mortality rates in the 40–64 years age group in Edinburgh by post code area (e.g. EH6). These rates provide some measure of general health in the area and could conceivably have some value for health care planning. However, mortality rates are subject to random variation and may be inaccurate, particularly in small areas. This problem can be alleviated by obtaining shrunken mortality estimates using a random effects model, with post code area taken as random. The data are analysed in binomial form, with the dispersion parameter fixed at one. This prevents the area variance component from becoming incorporated into the dispersion parameter. The data could alternatively have been analysed in Bernoulli form (i.e. a separate observation for each person), although this would have led to a very large dataset. If there had been no uniform area categories (i.e. with zero mortality), we would expect the results to have been identical, regardless of the data form. Since one of the areas does have a mortality of zero, the Bernoulli analysis might have been marginally preferable, since the (unfixed) dispersion parameter would then help to overcome any bias caused by random effects shrinkage (see Section 3.2.4). However, we found that the results changed only very slightly when the data were analysed in this form.

The overall mortality rate of 40–64-year-old people in 1991 in Edinburgh was 1.29% (2845/220,178). The variance component obtained for post code area is 0.0576, indicating that more variability occurs between the areas than expected by chance variation. This is not surprising, since the areas are known to differ in terms of socioeconomic factors and age distribution within the 40–64-years range. The shrunken mortality estimates are listed in Table 8.11 in increasing order. Note that they are calculated by $\{1 + \exp[-(\text{intercept} + \text{logit})]\}^{-1}$.

The ranking of the areas is quite different between the raw and shrunken mortality rates. As expected, greatest shrinkage towards the overall mortality rate of 1.29% occurs in the areas with smaller populations. For example, 0% mortality is observed in area EH38 with a population of 93, but the shrunken rate of 1.19% is close to the overall rate of 1.29%. The area with the highest raw mortality rate, EH24, with a relatively small population of 386, has its raw rate of 2.85%

Table 8.11 Mortality and shrunken mortality prediction of 40–64-year-old people in Edinburgh by post code area.

Post code sector	Deaths aged 40–64	Population aged 40–64	Shrunken logit estimate	Raw mortality rate per 100	Shrunken mortality rate per 100	Original rank
EH10	60	8298	−0.442	0.72	0.81	3
EH9	38	4704	−0.316	0.81	0.92	7
EH12	99	11105	−0.300	0.89	0.94	9
EH53	19	2456	−0.271	0.77	0.96	4
EH39	20	2399	−0.233	0.83	1.00	8
EH26	53	5550	−0.214	0.95	1.02	15
EH45	23	2534	−0.195	0.91	1.04	11
EH14	122	11922	−0.186	1.02	1.05	16
EH54	114	10956	−0.170	1.04	1.07	17
EH35	5	693	−0.148	0.72	1.09	2
EH32	52	4941	−0.139	1.05	1.10	19
EH51	47	4459	−0.134	1.05	1.11	20
EH34	5	644	−0.126	0.78	1.11	5
EH29	9	999	−0.124	0.90	1.12	10
EH55	24	2294	−0.112	1.05	1.13	18
EH37	4	512	−0.105	0.78	1.14	6
EH31	7	769	−0.102	0.91	1.14	12
EH46	8	853	−0.101	0.94	1.14	14
EH42	25	2315	−0.094	1.08	1.15	21
EH3	52	4630	−0.089	1.12	1.16	22
EH38	0	93	−0.064	0.00	1.19	1
EH41	40	3416	−0.053	1.17	1.20	23
EH30	32	2678	−0.036	1.19	1.22	24
EH20	25	2087	−0.032	1.20	1.22	25
EH19	45	3704	−0.028	1.21	1.23	26
EH21	77	6258	−0.022	1.23	1.24	28
EH36	1	108	−0.019	0.93	1.24	13
EH49	56	4529	−0.016	1.24	1.24	29
EH18	9	732	−0.009	1.23	1.25	27
EH48	102	8130	−0.006	1.25	1.25	30
EH43	3	237	0.000	1.27	1.26	31
EH40	7	548	0.003	1.28	1.27	33
EH25	14	1099	0.004	1.27	1.27	32
EH44	12	906	0.019	1.32	1.29	35
EH33	42	3233	0.020	1.30	1.29	34
EH4	220	16382	0.058	1.34	1.34	36
EH2	3	134	0.068	2.24	1.35	52
EH15	85	6106	0.082	1.39	1.37	37
EH28	14	918	0.082	1.53	1.37	40

(*continued overleaf*)

Table 8.11 (*continued*)

Post code sector	Deaths aged 40–64	Population aged 40–64	Shrunken logit estimate	Raw mortality rate per 100	Shrunken mortality rate per 100	Original rank
EH22	116	8297	0.089	1.40	1.38	38
EH47	113	7649	0.137	1.48	1.44	39
EH13	65	4251	0.152	1.53	1.47	41
EH5	73	4772	0.156	1.53	1.47	42
EH8	84	5480	0.162	1.53	1.48	43
EH52	78	5044	0.167	1.55	1.49	44
EH17	79	5084	0.171	1.55	1.49	45
EH27	12	569	0.191	2.11	1.52	51
EH6	121	7624	0.202	1.59	1.54	46
EH24	11	386	0.268	2.85	1.64	54
EH23	39	2031	0.287	1.92	1.67	49
EH16	147	7731	0.370	1.90	1.82	47
EH11	165	8604	0.383	1.92	1.84	48
EH7	143	7355	0.389	1.94	1.85	50
EH1	26	970	0.425	2.68	1.92	53

shrunken to 1.64% and no longer ranks the highest. Any planning decisions based on the mortality rates would differ widely depending on which estimates were taken. Use of shrunken estimates is an important aid to overcoming the problem of extreme estimates that can occur when estimates are based on small populations. However, it is important to bear in mind that the estimates will be sensitive to the assumption that the underlying practice rates, that is, the β, are normally distributed on the logit scale (see Section 2.4.6). Unfortunately, at present, it is difficult to check with confidence that this assumption is met.

It is possible to use t tests to determine whether the mortality estimates for each area are significantly different from the average. From the SAS output produced at the end of this section, we find that post code areas EH1, EH11, EH16, EH6 and EH7 have significantly higher mortality rates ($p = 0.01$, 0.0001, 0.0001, 0.04, 0.0001, respectively), and areas EH10, EH12, EH14 and EH9 have significantly lower mortality rates ($p = 0.0003$, 0.003, 0.05, 0.02, respectively). If required, the estimated standard errors produced by the analysis can be used to calculate confidence intervals for the shrunken mortality rates.

`SAS` code and output

Variables

`pc` = post code area,
`death` = number of deaths in post code area,
`pop` = population in post code area.

```
PROC GLIMMIX; CLASS pc;
MODEL death/pop= / LINK=LOGIT SOLUTION DDFM=KENWARDROGER;
RANDOM pc /SOLUTION;
```

Fit Statistics

−2 Res Log Pseudo-Likelihood	36.88
Generalized Chi-Square	52.61
Gener. Chi-Square / DF	0.99

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
Pc	0.05755	0.01709

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	−4.3591	0.04199	44.99	−103.82	<.0001

Solution for Random Effects

Effect	pc	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Pc	EH1	0.4250	0.1724	53	2.47	0.0170
Pc	EH10	−0.4418	0.1153	53	−3.83	0.0003
Pc	EH11	0.3825	0.08551	53	4.47	<.0001
Pc	EH12	−0.3002	0.09883	53	−3.04	0.0037
Pc	EH13	0.1519	0.1189	53	1.28	0.2071

ETC.

8.12 Estimating surgeon performance

The performance of nine surgeons undertaking mastectomy operations at an Edinburgh hospital was recorded in terms of whether any post-operative complications arose (reported by Dixon *et al.*, 1996). However, some of the surgeons only performed a few operations, and therefore their complication rates are likely to be unreliable. More appropriate shrunken estimates can be obtained by fitting a random effects model with surgeons taken as random. The data were analysed in binomial form, with the dispersion parameter fixed at one.

The overall complication rate was 42.6%. The surgeon variance component was 1.296, indicating greater variability between surgeons than expected by chance.

The raw and shrunken complication rates are given as follows. As in the previous example, the shrunken rates are calculated by $\{1 + \exp[-(\text{intercept} + \text{logit})]\}^{-1}$.

Surgeon number	Number of complications	Number of operations	Shrunken logit estimate	Complication rate (%)	Shrunken complication rate (%)	Original rank
1	7	91	−1.957	8	9	1.0
3	3	15	−0.825	20	25	2.0
9	1	3	−0.179	33	38	3.0
5	2	5	−0.057	40	41	4.0
4	7	13	0.375	54	52	5.0
6	3	5	0.438	60	53	6.0
8	2	3	0.479	67	54	7.5
2	10	15	0.818	67	63	7.5
7	2	2	0.908	100	65	9.0

Although the ranking of the surgeons does not change, some of the shrunken rates have changed noticeably from the raw rates. For example, surgeon 7 has a raw complication rate of 100% based on only two operations, but his shrunken rate of 65% is more acceptable. However, even after shrinkage, it is clear that complication rates differ widely between the surgeons. From the SAS output produced at the end of this section, we find that surgeon 1 has a significantly lower complication rate than average ($p = 0.008$) but that no surgeon has a significantly higher rate than average. If required, the estimated standard errors produced by the analysis can be used to calculate confidence intervals for the complication rate of each surgeon. As with the previous example, it is important to bear in mind that the estimates will be sensitive to the strong assumption that the underlying practice rates, that is, the β, are normally distributed on the logit scale (see Section 2.4.6).

SAS code and output

Variables

`surg` = surgeon ID
`outcome` = number of patients with post-operative complications
`n` = number of patients operated on by surgeon.

```
PROC GLIMMIX; CLASS surg;
MODEL outcome/n= /LINK=LOGIT SOLUTION DDFM=KENWARDROGER;
RANDOM surg / SOLUTION;
```

Fit Statistics

−2 Res Log Pseudo-Likelihood	29.82
Generalized Chi-Square	7.18
Gener. Chi-Square / DF	0.90

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
surg	1.2779	0.8543

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	−0.3115	0.4725	8	−0.66	0.5283

Solution for Random Effects

Effect	surg	Estimate	Std Err Pred	DF	t Value	Pr > |t|
surg	1	−1.9572	0.5614	8	−3.49	0.0082
surg	2	0.8180	0.6436	8	1.27	0.2394
surg	3	−0.8247	0.6780	8	−1.22	0.2585
surg	4	0.3750	0.6545	8	0.57	0.5824
surg	5	−0.05697	0.8218	8	−0.07	0.9464
surg	6	0.4377	0.8162	8	0.54	0.6064
surg	7	0.9078	0.9913	8	0.92	0.3866
surg	8	0.4789	0.9121	8	0.53	0.6138
surg	9	−0.1785	0.9214	8	−0.19	0.8512

8.13 Event history analysis

Sometimes, data on the exact times of a particular event (or events) are available on a group of patients. Examples of events would include asthma attacks, epilepsy attacks, myocardial infarctions and hospital admissions. Often, occurrence (and non-occurrence) of an event is available on a regular basis (e.g. daily), and the data can then be thought of as having a repeated measures structure. An objective may be to determine whether any concurrent events or measurements have influenced the occurrence of the event of interest. For example, daily pollen counts may influence the risk of asthma attacks; high blood pressure may precede a myocardial infarction.

8.13.1 Example

We will consider data from a placebo-controlled study of a treatment for eczema, which can cause itchiness. Thirty-four female subjects were asked to complete a diary, recording the severity of their eczema and days of menstrual bleeding. This was done daily during a 4-week run-in period and for 6 months following treatment. We will consider analysing the occurrence of severe itchiness. To assess whether itchiness was related to the menstrual cycle, a covariance pattern model was fitted with average pre-treatment itchiness, treatment (active or placebo) and cycle (menstrual bleeding, y/n) taken as fixed effects and with a compound symmetry covariance pattern to model the correlation between the repeated observations on each subject. Many of the subjects had not recorded their symptoms on every day or had stopped filling in their diaries before the end of the trial. However, such missing data did not pose problems in fitting a covariance pattern model.

This analysis produced a compound symmetry correlation parameter of 0.30, showing a moderate correlation between itchiness records on the same subjects. The treatment effect was clearly non-significant. The cycle effect was significant ($p = 0.02$), and subjects were more likely to experience severe itchiness during menstrual bleeding. The cycle odds ratio and 95% confidence interval based on empirical standard errors were 1.41 (1.05, 1.89). We suggest that the empirical standard errors are taken because more complex covariance patterns have not been explored, and it is possible that the true covariance pattern is not compound symmetry. In this instance, the empirical standard errors are appreciably larger than the model-based standard errors, making this the conservative approach. In the following code, we use one of the options for reducing the standard error bias with the empirical approach. All such methods gave broadly similar results with slightly higher standard errors than the classical method.

`SAS` code and output

Variables

`treat` = treatment (a = active, p = placebo)
`pat` = subject
`cycle` = menstrual bleeding (1 = yes, 2 = no)
`itch` = response variable (0 = no itch, 1 = itch)
`itch1` = proportion of days of severe itchiness pre-treatment
`one` = 1 for all observations

```
PROC GLIMMIX ABSPCONV=0.00001 EMPIRICAL=FIRORES;
CLASS pat treat cycle;
NLOPTIONS MAXITER=50;

MODEL itch/one= itch1 treat cycle / LINK=LOGIT SOLUTION
DDFM=SATTERTH CL OR;
RANDOM INT /SUBJECT=pat TYPE=CS RESIDUAL;
```

Note that the ABSPCONV and MAXITER options have been used, as the procedure did not converge using the default settings. Note also that we could have fitted this model using PROC GENMOD.

Fit Statistics

−2 Res Log Pseudo-Likelihood	21646.08
Generalized Chi-Square	3715.76
Gener. Chi-Square / DF	0.76

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
CS	pat	0.2957	0.07653
Residual		0.7616	0.01547

Solutions for Fixed Effects

Effect	treat	cycle	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept			−1.8636	0.4244	31	−4.39	0.0001	0.05
itch1			2.0262	1.3196	31	1.54	0.1348	0.05
treat	a		0.09362	0.4785	31	0.20	0.8462	0.05
treat	p		0	.4785	.	.38	.8462	.
cycle		1	0.3443	0.1446	33	2.38	0.0232	0.05
cycle		2	0	.4785	.	.38	.8462	.

Solutions for Fixed Effects

Effect	treat	cycle	Lower	Upper
Intercept			−2.7291	−0.9982
itch1			−0.6652	4.7176
treat	a		−0.8824	1.0696
treat	p		.	.
cycle		1	0.05016	0.6385
cycle		2	.	.

Odds Ratio Estimates

Effect	treat	cycle	_treat	_cycle	Estimate	DF	95% Confidence Limits	
itch1					7.59	31	0.514	111.90
treat	a		p		1.10	31	0.414	2.91
cycle		1		2	1.41	33	1.051	1.89

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
itch1	1	31	2.36	0.1348
treat	1	31	0.04	0.8462
cycle	1	33	5.67	0.0232

8.14 A laboratory study using a within-subject 4 × 4 factorial design

This example is part of a more extensive experiment investigating inflammatory processes in cells, including the effects of nicotine and smoking habit, from three categories of subjects: those with ulcerative colitis, Crohn's disease, or controls (Aldhous *et al.*, 2008). For each subject, cells are treated with one of four doses of nicotine (0, 1, 10 or 100 μg/ml) in combination with one of four stimuli (medium alone, lipopolysaccharide (LPS, a component of cell walls of bacteria that will stimulate cells, especially monocytes) at 1 μg/ml and phytohaemmaglutinin (PHA, a mitogenic agent that stimulates all T cells in a manner which gives a maximal response) at 0.5 μg/ml and 5 μg/ml; the two concentrations were used to assess sensitivity of T cells to a sub-optimal concentration, leading to 16 treatment combinations within each subject. There were 15 subjects with ulcerative colitis, 18 with Crohn's disease and 12 controls, of whom 1, 6 and 3, respectively, were current cigarette smokers. In this analysis, the response variable is the percentage of cells in the stage of apoptosis (cell death), 3 days after stimulation. Missing values are fairly extensive, with all 16 observations being present for only 10 of the subjects (206 of the possible 720 observations were missing). The reasons for missing values were associated with the practicalities of conducting the experiment and could be considered as missing at random.

The first stage of analysis of this dataset is to consider transformations of the outcome variable. The logit transformation was found to have good variance stabilising properties both for apoptosis and for other outcome variables not considered in this section. All subsequent results are therefore based on $\log(x/(100-x))$ where x denotes the percentage of cells showing apoptosis. There were no values of 0 or 100 leading to infinite values after transformation within the dataset.

There are four fixed effects that need to be considered in our model, together with their interactions:

Disease category	(ulcerative colitis, Crohn's disease, control)
Current smoker?	(yes, no)
Nicotine dose	(0, 1, 10, 100 μg/ml)
Stimulant	nicotine alone, LPS, PHA at 5 or 0.5 μg/ml.

We also need to consider the variance–covariance structure across the 16 observations on each subject. The simplest approach is to fit subjects as a random effect, producing equal correlations between all pairs of treatments. This could be seen as a 'natural' model to use as results arise from 16 physically distinct sets of cells. Further thought, however, leads us to consider slightly more complicated relationships. It is plausible that the four observations obtained from the same stimulant would be more highly correlated with each other than with other observations. This is achieved with the use of subject·stimulant as an additional random effect. By a similar argument, observations at the same dose of nicotine may be more highly correlated than those at different doses of nicotine. Thus subject·nicotine dose can also be fitted as a random effect.

In principle, the proposed correlation structure induced by fitting subject, subject·stimulant and subject·nicotine as random could be compared with more complicated covariance patterns. There are, however, a limited number of subjects and a high proportion of missing values, with a danger of overfitting. More complicated structures will therefore not be considered in this example.

As there are biologically plausible reasons for high-level interactions between the four fixed effects we have specified, our fitting strategy has been to start with a model including all possible main effects and interactions, then dropping the highest level interaction terms if these are non-significant. The random effects have consistently given positive estimates for all three terms, with a substantial subject·stimulant effect but only a tiny subject·nicotine dose effect. This has been retained, nevertheless, because of its plausibility.

The results of our final model are shown in Table 8.12. The variable representing current smoking status, and all interactions involving smoking, all showed no indication of any effect on the proportion of cells in apoptosis, but there was a significant three-way interaction between the disease category, the stimulant and the nicotine dose. For comparison, we show the corresponding model fitting only a simple correlation structure (subject·stimulant and subject·nicotine random effects omitted). It is apparent that the more appropriate covariance structure has yielded a more informative analysis.

`SAS` code and output

Variables

`smoker` = current smoker (yes, no),
`disease` = disease (CD = Crohn's Disease, UC = Ulcerative Colitis, HC = Healthy Control),
`stim` = stimulant used,
`dose` = nicotine dose (0, 1, 10, 100),
`patient` = patient identifier,
`result` = logit (proportion of cells in apoptosis).

Table 8.12 Tests of fixed effects in models with different random effects specified.

	4 Full model				Simple model			
Fixed effects	**Numerator DF**	**Denominator DF**	***F***	**Prob.**	**Numerator DF**	**Denominator DF**	***F***	**Prob.**
Stimulant	3	90.6	12.02	<0.0001	3	426	47.48	<0.0001
Nicotine dose	3	102	0.15	0.931	3	424	0.07	0.971
Disease	2	42.3	1.92	0.161	2	42.4	1.84	0.171
Stim·dose	9	253	1.43	0.181	9	423	0.48	0.891
Stim·disease	6	90.6	1.02	0.411	6	426	3.81	0.001
Dose·disease	6	102	1.85	0.101	6	424	0.71	0.641
Stim·dose·disease	18	253	2.08	0.007	18	423	0.93	0.551
Variance components								
Subject		0.365				0.424		
Subject·stim		0.185				–		
Subject·dose		0.005				–		
Residual		0.066				0.207		

The first PROC MIXED fits a model with all possible interactions among the fixed effects and the more complicated random effects modelling.

```
PROC MIXED; CLASS smoker disease stim dose patient;
MODEL result= smoker | disease | stim | dose/DDFM=KR
RANDOM patient patient*stim patient*dose;
```

The following SAS code fits the simple model for which the results are summarised in Table. 8.12:

```
PROC MIXED; CLASS smoker disease stim dose patient;
MODEL result=disease | stim | dose/DDFM=KR;
RANDOM patient;
```

The following LSMEANS statement allows the three-way interaction to be explored in more detail:

```
LSMEANS disease*stim*dose;
```

The details of the outputs do not introduce any new features and are not included in this section.

8.15 Bioequivalence studies with replicate cross-over designs

Bioequivalence studies are used to demonstrate whether two formulations of a drug will produce the same bioavailability. Typically, studies have a cross-over design, and healthy volunteers will take a single dose of a drug at each visit, following which their blood and/or urine drug concentrations are measured repeatedly. Bioavailability is then usually assessed using the summary statistics such as the area under the curve and maximum concentration. When a study has a two-way cross-over design, the methods described in Section 7.3 and Section 8.1 are appropriate. This design can be used for the determination of average bioequivalence. More recently, bioequivalence studies using replicate cross-over designs, where treatments are received more than once by each patient, have become popular. These can be used to assess the within-subject variability on each treatment, the variability of the within-subject treatment difference and the between-subject variability of each treatment. These parameters are used in the criteria to establish 'individual bioequivalence' and 'population bioequivalence'. It is beyond the scope of this book to examine these topics in depth, and a detailed description of the issues is available on the FDA Website (http://www.fda.gov/downloads/Drugs/Guidances/ucm070244.pdf). In this section, we will look at some of the possibilities for modelling data from such designs. (the FDA guidance is more prescriptive).

A common design has four periods, and patients are randomised to receive treatments in one of the sequences: ABAB, ABBA, BABA, BAAB. Usually, summary statistics calculated at each visit are analysed, and this is the approach we will consider. However, if the treatment interaction with time point (within visit) was of interest or if there were missing data within visits, it may alternatively be appropriate to consider analysing the individual data using the covariance structures suggested in Section 8.1. Treatment and period effects should be fitted as fixed. Sequence and sequence-by-period interaction can also be fitted as fixed; however, we do not see an advantage to it. The sequence effect will only reflect the between-subject group variation that we know is the direct consequence of the randomisation in the trial (unless there is carry-over). The sequence-by-period interaction will estimate within-subject variation, unless the model is mis-specified.

There are various structures that can be used for the covariance matrix. Some of the more plausible options are described as follows starting with the simplest structures. As with other designs where there are many possibilities for modelling covariance structure, the structure can be chosen by performing likelihood ratio tests to determine whether more complex features can be justified statistically. Alternatively, statistics such as Akaike's information criterion can be used, although this is not our preferred option (see Section 6.2.2). The models will all have a block diagonal form for the overall variance matrix, $\mathbf{V}$, with zero correlations between observations on different patients. For example, for a trial with nine patients, the $\mathbf{V}$ matrix can written as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_5} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_6} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_7} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V_9} \end{pmatrix}.$$

The $\mathbf{V_i}$ are blocks of covariances for observations on the ith patient. We will illustrate structures assuming a four-period trial; however, the structures can be extended for trials with more periods. For ease of interpretation, structures are given as if a patient had received sequence AABB. The terms can be appropriately reordered to give matrices for the actual sequences used: ABAB, ABBA, BABA, BAAB or BBAA.

Model 1. Constant covariances for observations on the same patient A very simple structure for $\mathbf{V_i}$ would assume a constant correlation between all

observations on the same patient regardless of the treatment or period,

$$\mathbf{V}_i = \begin{pmatrix} \sigma^2 & \theta & \theta & \theta \\ \theta & \sigma^2 & \theta & \theta \\ \theta & \theta & \sigma^2 & \theta \\ \theta & \theta & \theta & \sigma^2 \end{pmatrix},$$

where

θ = covariance between observations on same patient,
σ^2 = variance.

Model 2. Different variances for treatments It is possible that observations on the two drug formulations have different residual variances. This can be allowed for by modelling separate residuals for observations on each treatment,

$$\mathbf{V}_i = \begin{pmatrix} \sigma^2_{\mathrm{A}} & \theta & \theta & \theta \\ \theta & \sigma^2_{\mathrm{A}} & \theta & \theta \\ \theta & \theta & \sigma^2_{\mathrm{B}} & \theta \\ \theta & \theta & \theta & \sigma^2_{\mathrm{B}} \end{pmatrix},$$

where

θ = covariance between observations on the same patient,
σ^2_{A} = variance for treatment A,
σ^2_{B} = variance for treatment B.

Model 3. Extra covariance for observations on the same treatment The covariance pattern used for Model 2 takes no account of the possibility that observations taken on the same treatment may be more highly correlated than those taken on different treatments. This can be allowed for using the following structure, which builds on Model 1,

$$\mathbf{V}_i = \begin{pmatrix} \sigma^2 & \theta_T & \theta & \theta \\ \theta_T & \sigma^2 & \theta & \theta \\ \theta & \theta & \sigma^2 & \theta_T \\ \theta & \theta & \theta_T & \sigma^2 \end{pmatrix},$$

where

θ = covariance between observations on different treatments,
θ_T = covariance between observations on the same treatments,
σ^2 = variance.

Model 4. Difference covariances for treatments It is possible that observations on the two treatments have different variances and covariances.

This can be allowed for by including two extra parameters.

$$\mathbf{V}_i = \begin{pmatrix} \sigma_A^2 & \theta_A & \theta & \theta \\ \theta_A & \sigma_A^2 & \theta & \theta \\ \theta & \theta & \sigma_B^2 & \theta_B \\ \theta & \theta & \theta_B & \sigma_B^2 \end{pmatrix},$$

where

θ = covariance between observations on different treatments,
θ_A = covariance between observations on treatments A,
θ_B = covariance between observations on treatments B,
σ_A^2 = variance for treatment A,
σ_B^2 = variance for treatment B.

This model can, if required, be optionally constrained so that θ is less than or equal to $\sqrt{(\theta_A \theta_B)}$. Alternatively, the model can be further constrained so that the covariance between observations on the same patient and treatment is never less than that for observations on the same patient but different treatments (i.e. $\theta \leq \theta_A$ and $\theta \leq \theta_B$). As we will see later, this model facilitates the use of Bayesian methods in `PROC MIXED` for obtaining confidence bounds.

Model 5. Modelling covariance across time An alternative to Models 3 and 4, where the treatment groups have different variances and covariances, is to model covariance based on the order of the repeated measurements across time. For example, a general structure with a different covariance parameter for each period pair can be written as

$$\mathbf{V}_i = \begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{12} & \sigma_2^2 & \theta_{23} & \theta_{24} \\ \theta_{13} & \theta_{23} & \sigma_3^2 & \theta_{34} \\ \theta_{14} & \theta_{24} & \theta_{34} & \sigma_4^2 \end{pmatrix}.$$

where

θ_{ij} = covariance between observations at periods i and j,
σ_i^2 = variance for ith period.

Alternatively any of the other covariance structures described in Section 6.2 can be used.

Other models In addition, models can be constructed to model both the correlation structure across the repeated measurements (as in Model 5) and also

take into account whether observations are in the same treatment group. One such model is considered in the following example (Model 6).

Assessing average, population and individual bioequivalence

The current FDA recommendations describe three criteria for demonstrating bioequivalence: average, population and individual bioequivalence. Any of the models described can be used to demonstrate average bioequivalence; however, parameter estimates from Model 4 are required to obtain population and individual bioequivalence.

Average bioequivalence is always required and is demonstrated when the 90% confidence interval of the ratio of the treatment mean to reference mean falls between 0.80 and 1.25. This is equivalently demonstrated when the 90% confidence interval of the difference of the treatment mean to reference mean on the log scale falls between $\pm\log(1.25)$.

Population bioequivalence compares the expected squared difference between observations on different patients and different treatments, to the expected squared difference between observations on different patients who are both on the reference treatment. The criterion is defined by the difference in these expected squared differences divided by the variance for the reference treatment and can be written as

$$\frac{E(Y_{i,\mathrm{A}} - Y_{j,\mathrm{B}})^2 - E(Y_{i,\mathrm{B}} - Y_{j,\mathrm{B}})^2}{\sigma_{\mathrm{B}}^2}$$

with treatment B in this case assumed to be the reference treatment and $Y_{i,\mathrm{T}}$ representing an observation on patient i receiving treatment T. The usual specification of the criterion in terms of the covariance parameters for Model 4 can be obtained from the above expression as,

$$\frac{(\mu_{\mathrm{A}} - \mu_{\mathrm{B}})^2 + (\sigma_{\mathrm{A}}^2 - \sigma_{\mathrm{B}}^2)}{\sigma_{\mathrm{B}}^2}$$

To achieve population bioequivalence, the upper 95% confidence bound for the criterion based on a one-sided confidence interval should be no more than a maximum limit, θ_p. There is no generally recommended limit for θ_p, and the FDA guidance suggests that they (the FDA) can be contacted for more information on setting it. In addition, an upper limit, $\sigma^2_{\mathrm{B,Max}}$, for is set for the denominator variance. This is substituted as the denominator term in the criterion if $\sigma_{\mathrm{B}}^2 > \sigma^2_{\mathrm{B,Max}}$, and the criterion is then described as 'reference scaled'.

Individual bioequivalence compares the expected squared difference between observations on the same patient but different treatments, to the expected squared

difference between repeated observations on the same patient on the reference treatment. The criterion is defined by the difference in these expected squared differences divided by the residual variance for the reference treatment and can be written as

$$\frac{E\left(Y_{i,\mathrm{A}} - Y_{i,\mathrm{B}}\right)^2 - E\left(Y_{i,\mathrm{B}} - Y_{i,\mathrm{B}'}\right)^2}{\sigma^2_{\mathrm{R,B}}}$$

where $Y_{i,\mathrm{B}'}$ represents a repeated observation on patient i and treatment B. The criterion in terms of the covariance parameters in Model 4 can be obtained from this expression as,

$$\frac{\left(\mu_{\mathrm{A}} - \mu_{\mathrm{B}}\right)^2 + \sigma^2_{\mathrm{D}} + \left(\sigma^2_{\mathrm{R,A}} - \sigma^2_{\mathrm{R,B}}\right)}{\sigma^2_{\mathrm{R,B}}}$$

σ^2_{D} relates to the subject-by-treatment variance components and, using the parameterisation of Model 4 given earlier, can be calculated as $\theta_{\mathrm{A}} + \theta_{\mathrm{B}} - 2\theta$. $\sigma^2_{\mathrm{R,A}}$ and $\sigma^2_{\mathrm{R,B}}$ are the residual variances for treatments A and B and are estimated directly from the model. Note that the variance terms in Model 4 can be written as $\sigma^2_{\mathrm{A}} = \sigma^2_{\mathrm{R,A}} + \theta_{\mathrm{A}}$ and $\sigma^2_{\mathrm{R,B}} = \sigma^2_{\mathrm{R,B}} + \theta_{\mathrm{B}}$. To achieve individual bioequivalence, the upper 95% confidence bound of the criterion should be no more than a maximum limit, θ_I (θ_I is often set at 2.495). In addition, an upper limit, $\sigma^2_{\mathrm{R,B,Max}}$, for is set for the denominator variance. This is substituted as the denominator term in the criterion if $\sigma^2_{\mathrm{R,B}} > \sigma^2_{\mathrm{R,B,Max}}$ and, as with population bioequivalence, the criterion is then described as 'reference scaled'.

8.15.1 Example

This was a four-period trial to compare two formulations of an anti-anxiety agent where patients were randomised to receive treatments in the sequences ABAB, ABBA, BABA, BAAB. It is taken from the FDA website (http://www.fda.gov/drugs/scienceresearch/ucm301277.htm). In this case, analyses of the log area under the curve (AUC) will be carried out. Each of the covariance structures detailed previously is fitted, and treatment and period effects are fitted as fixed. The resulting covariance matrices along with −2REML, Akaike's information criterion (AIC) and the Bayesian criterion (BIC) are shown in Table 8.13. A covariance structure is selected by using likelihood ratio tests to statistically justify the inclusion of variance and covariance terms in the model.

Model 2 fits different residual variances for treatments. It is significantly better than Model 1 ($\chi^2 = 7.9, p < 0.05$) and also has lower information criteria. Model 3 reduces to Model 1 as zero additional covariance is estimated for observations on the same treatment. (Note that although SAS reduces the number of parameters in Model 3 from 3 to 2 our view is that the parameter should still be counted. Thus, the values of AIC and BIC shown in Table 8.13 are different to those produced by

Table 8.13 Comparison of models for an equivalence trial of anti-anxiety agents.

Model	Covariance	Number of parameters	−2REML	Akaike's information criterion	Bayes information criterion
1	$\begin{pmatrix} .59 & .45 & .45 & .45 \\ .45 & .59 & .45 & .45 \\ .45 & .45 & .59 & .45 \\ .45 & .45 & .45 & .59 \end{pmatrix}$	2	248.3	252.3	255.6
2	$\begin{pmatrix} .56 & .49 & .49 & .49 \\ .49 & .56 & .49 & .49 \\ .49 & .49 & .68 & .49 \\ .49 & .49 & .49 & .68 \end{pmatrix}$	3	240.4	246.4	251.5
3	$\begin{pmatrix} .59 & .45 & .45 & .45 \\ .45 & .59 & .45 & .45 \\ .45 & .45 & .59 & .45 \\ .45 & .45 & .45 & .59 \end{pmatrix}$	3	248.3	252.3	255.6
4 (unconstrained)	$\begin{pmatrix} .61 & .54 & .45 & .45 \\ .54 & .61 & .45 & .45 \\ .45 & .45 & .57 & .36 \\ .45 & .45 & .36 & .57 \end{pmatrix}$	5	236.2	246.2	254.7
4 (constrained such that $\theta \leq \sqrt{(\theta_A \theta_B)}$)	$\begin{pmatrix} .60 & .54 & .45 & .45 \\ .54 & .60 & .45 & .45 \\ .45 & .45 & .57 & .38 \\ .45 & .45 & .38 & .57 \end{pmatrix}$	5	236.7	244.7	251.5
4 (totally constrained so that $\theta \leq \theta_A$ and $\theta \leq \theta_B$)	$\begin{pmatrix} .65 & .49 & .45 & .45 \\ .49 & .65 & .45 & .45 \\ .46 & .46 & .56 & .49 \\ .46 & .46 & .49 & .56 \end{pmatrix}$	5	239.8	247.8	254.5
5 AR(1)	$\begin{pmatrix} .59 & .48 & .40 & .33 \\ .48 & .59 & .48 & .40 \\ .40 & .48 & .59 & .48 \\ .33 & .40 & .48 & .59 \end{pmatrix}$	2	238.6	242.6	256.6
5 Toeplitz (TOEP)	$\begin{pmatrix} .59 & .48 & .42 & .38 \\ .48 & .59 & .48 & .42 \\ .42 & .48 & .59 & .48 \\ .38 & .42 & .48 & .59 \end{pmatrix}$	4	237.2	245.2	252.0
5 Heterogeneous compound symmetry (CSH)	$\begin{pmatrix} .61 & .45 & .45 & .47 \\ .45 & .56 & .44 & .46 \\ .45 & .44 & .56 & .45 \\ .47 & .46 & .45 & .61 \end{pmatrix}$	5	248.0	258.0	266.4
5 General (UN)	$\begin{pmatrix} .58 & .44 & .44 & .38 \\ .44 & .56 & .54 & .43 \\ .44 & .54 & .56 & .49 \\ .38 & .43 & .49 & .61 \end{pmatrix}$	10	228.9	248.9	265.8
6 AR(1) with different residual variances for treatments	$\begin{pmatrix} .63 & .48 & .44 & .40 \\ .48 & .63 & .48 & .44 \\ .44 & .48 & .56 & .48 \\ .40 & .44 & .48 & .56 \end{pmatrix}$	4	235.4	243.4	235.4 (`PROC MIXED` appears to give wrong value here)

PROC MIXED.) Slightly different covariance parameters are obtained for Model 4 depending on whether and how θ is constrained. As with Model 2, we count θ as a parameter when calculating AIC and BIC. None of the constrained or unconstrained versions of Model 4 leads to a significant improvement over Model 2. Model 5 examines covariance matrices that are structured by time. The first order autoregressive structure cannot be compared to Model 2 using a likelihood ratio test, as the two models are not nested. However, its information criteria are lower, demonstrating that structuring the covariance matrix by periods has led to a better model than structuring it by treatment. None of the other structures for Model 5 shows any improvement over the autoregressive structure. Thus, in this example, more has been gained from fitting an autoregressive structure across all the repeated measurements than from using different covariance terms for treatments. To use the features of both Model 2 and Model 5, Model 6 is fitted with an autoregressive structure for the repeated measurements and also fitting different residual variances for each treatment. However, this model does not lead to an improvement over Model 5.

The FDA currently requires the 90% confidence interval of the ratio of the mean of the new formulation to the reference formulation to be within 0.80 and 1.25 to demonstrate average bioequivalence. The ratio and confidence interval were obtained by taking the exponential of the difference in treatment means (on the log scale) and its confidence interval based on standard errors obtained using the Kenward–Roger adjustment. To provide consistency, we consider results from Model 4 (totally constrained), as this model will be used to assess population and individual bioequivalence. The ratio and confidence interval resulting from this model were 0.98 (0.88, 1.09), and thus average bioequivalence is achieved. In addition, each of the other models considered led to confidence intervals that demonstrated bioequivalence.

The population bioequivalence criterion was calculated from maximum likelihood estimates obtained from the REML analysis of the totally constrained version of Model 4 (see SAS output) as,

$$\begin{aligned}\frac{(\mu_A-\mu_B)^2+(\sigma_A^2-\sigma_B^2)}{\sigma_B^2} &= \frac{0.0208^2+(0.4605+0.0272+0.0710)-(0.4605+0+0.1851)}{(0.4605+0+0.1851)}\\ &= -0.134\end{aligned}$$

The upper 95% confidence bound for the criterion was taken as the 95% centile point in the Bayesian analysis of Model 4 (totally constrained). A value of 0.036 was obtained, and individual bioequivalence can then be assessed by comparing this value to a maximum limit, θ_p. Recommendations for setting θ_p are given in the FDA guidance. The Bayesian model can also be used to provide an alternative point estimate for the criterion. The median value was −0.110 and differs from

the value calculated using maximum likelihood estimates. However, a difference is expected between the two approaches because the distribution of the criterion is likely to be skewed and its median unlikely to coincide with the value calculated from maximum likelihood parameter estimates. Also, the Bayesian analysis used the full joint posterior distribution of the parameters to obtain the distribution of the criterion, whereas REML does not take this into account. However, in the context of assessing bioequivalence, the point estimate is of minor interest.

The individual bioequivalence criterion was calculated from the maximum likelihood estimates obtained in the REML analysis of Model 4 (totally constrained) as,

$$\frac{(\mu_A - \mu_B)^2 + \sigma_D^2 + (\sigma_{R,A}^2 - \sigma_{R,B}^2)}{\sigma_{R,B}^2}$$

$$= \frac{0.0208^2 + (0.0272 + 0) + (0.0710 - 0.1851)}{0.1851} = -0.467$$

The upper 95% confidence bound for the criterion was taken as the 95% centile point in the Bayesian analysis of Model 4 (totally constrained), and a value of 0.504 was obtained. This is lower than the generally recommended limit of 2.495 for θ_p and thus indicates individual bioequivalence has been demonstrated. However, if the value of $\sigma_{R,B}^2$ (0.1851) exceeded the value set for $\sigma_{R,B,Max}^2$, it would be necessary to assess individual bioequivalence based on the reference scaled criterion where $\sigma_{R,B,Max}^2$ was substituted for $\sigma_{R,B}^2$ in the denominator of the criterion. The Bayesian model can also be used to provide an alternative location estimate for the criterion. The median value was 0.177 and, as with the individual criterion, differs from the value calculated using maximum likelihood estimates.

SAS code

The SAS code to fit each model is shown as follows. For some models, alternative code is also given, which sometimes leads to different parameterisations of the same model.

Variables

`lauc` = log(AUC)
`treat` = drug formulation
`per` = period
`pat` = patient

The following statements are used in all the models.

```
PROC MIXED; CLASS pat treat per;
MODEL lauc = treat per/ DDFM=KR;
LSMEANS treat/ DIFF PDIFF CL;
```

Sequence and sequence. period interaction effects can be additionally fitted as fixed in the MODEL statement if required. The following RANDOM and REPEATED statements are then included for each model.

Model 1

```
RANDOM pat;
```

Model 2

```
RANDOM pat;
REPEATED /GROUP=treat TYPE=SIMPLE R;
```

Or equivalently using the following statement that allows the variance matrix to be viewed more easily via the V option.

```
RANDOM int treat/ SUB=pat TYPE=SIMPLE V;
```

Model 3

```
RANDOM pat pat*treat;
```

Or equivalently using,

```
RANDOM int treat/ SUB=pat TYPE=SIMPLE V;
```

Model 4 (unconstrained)

```
RANDOM treat/ SUB=pat TYPE=UN V;
REPEATED /GROUP=treat TYPE=SIMPLE;
```

Model 4 (constrained so that $\theta \leq \sqrt{\theta_A \theta_B}$)

```
RANDOM treat/ SUB=pat TYPE=CSH V;
REPEATED /GROUP=treat TYPE=SIMPLE;
```

Or equivalently substituting the random statement,

```
RANDOM treat/ SUB=pat TYPE=FA0(2) V;
```

Model 4 (totally constrained so that $\theta \leq \theta_A$ and $\theta \leq \theta_B$)

```
RANDOM pat;
RANDOM pat*treat/ GROUP=treat;
REPEATED /GROUP=treat TYPE=SIMPLE;
```

Model 5

The covariance structure across periods can be modelled using one of the following REPEATED statements,

```
REPEATED per/ SUBJECT=pat TYPE=AR(1) R; * autoregressive;
REPEATED per/ SUBJECT=pat TYPE=TOEP R; * Toeplitz;
REPEATED per/ SUBJECT=pat TYPE=CSH R; * heterogeneous
   compound symmetry;
REPEATED per/ SUBJECT=pat TYPE=UN R; * general;
```

To allow extra covariance for observations on the same treatment, the following RANDOM statement should be included,

```
RANDOM pat*treat;
```

Model 6

To fit different variances for treatments include,

```
RANDOM int/ sub=obs group=treat type=simple v;
```

Note that 'obs' is a variable denoting the observation number (obs=_N_;).

SAS output

SAS output is listed from three of the models fitted (2, 4 and 5) in order to demonstrate some of the different parameterisations used by PROC MIXED.

Model 2

Estimated R Matrix for Index 1

Row	Col1
1	0.07529

Estimated V Matrix for pat 1

Row	Col1	Col2	Col3	Col4
1	0.5642	0.4889	0.4889	0.4889
2	0.4889	0.5642	0.4889	0.4889
3	0.4889	0.4889	0.6847	0.4889
4	0.4889	0.4889	0.4889	0.6847

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Intercept	pat		0.4889
Residual		treat 1	0.07529
Residual		treat 2	0.1958

Fit Statistics

−2 Res Log Likelihood	240.4
AIC (smaller is better)	246.4
AICC (smaller is better)	246.6
BIC (smaller is better)	251.5

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	103	0.12	0.7264
per	3	90.6	1.93	0.1297

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
treat	1	2.2724	0.1162	39.4	19.56	<.0001	0.05	2.0375	2.5074
treat	2	2.2931	0.1227	46.4	18.69	<.0001	0.05	2.0463	2.5400

Differences of Least Squares Means

Effect	treat	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
treat	1	2	−0.02073	0.05907	103	−0.35	0.7264	0.05	−0.1379	0.09641

Model 4 (totally constrained) Only output relating to the covariance structure is shown for this model. Note that output listing the individual $\mathbf{V}_i$ matrices is not readily available in SAS for this model.

Covariance Parameter Estimates

Cov Parm	Group	Estimate
pat		0.4605
pat*treattreat	1	0.02723
pat*treattreat	2	0
Residual	treat 1	0.07104
Residual	treat 2	0.1851

Fit Statistics

−2 Res Log Likelihood	239.8
AIC (smaller is better)	247.8
AICC (smaller is better)	248.1
BIC (smaller is better)	254.5

Model 5 Only output relating to the covariance structure is shown.

Estimated R Matrix for pat 1

Row	Col1	Col2	Col3	Col4
1	0.5867	0.4831	0.3978	0.3275
2	0.4831	0.5867	0.4831	0.3978
3	0.3978	0.4831	0.5867	0.4831
4	0.3275	0.3978	0.4831	0.5867

```
 Covariance Parameter Estimates
Cov Parm   Subject     Estimate
AR(1)        Pat        0.8234
Residual                0.5867

2 Fit Statistics
−2 Res Log Likelihood      238.6
AIC (smaller is better)    242.6
AICC (smaller is better)   242.7
BIC (smaller is better)    245.9
```

SAS code to assess population and individual bioequivalence criteria

These criteria are calculated by fitting the totally constrained version of Model 4 using the Bayesian approach to allow probability points to be obtained. In contrast to Bayesian analyses carried out in earlier sections using PROC MCMC, this analysis is carried out using the PRIOR in PROC MIXED (see Section 9.2). Although in general we prefer the use of PROC MCMC for Bayesian analysis, in this case, the MIXED procedure is used because it allows different residual variances for the treatment groups to be specified.

```
DATA b; SET a;
obs=_n_;
PROC MIXED CONVH=0.000001; CLASS pat treat per obs;
MODEL lauc = treat per/ DDFM=KR S;
RANDOM pat;
RANDOM pat*treat/ GROUP=treat V;
RANDOM obs/ GROUP=treat;
PRIOR/ NSAMPLE=11000 OUT=bioeq;
LSMEANS treat/ DIFF PDIFF;
```

The last RANDOM statement has replaced the REPEATED statement used in the previous version of this model. This is because a Bayesian analysis is not available in SAS when a REPEATED statement is used. The model results in identical estimates to the earlier model fitted using a REPEATED statement but it is parameterised slightly differently. An overall residual term is included and needs to be added to each of the treatment residuals.

The unconstrained version of Model 4 can alternatively be obtained by adding a PARMS statement with the NOBOUND option to the code above,

```
PARMS / NOBOUND;
```

SAS output

Only output relating to the covariance parameters is shown.

```
 Covariance Parameter Estimates
Cov Parm     Group      Estimate
pat                      0.4605
pat*treat   treat 1     0.02724
pat*treat   treat 2           0
obs         treat 1    3.58E-19
obs         treat 2      0.1141
Residual                0.07105
```

SAS code to analyse the posterior densities

The PRIOR statement in the SAS code causes a large sample of the model parameters to be output to a new dataset 'bioeq'. This dataset contains seven fixed effects parameters, beta1-beta7, representing the intercept, two treatment effects and the four period effects. Note that the last treatment effect and period effect parameters (beta3 and beta7) are redundant, and all their samples take a value of zero. The six covariance parameters, covp1-covp6, represent the patient variance component, patient. treatment variance components for the two treatments, residual variance components for each treatment and an overall residual variance as shown previously in the 'Covariance Parameter Estimates' output. Note that the residual term needs to be added to each of the treatment residuals to obtain the treatment residual variances shown in the earlier output for Model 4 where a REPEATED statement was used. The parameters required for calculating population and individual bioequivalence criteria are then obtained from the sampled model parameters in SAS as,

$$\begin{aligned}
\mu_{\mathrm{A}} - \mu_{\mathrm{B}} &= \mathrm{beta2} \\
\theta &= \mathrm{covp1} \\
\theta_{\mathrm{A}} &= \mathrm{covp2} + \theta \\
\theta_{\mathrm{B}} &= \mathrm{covp3} + \theta \\
\sigma^2_{\mathrm{R,A}} &= \mathrm{covp4} + \mathrm{covp6} \\
\sigma^2_{\mathrm{R,B}} &= \mathrm{covp5} + \mathrm{covp6} \\
\sigma^2_{\mathrm{A}} &= \mathrm{covp1} + \mathrm{covp2} + \sigma^2_{\mathrm{R,A}} \\
\sigma^2_{\mathrm{B}} &= \mathrm{covp1} + \mathrm{covp3} + \sigma^2_{\mathrm{R,B}} \\
\sigma^2_{\mathrm{D}} &= \theta_A + \theta_{\mathrm{B}} - 2\theta = \mathrm{covp2} + \mathrm{covp3}
\end{aligned}$$

These parameters and values of the individual and population and individual bioequivalence criteria are calculated for each of 10,000 samples using the following SAS code. The first 1000 samples are assumed to be 'burn in' samples, which may be unreliable and are therefore deleted.

```
DATA a; SET data.bioeq;
IF _n_<1000 THEN DELETE;
trt_diff=beta2;
resida = covp4+covp6;
residb = covp5+covp6;
resid_diff=resida-residb;
sig2d = covp2+covp3;
sig2a = covp1+covp2+resida;
sig2b = covp1+covp3+residb;
tot_diff=sig2a-sig2b;
trt2=trt_diff**2;
popn_crit = ( trt_diff**2 + tot_diff ) / sig2b;
indiv_crit = ( trt_diff**2 + sig2d + resid_diff ) / residb;
PROC UNIVARIATE FREQ ROUND=0.001 DATA=a;
VAR trt_diff popn_crit indiv_crit;
```

The mean, median and upper 95% centile are obtained from the PROC UNIVARIATE output. The 90% probability interval for the treatment difference is given by the 5% and 95% centile points.

8.16 Cluster randomised trials

In the clinical trials we have considered so far, subjects have been randomised directly to treatments or health care interventions. However, in some situations, subjects are grouped within 'clusters', and it may not be practical or ethical for subjects to receive different treatments within the same cluster (e.g. centre, hospital, clinic or general practice). For example, in a trial of breast screening carried out in Edinburgh (Alexander *et al.*, 1999), it was not considered ethical to offer screening to some women within a general practice and not to others. In this situation, clusters rather than subjects can be randomised to treatment groups to form a so-called 'cluster randomised' design. An appropriate analysis for this design can be achieved using a mixed model fitting treatment effects as fixed and cluster effects as random. Compared with a multi-centre, individually randomised design this leads to some loss of efficiency as treatment effects are compared between clusters rather than within centres, and hence their standard errors are larger whenever the cluster variance component is positive. The variance of the difference between treatments is given by

$$\text{var}(t_i - t_j) = \sigma^2(1/n_i + 1/n_j) + \sigma_c^2/(1/c_i + 1/c_j)$$

where σ^2 is the residual variance, n_i and n_j are the numbers of subjects receiving treatments i and j, σ_c^2 is the cluster variance component, and c_i and c_j are the number of clusters allocated to groups i and j. In a between-subject analysis taking no account of cluster effects, the variance would have been

$\text{var}(t_i - t_j) = \sigma^2(1/n_i + 1/n_j)$. Note that unlike the multi-centre trial where there is a choice of whether to fit centre effects either as fixed or random giving results with alternative interpretations, a cluster randomised design should always be analysed comparing treatments at the cluster level of variation by fitting cluster effects as random.

8.16.1 Example: A trial to evaluate integrated care pathways for treatment of children with asthma in hospital

Within a children's hospital, there is a wish to evaluate the use of 'integrated care pathways' for children attending the Accident and Emergency Department with attacks of asthma. This more structured form of care is being compared with usual hospital practice. It is not administratively possible to simultaneously use one system for one child and the alternative for another, and so the trial was set up as a cluster randomised trial based on one system being chosen randomly to be used for each week. All children seen during the same week form a cluster.

Analysis of a cluster randomised trial using a mixed model is extremely simple, as we have seen previously. The feature that distinguishes it from a more conventional parallel group trial is that individuals within a cluster may exhibit intra-cluster correlation, as a consequence of individuals within the cluster producing results that are more similar than those from subjects from different clusters. We can achieve the desired analysis in `SAS` by simply having a variable *cluster* to identify cluster membership, and then adding *cluster* as a random effect to the fixed effects model that would otherwise be used.

This particular example has been chosen, though, because it demonstrates an additional benefit of the mixed models approach. Some children have repeated attacks of asthma and will therefore attend the Accident and Emergency Department on multiple occasions. In a conventional parallel group trial, such children would not be considered for the trial again after they had been randomised once. In this setting, we can utilise every visit made by a child. Our analysis has to recognise that the results from the same child cannot be regarded as independent. If we assume a constant correlation between all observations on a child, then we can handle the analysis by simply fitting the patient identifier as a random effect as well as cluster.

The primary outcome variable for this trial is the interval from arrival to discharge for patients admitted to the ward. We do not discuss the results in detail, (see Cunningham *et al.* (2008)) but the code and key parts of the output are summarised as follows.

Also, analysis is presented for all patients attending the hospital, in order to maximise data for analysis. In this example, we see from the output presented as follows that the variance component for the subject is estimated as zero, with a moderate variance component for the clusters. In this analysis, there is no evidence for a treatment effect, but the confidence interval for the treatment difference is still relatively wide.

SAS code and output

Variables

`weekno` = week number that defines clusters,
`subject` = subject identifier,
`treat` = treatment,
`logmins` = log of duration of stay in minutes.

```
PROC MIXED NOCLPRINT;
CLASS treat subject weekno;
MODEL logmins= treat / DDFM=KR;
RANDOM subject weekno;
LSMEANS treat / DIFF PDIFF CL;
```

Covariance Parameter Estimates

Cov Parm	Estimate
subject	0
WEEKNO	0.09455
Residual	2.1708

Fit Statistics

−2 Res Log Likelihood	543.4
AIC (smaller is better)	547.4
AICC (smaller is better)	547.5
BIC (smaller is better)	553.4

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Treat	1	5.89	0.81	0.4034

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Treat	CONTROL	6.8962	0.2468	6.14	27.94	<.0001	0.05
Treat	ICP	6.6019	0.2145	5.57	30.78	<.0001	0.05

Least Squares Means

Effect	treat	Lower	Upper
treat	CONTROL	6.2958	7.4966
treat	ICP	6.0672	7.1366

```
                Differences of Least Squares Means
                                    Standard
Effect  treat  _treat Estimate   Error    DF  t Value Pr > |t| Alpha
treat  CONTROL  ICP    0.2943   0.3270  5.89   0.90   0.4034  0.05

                Differences of Least Squares Means
             Effect    treat    _treat    Lower    Upper
             treat     CONTROL    ICP    −0.5094   1.0980
```

8.16.2 Example: Edinburgh randomised trial of breast screening

This study involved 54 654 women aged between 45 and 64 years from 87 general practices in Edinburgh. Women were recruited between 1978 and 1981 at a time when there was no breast screening programme in Scotland. Practices were randomly assigned either to an intervention or to a control group. Women in the intervention group were invited to participate in a screening programme involving an annual screen for breast cancer, while those in the control group received normal medical care. All subjects in the intervention practices were considered part of the intervention group, whether they attended screening or not. The primary endpoint was death because of breast cancer during the following 14 years. However, as the practice variance component was zero for this endpoint, death from all causes will be considered, since a positive practice variance component and hence a more interesting analysis is obtained. Age and deprivation category were available for individual women and are considered as covariates in the analyses. Table 8.14 shows the models fitted and their results. The first set of models (1a–4a) is unadjusted for age and deprivation, and the second set of models adjust for these factors. Analyses of the data are considered with and without fitting practice effects as random and with the data in both Bernoulli and binomial form. Individual subject values for age and deprivation are fitted when data are in Bernoulli form (Models 3a, 3b). However, this is not possible when the data are in binomial form (Models 4a, 4b) and instead average age and deprivation for each practice are fitted.

All the unadjusted analyses showed a significant reduction in mortality in the intervention group. The dispersion parameter is greater than one in Model 2a, and the practice variance component is positive in Models 3a and 4a, indicating that more than chance variation has occurred between the practices. The odds ratio in Model 1a erroneously treats all subjects as independent, ignoring the fact that they come from different practices. This has led to a smaller confidence interval than for the other models. Results from Models 2a and 4a are similar. Thus, in this case, it has made little difference whether practice variation is modelled using a random effect or a dispersion parameter.

Age and deprivation were highly significant in all the adjusted analyses (Models 1b and 4b), and the odds ratios were higher than in the unadjusted models. The

Table 8.14 Results of analyses of Edinburgh breast screening trial.

Model	Data	Fixed effects	Random effects	Practice variance component	OR (95% CI)
1a	Bernoulli	Group	–	–	0.79 (0.75, 0.83)
2a	Binomial	Group	ϕ[a]	($\phi = 1.30$)	0.83 (0.74, 0.94)
3a	Bernoulli	Group	Practice	0.056 ($\phi = 0.99$)	0.82 (0.73,0.92)
4a	Binomial	Group	Practice	0.054 ($\phi = 1.35$)	0.84 (0.75, 0.95)
1b	Bernoulli	Group, age, deprivation	–	–	0.86 (0.82, 0.91)
2b	Binomial	Group, age, deprivation	ϕ[a]	($\phi = 1.48$)	0.95 (0.87, 1.03)
3b	Bernoulli	Group, age, deprivation	Practice	0.034 ($\phi = 0.98$)	0.87 (0.79,0.96)
4b	Binomial	Group, age, deprivation	Practice	0.020 ($\phi = 0.77$)	0.93 (0.86, 1.01)

[a] ϕ indicates the dispersion parameter is fitted.

odds ratios were notably higher (and non-significant) in Models 2b and 4b, where average age and deprivation covariates were used to analyse the data in binomial form, than in Models 1b and 3b where data were analysed in Bernoulli form. Thus, in this case, it has been important to adjust for age and deprivation at the subject level. Modelling the practice effect as random in Models 3b and 4b is preferable to Model 2b where a dispersion parameter is used. This is because differing sizes of the practices are taken into account when determining the variance component. The results from the Bernoulli model fitting practice as random (Model 3b) showed a significant difference in all-cause mortality between the treatment groups. This was not accounted for by the breast screening intervention, and this could indicate that the randomisation may not have been strictly adhered to or this may be one of the occasions where chance has resulted in a false positive.

`SAS` code and output

Datasets

`bern` = dataset in Bernoulli form,
`bin` = dataset in binomial form (frequencies by general practice).

Variables

`death` = dead or alive after 14 years' follow-up,
`group` = intervention or control group,
`dep` = deprivation category,
`gp` = general practice,
`age` = age (divided by 10 for easier fitting),
`one` = 1,
`n` = numbers of subjects.

SAS code is given for the analyses adjusted for age and deprivation (Models 1b–4b). Models 1a–4a are obtained by using the same code, with ‘age’ and ‘dep’ omitted in the MODEL statements. Relevant output is given for Models 1b and 3b with Models 2b and 4b having a similar form.

Model 1b

```
PROC GENMOD DATA=bern; CLASS group dep;
MODEL death/one=group dep age / DIST=B WALD TYPE3;
ESTIMATE 'group' group 1 -1/ EXP ALPHA=0.05;
```

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence	Limits	Chi-Square	Pr > ChiSq
Intercept		1	−1.7696	0.0177	−1.8042	−1.7349	10032.8	<.0001
GROUP	1	1	−0.2398	0.0255	−0.2898	−0.1898	88.35	<.0001
GROUP	2	0	0.0000	0.0000	0.0000	0.0000	.35	.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Wald Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
GROUP	1	88.35	<.0001

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence	Limits	Chi-Square	Pr > ChiSq
group	−0.2398	0.0255	0.05	−0.2898	−0.1898	88.35	<.0001
Exp(group)	0.7868	0.0201	0.05	0.7484	0.8271		

Model 2b

```
PROC GENMOD DATA=bin; CLASS group dep;
MODEL death/n=group dep age / DIST=B WALD TYPE3 DSCALE;
ESTIMATE 'group' group 1 -1/ EXP ALPHA=0.05;
```

Model 3b

```
PROC GLIMMIX DATA=bern; CLASS gp group dep;
MODEL death=group age dep/ DIST=B S DDFM=KR;
RANDOM _RESIDUAL_;
RANDOM gp;
INITITER=20;
NLOPTIONS MAXITER=100;
ESTIMATE 'group' group 1 -1/ CL OR;
```

Fit Statistics	
−2 Res Log Pseudo-Likelihood	283832.8
Generalized Chi-Square	53007.66
Gener. Chi-Square / DF	0.98

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
GP	0.03404	0.008147
Residual (VC)	0.9803	0.005966

Solutions for Fixed Effects

Effect	GROUP	DEP	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			−8.3411	0.1414	10704	−58.99	<.0001
GROUP	1		−0.1427	0.04953	79.97	−2.88	0.0051
GROUP	2		0	.	.	.88	.0051
AGE			0.1254	0.002240	53971	56.01	<.0001
DEP		1	−0.6231	0.07654	13289	−8.14	<.0001
DEP		2	−0.4736	0.07079	17540	−6.69	<.0001
DEP		3	−0.3987	0.07100	20476	−5.62	<.0001
DEP		4	−0.3049	0.06510	18629	−4.68	<.0001
DEP		5	−0.2061	0.07398	19757	−2.79	0.0054
DEP		6	−0.09823	0.07552	32847	−1.30	0.1934
DEP		7	0	.	.	.88	.0051

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
GROUP	1	79.97	8.30	0.0051
AGE	1	53971	3136.84	<.0001
DEP	6	22689	18.85	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
group	−0.1427	0.04953	79.97	−2.88	0.0051	0.05	−0.2413	−0.04412

Estimates

Label	Odds Ratio	Lower Odds Ratio	Upper Odds Ratio
group	0.867	0.786	0.957

Model 4b

```
PROC GLIMMIX DATA=bin; CLASS gp group dep;
MODEL death/n=group age dep/ DIST=B S DDFM=KR;
RANDOM _RESIDUAL_;
RANDOM gp;
ESTIMATE 'group' group 1 -1/ CL OR;
```

8.17 Analysis of bilateral data

The near symmetry of the human body, and indeed of other animals, ensures that measurements in human and animal research are commonly bilateral. Opthalmology, rhinology and dentistry are candidate disciplines for this type of data. Most examples, though, arguably occur in orthopaedics. Observations on gait, muscle strength and joint mobility provide examples where measurements may be made, but there are also subjective measures such as pain where observations may be bilateral.

We will present two examples. The data for the first of these examples come from a large database of children and young adults with gait problems. It is used purely to illustrate the differences in conclusions that can arise from alternative methods of analysis when we wish to compare measurements in two groups of subjects. Subjects are grouped according to whether they had received previous femoral derotation osteotomy (FDO) (an operation to re-position the ball of the femur in the hip socket) and whether there had been surgery to remove implants used in earlier operations to correct gait. The outcome variables considered are mean stance hip rotation (the amount the hip rotates when the foot is in contact with the ground during walking) and hip external rotation (a clinical measurement when the patient is static).

Within the medical literature, it has been common practice to treat observations on each limb, for example, as though they are independent and to apply standard statistical methods for analysis. In most instances, this assumption of independence will be unsound, as it will be unusual for an observation on the right and left side of an individual to be uncorrelated. The consequence will usually be to obtain standard errors that are smaller than is justified and to obtain p-values that are too small (i.e. biased towards statistical significance). A valid but potentially inefficient approach to analysis in the situation we are considering is to revert to the individual as the subject of analysis, either by randomly deleting one observation in each pair or, more commonly, by using the mean of observations on the left and right. Thus standard errors may be larger than necessary with higher p-values. A mixed models approach, in which subjects are fitted as random, takes into account any correlation and provides a fully efficient analysis. We will explore the practical effect of undertaking such analyses in the following simple example and, at the end of this section, we show the simple `SAS` code to conduct such an analysis.

Example 1

Our dataset contains results from 979 hips from 573 subjects (167 unilateral and 406 bilateral). In Table 8.15, we illustrate the results obtained with all 3 methods, including the invalid 'individual value' approach, when applied to the two outcome variables when dichotomised by prior FDO (yes or no) and by instrumentation removal (yes or no). We see a general pattern that the results from the mixed model are intermediate between those from analysis of individual

Table 8.15 Illustration of results from three methods of analysis of bilateral data.

		No FDO $N = 814$	FDO $N = 165$	p-value	No removal $N = 876$	Removal $N = 103$	p-value
Mean stance	I	12.6(0.5)	9.1(1.1)	0.003	12.2(0.5)	10.5(1.3)	0.24
Hip rotation	A	12.9(0.5)	10.2(1.3)	0.04	12.6(0.5)	11.3(1.6)	0.43
	M	12.7(0.5)	9.7(1.2)	0.02	12.3(0.5)	10.8(1.5)	0.34
Hip External	I	29.2(0.5)	33.3(1.2)	0.0007	29.5(0.5)	32.8(1.5)	0.03
Rotation	A	29.0(0.6)	33.6(1.4)	0.0011	29.5(0.6)	32.6(1.7)	0.07
	M	29.2(0.6)	33.2(1.3)	0.0044	29.5(0.6)	32.7(1.6)	0.06

I=analysis of individual values
A=analysis of averages
M=mixed model analysis

data (invalid) and analysis of averages per patient. In relation to the precision of the estimates (the standard errors), this is always true. Although this is also the usual situation with the p-values, note that it does not always happen. The analysis of hip external rotation comparing those with and without prior FDO yields the highest p-value with the mixed models analysis. This is due to the particular pattern of missing values, which gives a smaller estimate of the between-group difference with the mixed models analysis compared to using the averages, although the standard errors are intermediate as we would expect.

The same principle holds with regression situations, whether this is a univariate or multivariable association that is being investigated using bilateral data. Simply amalgamating all data points will violate the assumption of independence of the observations and produce an invalid analysis. Using averages or deleting observations are the only valid approaches using basic statistical methods, but the most efficient analysis will use mixed models. If we regress mean stance hip rotation on hip external rotation, the regression coefficients (standard errors) are −0.408 (0.027) erroneously using data points as independent, −0.374 (0.036) using averages and −0.428 (0.028) with mixed models.

Mixed models also allow the investigation of far more complicated questions that might not be possible at all with an approach of averaging results from the left and right sides. When there are multiple covariates recorded on each side (such as previous operations, previous fractures, dominant side, range of joint movements and muscle strengths), plus systemic covariates (such as age and gender) that might affect the outcome variable of interest, mixed models can allow for these effects while simultaneously allowing for the correlation between the bilateral observations.

Example 2

Our second example of bilateral data comes from a database used to study patella tracking after surgery to implant an artificial knee joint. Patella tracking is the

examination of the path of the patella or kneecap bone as the leg is bent. It may be assessed by radiographic examination of the knee from the lateral view at a sequence of knee flexion angles. From each radiograph, the position of the patella is assessed by a measure of distance, obtained between fixed landmarks on the femur and tibia, and an associated angle. In this study, these measurements were made at full extension of the leg, with the leg bent to an angle of 30° and at full flexion (maximum angle). In knee replacement surgery, the kneecap may be retained or it may be resurfaced on the articular side with an artificial surface. In this example, we will be looking at the effect of patella resurfacing on tracking of the patella in the sagittal plane. Observations were made pre-operatively, at 6 months post-operatively and at 12 months. The original dataset contained a mixture of patients who were studied prospectively from the preoperative phase onwards and others who were examined in a less systematic way for clinical reasons. For the latter group of patients, any missing observations are unlikely to be missing at random. We have therefore taken the approach of including only those patients for whom a full set of preoperative patella tracking observations was taken, as this indicates a prior intention to obtain full tracking data. The assumption of missing at random is more reasonable for any subsequent missing values, which may be due to technical problems such as obtaining measurements from the radiograph. There are 31 such patients, 24 of whom had unilateral operations and 7 of whom had bilateral operations.

In principle, both the angular and distance data could be analysed simultaneously, but this will be far more complicated to analyse and interpret, and so each will be analysed separately. We will concentrate principally on the angular data, as this shows the more interesting results. We can expect correlation between bilateral observations on the same subject, and so the patient identifier (id) will be fitted as a random effect. In addition, observations at the two postoperative time points, within a patient, are likely to be correlated, and so id·time will also be fitted as a random effect. The data at the three knee angles (flex) within a patient could be fitted as a repeated term, with a choice of covariate pattern but will be analysed in this case by fitting id·flex as a third random effect term, generating the compound symmetry structure. The relationship between results at each of the time points and at each flexion may be related to the corresponding preoperative measurements (anglebase), but the coefficients for these covariate effects may not necessarily be identical. As well as a covariate term in the model we will also initially fit an anglebase·time and an anglebase·flex interaction. The initial model will also contain fixed effects for time, flex, time·flex, resurface (the variable indicating whether resurfacing was carried out), resurface·time, resurface·flex and resurface·flex·time.

The initial model confirmed the correlation across bilateral observations, across time and across the three angles. The anglebase·time interaction ($p = 0.84$) and anglebase·flex interaction ($p = 0.36$) indicated a similar baseline effect throughout the postoperative observations, and the three-way interaction was also non-significant ($p = 0.43$). Successive fitting of reduced models also led to the

removal from the model of resurface·time (p = 0.23) and time·flex (p = 0.21). In the final model (see output at the end of this section), time is also non-significant with an estimate that at 12 months, the angles are, on average 0.5° higher than at 6 months (SE 0.9, p = 0.63). The baseline is having only a modest effect on postoperative readings (coefficient 0.16, SE 0.07, p = 0.02). The most noticeable effect in the model is the highly significant resurface·flex interaction (p < 0.0001), which is summarised as follows.

	LSMEANS for angle (°)(SE)	
Flexion	**Not Resurfaced**	**Resurfaced**
Full extension	19.4 (1.3)	14.8 (1.4)
30°	22.4 (1.4)	21.1 (1.5)
Full flexion	23.5 (1.4)	29.3 (1.5)

Preoperatively, the angles did not differ significantly with flexion ($p = 0.08$) with overall means and standard errors of 20.2° (1.4), 22.3° (1.3) and 18.2° (1.3). Thus, the pattern seen is that, after operation, if the patella has been retained intact, there is no important change in the angles from preoperative levels, while after resurfacing, the angle is relatively decreased at full extension and increased at full flexion. We can examine the statistical significance of these changes by modifying our analysis to consider change in angle as our outcome variable instead of actual angle. As the baseline angle is included in both models as a covariate, the analysis is just a reparameterisation of the original model. This shows the mean adjusted changes (SE) (p), as −5.3° (1.4) ($p = 0.0003$) at full extension after resurfacing and 9.2° (1.5) ($p < 0.0001$) at full flexion. In the group without resurfacing, the mean adjusted changes are also statistically significant at full flexion (3.5 (1.4), $p = 0.02$), though we retain the conclusion that, overall in this group, there are no important changes in the angles from preoperative levels.

The analysis of the distance measure (dist) showed no statistically significant effect for any of the terms involving resurfacing (minimum $p = 0.07$ for distbase·flex interaction) and the final model (see code at the end of this section) included only effects for time ($p = 0.13$), flex ($p < 0.001$) and time by flex ($p = 0.01$). The corresponding least squares means are shown as follows.

	LSMEANS for distance (mm)(SE)	
Flexion	**6 months postoperatively**	**12 months postoperatively**
Full extension	64.1 (2.1)	64.0 (2.1)
30°	19.2 (2.4)	26.8 (2.2)
Full flexion	48.0 (2.4)	47.7 (2.1)

At 6 months, the least squares means are very similar to the corresponding preoperative means (SE) of 62.1 (1.0), 19.3 (1.6) and 47.4 (2.0). At 12 months, there has been a moderate increase in the distance at 30° with no change at the other flexions.

Checks were made on outliers and influence, but neither outcome variable produced any cause for concern.

Sample SAS code for first example (output not shown)

```
PROC MIXED;
CLASS subject priorfdo;
MODEL meanstancehiprotation=priorfdo / DDFM=KR;
RANDOM subject;
LSMEANS priorfdo / DIFF PDIFF;
```

SAS code and output for second example

Variables

`id` = patient identifier,
`resurf` = was the patella resurfaced?,
`time` = time of measurement (2=6 months, 3=12 months),
`flex` = angle of flexion (1=full extension, 2=30°, 3=full flexion),
`angle` = measurement of angle,
`anglebase` = corresponding preoperative measurement of angle,
`dist` = measurement of distance,
`distbase` = corresponding preoperative measurement of distance.

Final model for angles and output

```
PROC MIXED; CLASS time flex id resurf;
MODEL angle= time flex resurf resurf*flex anglebase/
  DDFM=KR SOLUTION;
RANDOM id id*flex id*time;
LSMEANS resurf*flex;
```

Model Information

Data Set	WORK.C
Dependent Variable	angle
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Kenward-Roger
Degrees of Freedom Method	Kenward-Roger

Class Level Information

Class	Levels	Values
time	2	2 3
flex	3	1 2 3
id	31	2 4 5 12 13 16 17 23 26 32 39 43 44 46 49 50 51 56 60 61 62 66 68 69 74 75 78 80 83 85 86
resurf	2	No Yes

Dimensions

Covariance Parameters	4
Columns in X	15
Columns in Z	186
Subjects	1
Max Obs Per Subject	228

Number of Observations

Number of Observations Read	228
Number of Observations Used	185
Number of Observations Not Used	43

Iteration History

Iteration	Evaluations	−2 Res Log Like	Criterion
0	1	1168.99570666	
1	2	1126.22961000	0.00069923
2	1	1125.92178949	0.00002074
3	1	1125.91329805	0.00000003
4	1	1125.91328730	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
id	7.1001
flex*id	9.8319
time*id	6.2011
Residual	14.7438

Fit Statistics

−2 Res Log Likelihood	1125.9
AIC (smaller is better)	1133.9
AICC (smaller is better)	1134.1
BIC (smaller is better)	1139.6

Solution for Fixed Effects

Effect	resurf	time	flex	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept				26.2697	1.9616	95.4	13.39	<.0001
time		2		−0.4610	0.9489	28.7	−0.49	0.6308
time		3		0	.	.	.	.
flex			1	−14.4650	1.6206	50.2	−8.93	<.0001
flex			2	−8.1253	1.7043	52	−4.77	<.0001
flex			3	0	.	.	.	.
resurf	No			−5.7440	2.0115	67	−2.86	0.0057
resurf	Yes			0	.	.	.	.
flex*resurf	No		1	10.3026	2.2089	45.9	4.66	<.0001
flex*resurf	Yes		1	0	.	.	.	.
flex*resurf	No		2	7.0202	2.3030	48.2	3.05	0.0037
flex*resurf	Yes		2	0	.	.	.	.
flex*resurf	No		3	0	.	.	.	.
flex*resurf	Yes		3	0	.	.	.	.
anglebase				0.1610	0.07067	122	2.28	0.0244

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time	1	28.7	0.24	0.6308
flex	2	49.2	34.89	<.0001
resurf	1	28.9	0.00	0.9842
flex*resurf	2	47.3	11.23	0.0001
anglebase	1	122	5.19	0.0244

Least Squares Means

Effect	resurf	_flex	Estimate	Standard Error	DF	t Value	Pr > |t|
flex*resurf	No	1	19.3654	1.3165	59.1	14.71	<.0001
flex*resurf	Yes	1	14.8068	1.3783	62.7	10.74	<.0001
flex*resurf	No	2	22.4227	1.4197	69.3	15.79	<.0001
flex*resurf	Yes	2	21.1465	1.5219	74	13.90	<.0001
flex*resurf	No	3	23.5279	1.3830	62.8	17.01	<.0001
flex*resurf	Yes	3	29.2718	1.4817	72	19.76	<.0001

Code for final model for distance data (output not shown)

```
PROC MIXED;  CLASS time flex id resurf;
MODEL dist=time flex time*flex / DDFM=KR SOLUTION;
RANDOM id id*flex id*time
```

8.18 Incomplete block designs

8.18.1 Introduction

Incomplete block designs often provide an efficient design where it is not practical for subjects to receive all available treatments. For example, in a trial of a topical skin treatments, subjects may receive two of four possible treatments on each hand, or in a crossover trial, subjects may only receive some of the available treatments (see Section 7.5). The design is often suitable in situations where it is practical to observe more than one treatment per subject, but the number of assessments available per subject is limited. It is almost always more efficient than assessing each treatment pair in separate trials, although, of course, there may be other considerations that would make separate trials preferable: for example, the possibility of carryover effects or the risk of subjects dropping out in a crossover study.

A mixed model is usually the most efficient way to analyse data from incomplete block studies because information is combined from the within- and between-subject error strata. However, the mixed model variance for treatment comparisons is frequently not taken into account in study design. This is likely to be because formulae for these variances involve matrix multiplication and have not been readily available in a simpler algebraic form. However, the most appropriate sample size estimate will be based on the mixed model variance, and knowledge of this variance allows the efficiency of alternative incomplete block designs to be compared for a particular set of study objectives and practical constraints. For example, there may be a choice in the number of treatments to be studied or in the number of treatments that may be received per subject. Alternatively, a design providing a lower variance for certain treatment comparisons may be desirable when comparisons to a benchmark treatment are of greater interest or the difference between a particular treatment pair is expected to be smaller than for other pairs. In these situations, the more usual balanced incomplete block (BIB) design may be adapted to achieve the desired variances for the different treatment pairs.

As with the majority of clinical trials, the primary aim is usually to detect a significant difference between treatment pairs, and their expected variance is usually used to estimate sample size. Algebraic expressions for this variance may be obtained for some designs. However, algebraic expressions for more complex designs can become more cumbersome and may be difficult to obtain. It is then usually easier to calculate the variance using matrix multiplication within suitable software. This may be achieved using either software that carries out matrix multiplication (e.g. Matlab), custom software for trial design, or mixed models software (e.g. PROC MIXED), with a suitably constructed dataset to reflect a specific design. The latter will sometimes be more accessible to the practising scientist or statistician involved in designing the study who may not have ready access to specialist software for matrix multiplication or trial design.

A range of designs will be considered in Sections 8.18.2–8.18.5 and, where feasible, formulae for the mixed model variances of the treatment differences will be provided. The variances are computed based on the generic formula for a difference in fixed effects is $\mathbf{c}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{c}$, where $\mathbf{c}$ is the contrast vector defining the difference (see Section 2.2.2). For example, in a trial comparing three treatments A, B and C, a pairwise comparison of treatments A and B may be given by $\boldsymbol{c} = (0\ 1\ -1\ 0)'$ (see Section 2.4.4). The $\mathbf{V}$ matrix has a diagonal form as described in Section 2.1 and is specified in terms of the residual variance and the ratio of the subject variance component to the residual. The $\mathbf{X}$ matrix encapsulates the study design, and the rows specify each combination of treatments received by subjects in the study. In order to obtain algebraic formulae for the variances of treatment differences, the $\mathbf{X}$ and $\mathbf{V}$ matrices were assembled in Matlab and evaluated algebraically using the Matlab Symbolic Math toolbox (a software toolbox for the simplification and manipulation of symbolic expressions, http://www.mathworks.co.uk/products/symbolic/). The resulting algebraic formulae obtained were simplified into ratios of polynomials using further Matlab computational algebra and manual manipulation.

In Section 8.18.2, the BIB design is considered where the aim is to compare all treatment pairs with equal importance, but there may be scope to vary either the number of treatments per subject or the number of treatments included in the study. Section 8.18.3 covers designs for the situation where comparisons to a particular treatment are of primary interest. Section 8.18.4 looks at designs for the situation where it is desirable for comparisons between a particular treatment pair to have a lower variance than for other treatment pairs. This may be either because the treatment pair is of primary interest or because a smaller difference is expected between the treatments and thus a smaller variance required to obtain statistical significance. Section 8.18.5 examines designs for the situation where it is desirable for comparisons between two particular treatment pairs to have lower variance than for other treatment pairs. Although the most obvious use for the formulae will be for sample size estimation, we will also consider examples of their use as an aid to study design choice in the second part of each section.

8.18.2 Balanced incomplete block (BIB) designs

In a BIB design, each treatment is received an equal number of times, although only a subset of the treatments will be received by each subject. The variance of the mean difference between two treatments, where subjects receive T of N possible treatments and where there is an equal allocation of subjects to all possible treatment combinations, may be written

$$\text{Var(treatment difference)} = \frac{2(T\gamma + 1)\sigma_r^2}{R(Nn_1\gamma + n_1 + n_2)}$$

Where

σ_r^2 = residual (within subject) variance
σ_s^2 = subject variance component
$\gamma = \sigma_s^2/\sigma_r^2$
T = number of treatments received by each subject
N = total number of treatments in experiment
R = number of replicates of the set of all possible combinations of T treatments (e.g. for $T = 2$ and $N = 4$, R replicates of AB, AC, AD, BC, BD, and CD will lead to a total of $6R$ subjects)

$n_1 = \binom{N-2}{T-2}$ = number of instances of the treatment pair occurring on the same subject (e.g. for $T = 3$ and $N = 4$ there are 2 instances of treatment pair AB: ABC and ABD), note $n_1 = 1$ when $T = 2$.

$n_2 = \binom{N}{T} - n_1 - \binom{N-1}{T}$ = number of instances of one treatment from a pair (e.g. A or B from AB) occurring without the other, (e.g. for $T = 2$ and $N = 4$ there are two instances where just A or B are received: AC and AD for A, and BC and BD for B).

We will now consider scenarios where: the number of treatments (N) has been decided and there is flexibility in the number of treatments each subject may receive (T) (and hence the duration of the study); and where subjects receive a fixed number of treatments (T) but there is flexibility in the total number of treatments to be included in the study (N); and use the mixed model variance formula to examine alternative designs.

Varying the number of treatments per subject (T) when the total number of treatments (N) has been set

We assume that there is flexibility in the number of treatments that subjects can receive (T) but that a fixed number of treatments (N) are used in the study. It is assumed that all treatment comparisons are of equal interest, and so the BIB design is the best choice of design when it is not possible for subjects to receive all treatments. We consider a scenario where eight treatments in total are to be assessed ($N = 8$) and consider how the variance of the treatment difference is affected by varying numbers of treatments per subject (T) and varying values of the ratio of the between- and within-subject variances (γ). Variances are shown in Figure 8.4 as a ratio to the variance for the simplest design with two treatments per subject ($T = 2$) and assuming a fixed number of assessment sessions. The number of assessment sessions is the number of subjects multiplied by the number of treatments received per subject (T). It is fixed so that there is a fair comparison between the designs. However, comparisons could alternatively have been made between

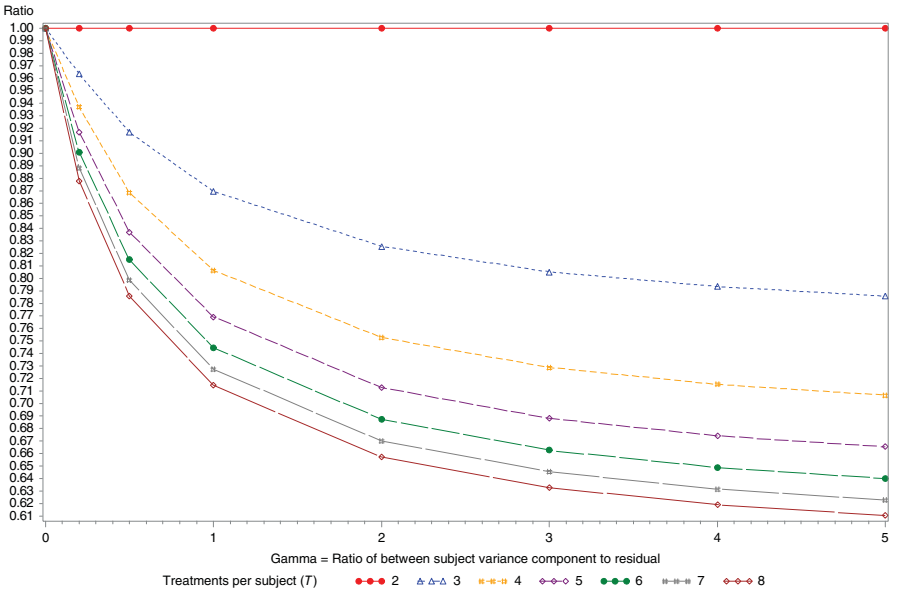


Figure 8.4 Variance of treatment difference for varying T with N fixed at 8. Expressed as ratio of variance for study with $T = 2$ and for a fixed number of assessment sessions.

designs using a fixed number of subjects if this was of more importance. The SAS code for calculating these variances will be provided at the end of the section.

If γ is positive, there is a reduction in the variance of the treatment difference as the number of treatments per subject (T) increases. The reduction is greater for larger values of γ (i.e. the between-subject variance is high compared to the residual, and so there is a higher correlation between the repeated observations on the same subject). Thus, using as many treatments as possible per subject provides the most efficient design. Of course, the choice of T may be limited by the number of assessments on subjects that is practical, and the increased risks associated with using more treatments such as more subjects dropping out would need to be weighed against the expected gain in efficiency. Note that the most efficient study where all treatments are received by subjects (i.e. $T = N$) would be no longer 'incomplete'.

Varying the number of treatments to be assessed (N) when the number of treatments per subject (T) is fixed

Sometimes, the number of treatments that may be used per subject (T) will be limited by practical constraints, for example, if measurements are taken from each eye of a subject, then only two treatments per subject are possible. However, there may still be flexibility in the total number of treatments to be assessed in the study (N).

In this section, we consider a design where the number of treatments per subject (T) is limited to two and consider variances for varying numbers of treatments (N). The design will be compared to the situation where all treatment pairs are evaluated in separate studies (i.e. $N = 2$, not necessarily a sensible approach when N becomes too large). The variance is expressed as a ratio of the variance obtained from separate studies and for a fixed total number of assessment sessions per treatment.

The variance is notably reduced as more treatments are assessed (Figure 8.5), although the reduction becomes a little less pronounced for larger values of γ. However, the reduced variance achieved by including more treatments in a single study would need to be weighed against the possible disadvantages of a carrying out one larger study, rather than several smaller ones. Note comparisons could alternatively be made between designs with a fixed number of subjects or a fixed total number of sessions over all treatments if either of these a more relevant measure of study cost.

8.18.3 Studies where only comparisons to a particular treatment are of interest

In some studies, the main interest is in comparing treatments to one particular treatment, which we will denote by T^*. This may be an established comparator treatment or, alternatively, it may be a new treatment that needs to be compared to

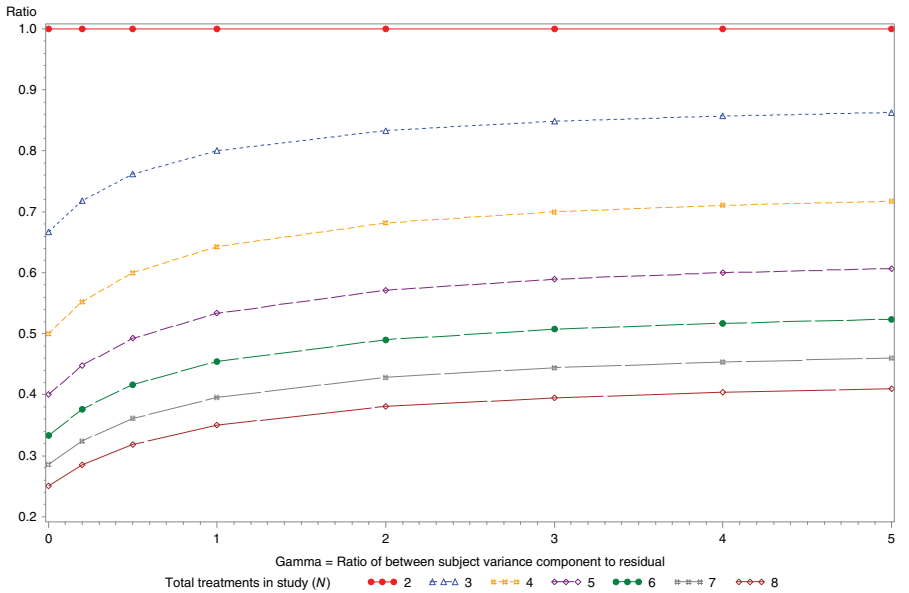


Figure 8.5 Variance of treatment difference for varying N and with T fixed at 2. Expressed as ratio of variance for study with $N = 2$, and for studies with a fixed number of assessment sessions per treatment.

several alternatives. A simple design to produce a lower variance for comparisons to T^* might have two treatments per subject ($T = 2$), with T^* always being one of the treatments received. For example, in a design to compare three comparator treatments (B, C, D) to treatment A (T^*), subjects would receive treatment pairs AB, AC and AD, and no subjects would receive the other pairs BC, BD and CD. We refer to the comparisons to T^* as 'primary comparisons', and to this design as a 'selected comparison design'. It is still an incomplete block design, but it is not balanced because treatments are not allocated equally to subjects. A selected comparison design will always be more efficient than a BIB design for the primary comparisons. It is helpful to know the size of the reduction in variance, both for sample size estimation and for choosing between different designs when there is flexibility in T or the number of comparator treatments to compare to T^* (M).

Comparing selected comparison designs to a balanced incomplete block design, for $T = 2$

In this simple situation and assuming an equal allocation of the non-T^* treatments, the variance of the primary comparisons may be written as

$$\text{Var}(\text{Treatment} - T^*) = \frac{(M(1 + 2\gamma) + 1)\sigma_r^2}{RM(1 + \gamma)}$$

Where

σ_r^2 = residual (within unit) variance
$\gamma = \sigma_s^2/\sigma_r^2$
σ_s^2 = subject variance component
M = number of non-T^* treatments
R = number of subjects receiving each set of treatment pairs (so the total number of subjects is RM)

Although no subjects receive the pairs of non-T^* treatments (e.g. BC, BD and CD in a design involving four treatments where $T^* = $ A), comparisons between these pairs may still be made, and their variance is readily obtained from the between-subject variance $(1 + \gamma)\sigma_r^2$ as

$$\text{Var}(\text{Treatment difference}) = \frac{2(1 + \gamma)\sigma_r^2}{R}$$

We refer to these as 'secondary comparisons'. If the secondary comparisons of non-T^* treatments are also of some interest, their increase in variance compared to the BIB would need to be weighed against the reduction in variance for the primary comparisons.

The variance of the primary comparisons is shown as a ratio of the variance for a BIB design in Figure 8.6, for varying values of M and γ. Although there is always some benefit in the selected comparison design, the reduction in variance is very small. Interestingly, the ratios of variances are identical when the number

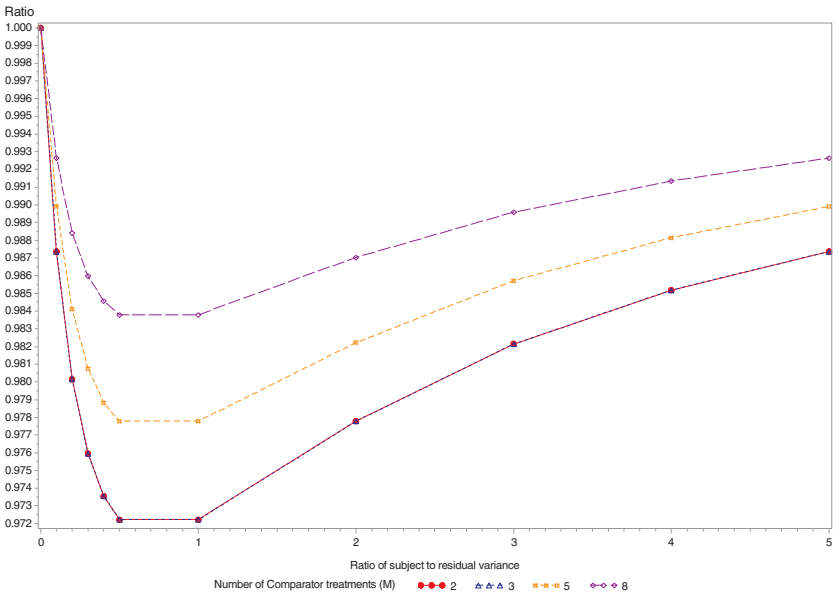


Figure 8.6 Ratio of variance for primary comparisons to BIB design variance, for $T = 2$ and a fixed number of subjects (and assessment sessions). M = number of comparator treatments (excluding T^*).

of comparator treatments (M) is either 2 or 3. As expected, the variance of the secondary comparisons is increased compared to the BIB design (Figure 8.7). If secondary comparisons are also of interest, it is debatable whether the selected comparison design is preferable, given the increase in variance of secondary comparisons is relatively large compared to the small reductions in variance obtained for the primary comparisons.

How much is the variance of primary comparisons reduced in a selected comparison design compared to a balanced incomplete block design with $T = 3$ and $M = 5$?

Next, we consider a scenario where five treatments are to be compared to T^* (i.e. $M = 5$), and three treatments are received by each subject ($T = 3$). Only treatment combinations that include T^* will be allocated to subjects in the selected comparison design. For example, if the comparator treatment (T^*) is A, the design would include only treatment combinations: ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF and AEF, and not the other possible combinations. For this design, it was more straightforward to compute variances numerically using `PROC MIXED` to perform matrix multiplication, than to derive an algebraic formula. Variances for the primary and secondary comparisons as ratios to the BIB variance are shown in Figure 8.8. The variance of the primary comparisons is always less than the BIB variance, although the reduction is never more than 20%. Thus, again, there is always benefit in the selected comparison design if the only objective is to compare treatments to T^*. Variances of secondary comparisons are increased compared to the BIB design as expected. As with the previous scenario, if the secondary comparisons are also of some interest, the increase in the variance would need to be weighed against the reduction in variance achieved for the primary comparisons.

How efficient is a selected comparison design (with $T = 2$) compared to the use of separate trials for each pair of treatments?

The variance of the primary comparisons from a selected comparison design with $T = 2$ is shown as a ratio for the variance that would be obtained by using separate trials for each primary comparison and assuming an equal number of subjects in total (Figure 8.9). The selected comparison design is always more efficient, particularly for higher values of γ, and has the advantage over separate studies that secondary comparisons of treatments are also available. Of course, any practical difficulties associated with assessing all treatments in a single study would need to be weighed against the expected gain in efficiency.

8.18.4 Designs to produce lower variances for a specific treatment pair

Sometimes, the comparison of a particular pair of treatments is of primary interest, while other treatment comparisons are of secondary interest. In another situation,

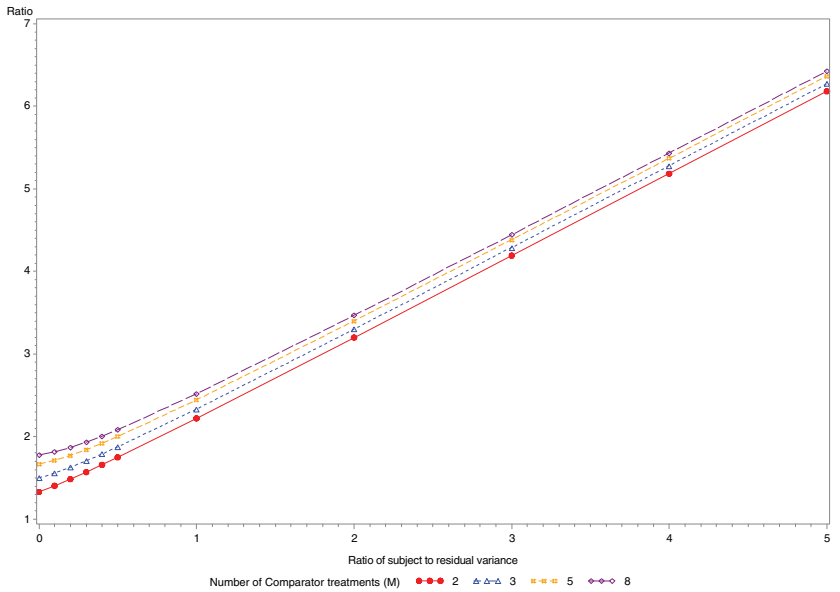


Figure 8.7 Ratio of variance for secondary comparisons to BIB design variance, for $T = 2$ and a fixed number of subjects (and assessment sessions). M = number of comparator treatments (excluding T^*).

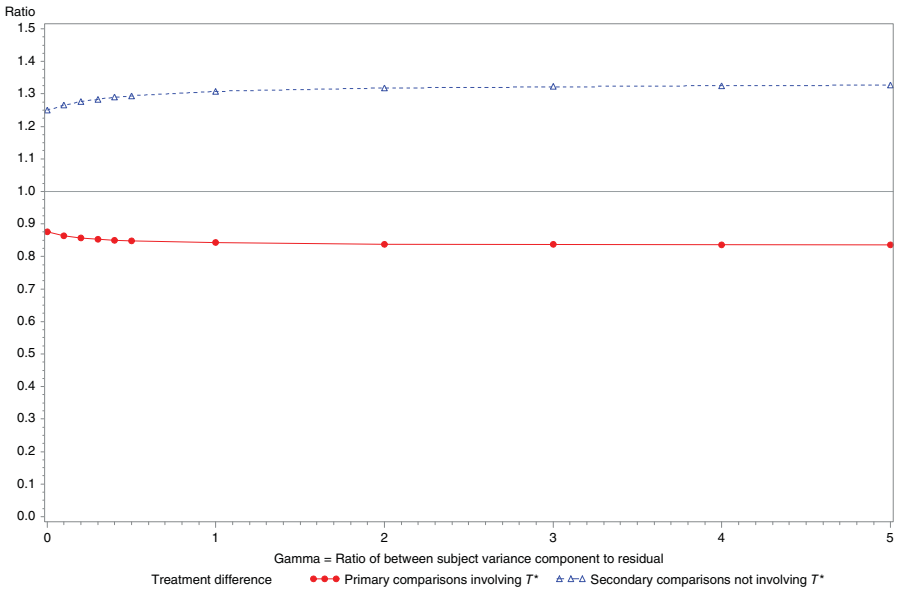


Figure 8.8 Variance of primary and secondary comparisons as ratio of BIB variance, for $T = 3$ and $m = 5$, comparing studies with same number of subjects.

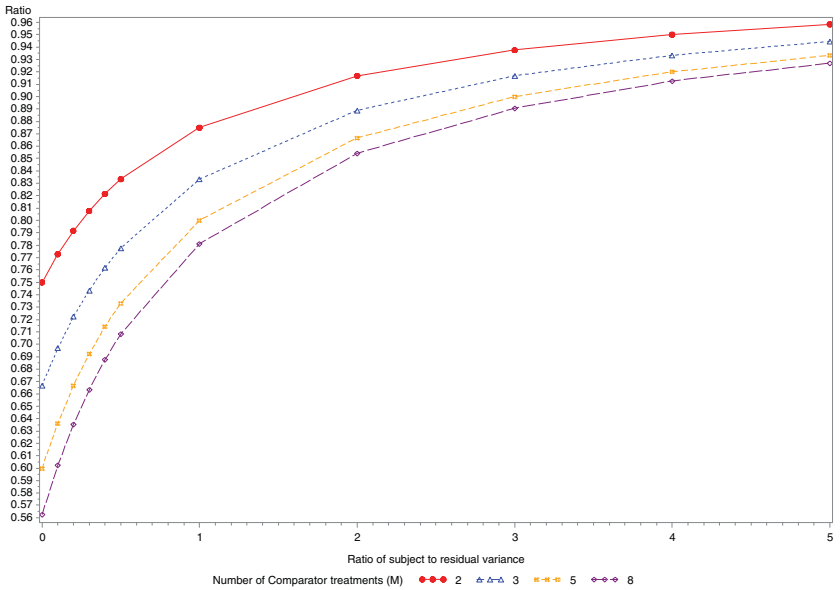


Figure 8.9 Ratio of variance for primary comparisons to variance from separate studies for each primary comparison, for $T = 2$ and a fixed number of subjects (and assessment sessions). M = number of comparator treatments (excluding T^*).

the difference between two particular treatments is expected to be smaller, but it is still important to detect the difference as significant. In both these situations, it is desirable for the variance of the treatment difference to be smaller, and we refer to the treatment pair of interest as the 'primary pair'. Two different approaches to achieving a reduced variance are considered.

Approach 1: The design includes **only** treatment combinations containing the comparison pair of interest. For example, if $T = 3$ and $N = 4$ and treatments AB form the primary pair, the design will allocate only combinations ABC and ABD to subjects. Note this design is only of interest if the number of treatments per subject (T) is three or more, otherwise all subjects would receive the primary pair, and the study would not have an incomplete block design.

Approach 2: The BIB design is adapted by including extra subjects receiving the primary treatment pair to the usual BIB treatment combinations. For example, in a design comparing four treatments (A, B, C and D), if AB is the primary pair, then extra subjects would receive AB on top of the BIB allocations (AB, AC, AD, BC, BD and CD).

Both these approaches will result in lower variances for the primary comparison (e.g. A−B) compared to the BIB design but higher variances for secondary comparisons not involving A or B (e.g. C-D if 4 treatments). Variances for comparisons only involving one of the primary pair (e.g. A−C, A−D, B−C and B−D) will be also be higher, but not as high as for the secondary comparison. We note, however, that if **only** comparison A-B was of interest, then obviously a study where all subjects receive AB would be preferable. However, this would not be a BIB design. Here we assume that there is still some interest in the other treatment comparisons.

When Approach 2 is used for a design where subjects receive two of four treatments ($T = 2$ and $N = 4$) and an extra m subjects receive the primary pair in addition to one full set of the BIB treatment allocations, the variance of the primary treatment comparison (assumed to be A−B) may be written as

$$\mathrm{Var}(A - B) = \frac{2(2\gamma + 1)\sigma_r^2}{2(m + 2)\gamma + m + 3}$$

Comparison C-D is the only comparison not involving A or B and is unaffected by the extra subjects added with AB. It has variance as

$$\mathrm{Var}(C - D) = \frac{2(2\gamma + 1)\sigma_r^2}{4\gamma + 3}$$

Comparisons involving just one of A or B have a variance in between these two variances

$$\mathrm{Var}(A - C) = \mathrm{Var}(A - D) = \mathrm{Var}(B - C) = \mathrm{Var}(B - D)$$
$$= \frac{(2\gamma + 1)(2(3m + 8)(m + 6)\gamma^2 + 3(m + 3)(3m + 16)\gamma + 3(m + 3)(m + 6)\sigma_r^2}{(4\gamma + 3)(2(m + 2)\gamma + m + 3)(2(m + 6)\gamma + 3(m + 3))}$$

If this set of subjects (i.e. the full set of the BIB treatment allocations plus m subjects receiving the primary pair) is replicated R times, then the variances are, of course, divided by R.

For more complex designs where $T > 2$ or $N > 4$, it is usually more straightforward to compute variances directly using matrix multiplication.

Design with $T = 2$ and $N = 4$ with extra subjects receiving primary comparison

The variance for the primary comparison (A–B) is always less when extra subjects receive AB than the BIB design (Figure 8.10). However, there is an increase in the variance of the secondary comparison (C–D) compared to BIB design (Figure 8.11), which is unrelated to the size of γ. Comparisons involving only one treatment from the primary pair (A–C, A–D, B–C and B–D) also have an increase in variance, which increases with increasing γ, but this is noticeably smaller than the increase for the secondary comparison, C–D (Figure 8.12). The reduction in variance for the primary comparison would need to be weighed against the amount of increase in variance for the other comparisons. In this case, it is possible to vary the proportion of subjects receiving allocations with AB by varying m, so that the desired differential of variances is achieved.

Design with $T = 3$ and $N = 6$ using only treatment combinations including the primary pair

Next, we consider a design where six treatments are to be assessed ($N = 6$) and three treatments are received per subject ($T = 3$). Only treatment combinations containing the primary pair (AB) are allocated to subjects (ABC, ABD, ABE and ABF). The variances of treatment differences are calculated using matrix multiplication for different values of γ. The variance for the primary comparison (A–B) is always less than the BIB variance, but, as with the previous scenario, there is an increase in the variance of the secondary comparisons compared to the BIB design and a smaller increase for comparisons involving only one of the primary pair (Figure 8.13). The lower variances achieved for A–B will need to be weighed against the increased variances for comparisons not involving A and B or only involving one of A or B. Again, it is possible to vary the proportion of subjects receiving allocations including AB, so that the desired variance differentials are achieved.

8.18.5 Design to produce lower variances for more than one treatment pair

Sometimes comparisons between more than one pair of treatments are of primary interest or alternatively the differences between more than one pair of treatments

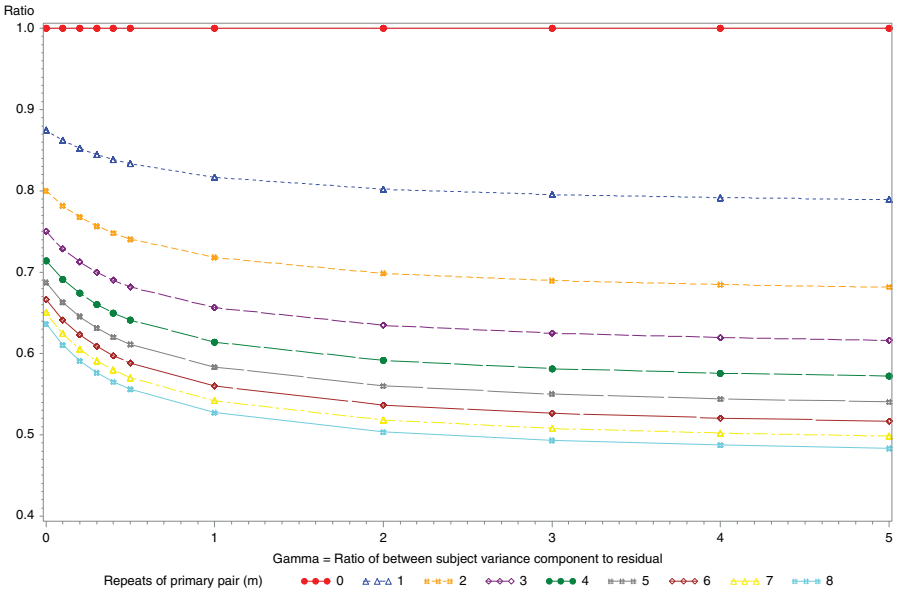


Figure 8.10 Variance of primary comparison as ratio to BIB variance, for $T = 2$, $N = 6$ and the primary pair repeated m times, comparing designs with same numbers of subjects (and assessment sessions).

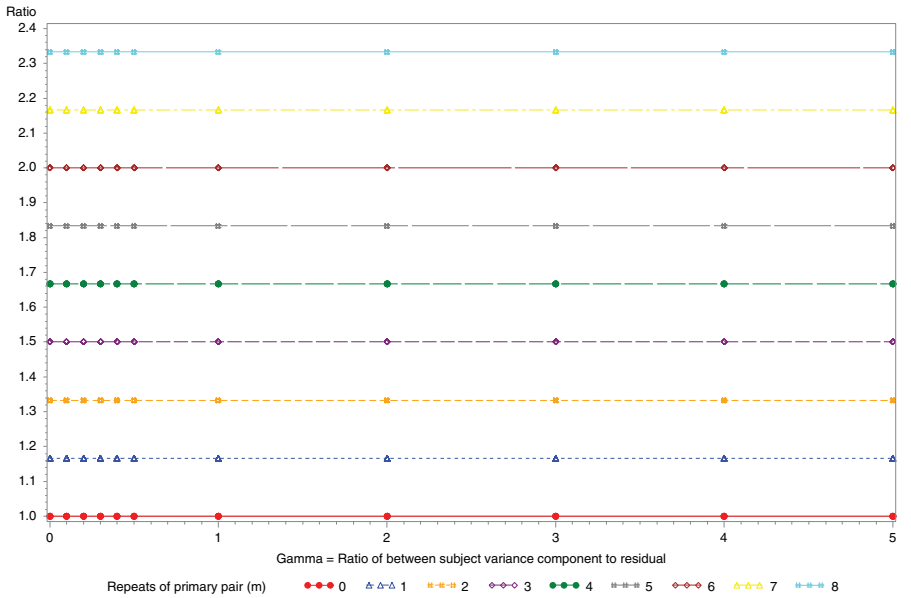


Figure 8.11 Variance of secondary comparisons as ratio to BIB variance, for $T = 2$, $N = 6$ and primary pairs repeated m times, comparing designs with same numbers of subjects (and assessment sessions).

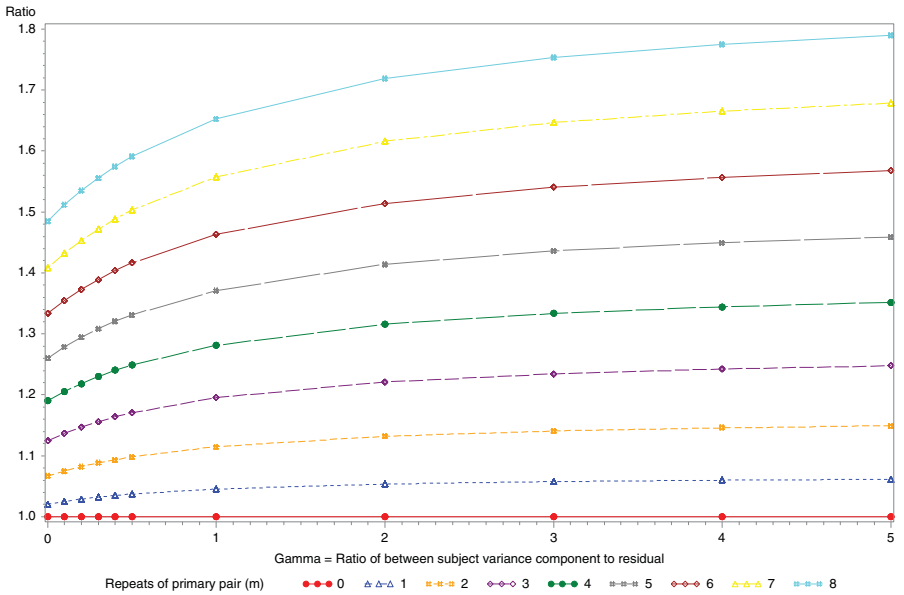


Figure 8.12 Variance of comparisons involving only one of primary pair treatments as ratio to BIB variance, for $T = 2$ and $N = 6$. Primary pairs repeated m times, comparing designs with same numbers of subjects (and assessment sessions).

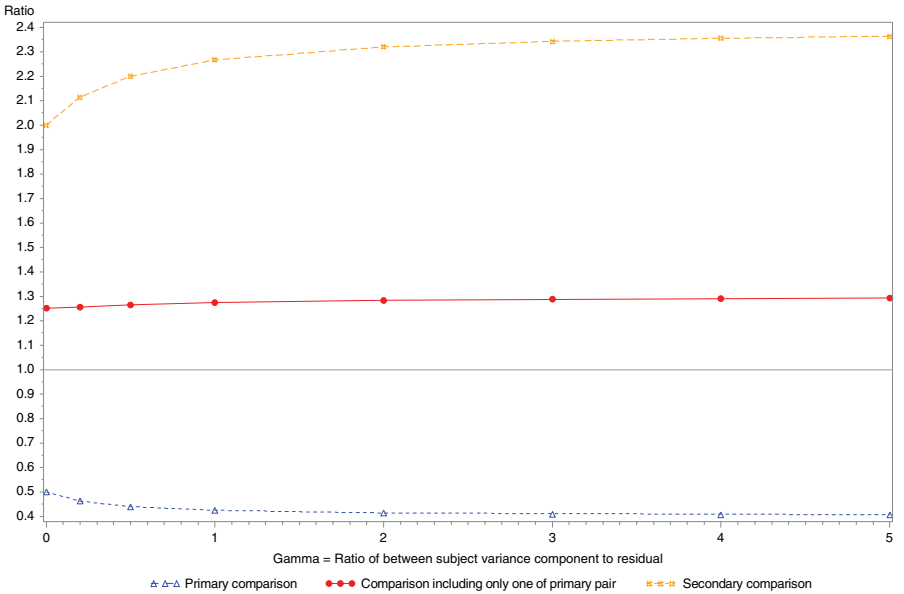


Figure 8.13 Ratio of variances of treatment differences to BIB variance. Design including only combinations with A and B, $T = 3$ and $N = 6$.

may be expected to be smaller than those between other treatments. In both these scenarios, we refer to the treatment pairs of interest as the 'primary pairs', and it is desirable for the variances of their differences to be smaller than for the other treatment pairs. However, we assume that there is still some interest in comparing the other pairs. Two approaches are considered for achieving a reduced variance for the primary pairs.

Approach 1: A simple strategy is to include **only** treatment combinations involving the primary pairs. For example, consider a design where subjects receiving two treatments ($T = 2$) and four treatments (A, B, C and D) are compared. If AB and CD are the primary pairs, then all subjects receive either AB or CD, and no subjects receive the other secondary pairs (AC, AD, BC and BD). The variance of the primary treatment comparisons is simply $2\sigma_r^2/R$ based on the residual variance. The variance of the secondary treatment comparisons is $2(\gamma + 1)\,\sigma_r^2/R$ based on the between-subject variance. Thus, the variance of the primary treatment comparisons is lower than in the BIB variance, and for secondary comparisons, it is higher. This design has the same efficiency as the use of a separate study for each primary treatment pair but has the advantage of additionally providing comparisons of the secondary treatment pairs (AC, AD, BC and CD).

In designs where subjects receive more than two treatments ($T \geq 3$), all combinations involving the primary treatment pairs would be allocated to subjects. For example, if T = 3 and the study includes four treatments (A, B, C and D) with primary pairs AB and CD, the design will allocate only combinations ABC, ABD, ACD and BCD to subjects. Again, this will result in lower variances for A−B and C−D compared to the BIB design, but higher variances for the secondary treatment comparisons (A−C, A−D, B−C and B−D). For these designs, algebraic formulae are cumbersome, and it is easier to compute variances directly using matrix multiplication.

Approach 2: An alternative approach that also leads to lower variances for the primary treatment pairs is to add extra subjects receiving the primary treatment pairs to the full complement of treatment pairs in the BIB. For example, in a design comparing four treatments (A, B, C and D) with primary treatment pairs AB and CD, extra subjects receive AB or CD on top of the BIB allocations (AB, AC, AD, BC, BD and CD). The reduction in primary comparison variances in this design will not be as great as for Approach 1, but there is more scope to achieve the desired variance differential by varying the number of extra subjects receiving the primary pairs. The differential will become greater as the number of extra subjects receiving combinations containing the primary treatment pairs increases, with the maximum differential is achieved when all subjects receive the primary pairs (i.e. reverting to Approach 1).

In a study of this type where subjects receive two of four treatments ($T = 2$ and $N = 4$) and an extra m subjects receive each of the primary pairs in addition to each set of the BIB allocations (AB, AC, AD, BC, BD and CD), the variance for primary

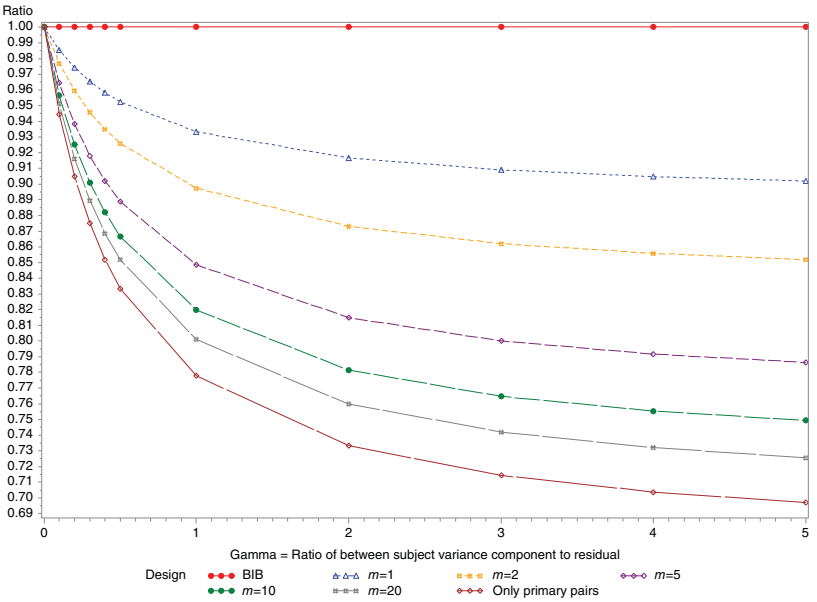


Figure 8.14 Variance of primary comparisons as ratio of BIB variance, for studies with same number subjects. Designs using *m* extra subjects receiving primary pairs in addition to BIB allocations.

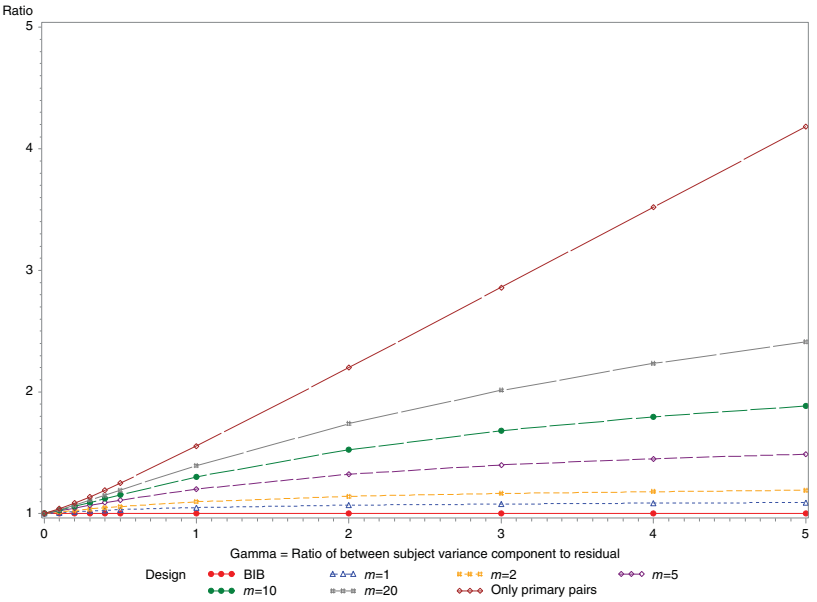


Figure 8.15 Variance of secondary comparisons as ratio of BIB variance, for studies with same number subjects. Designs using *m* extra subjects with each primary pair, in addition to BIB allocations.

comparisons (assumed to be A−B and C−D) is given by

$$\mathrm{Var}(A-B) = \mathrm{Var}(C-D) = \frac{4(2\gamma+1)\sigma_r^2}{2m(2\gamma+1)+2(4\gamma+3)}$$

For the secondary treatment comparisons it is

$$\mathrm{Var}(A-C) = \mathrm{Var}(A-D) = \mathrm{Var}(B-C) = \mathrm{Var}(B-D) = \frac{2(((m+4)\gamma+m+3)(2\gamma+1))\sigma_r^2}{2((m+2)\gamma+m+3)(4\gamma+m+3)}$$

For other designs involving more treatments per subject ($T > 2$) or total treatments ($N > 4$), the algebraic formulae are cumbersome, and it is easiest to compute variances directly using matrix multiplication.

Comparing variances of alternative designs when $T = 2$ and $N = 4$

The variance for the differences between the primary treatment comparisons (A−B and C−D) is reduced in a design including only the primary pairs (Approach 1) and to a lesser extent in designs adding additional subjects receiving primary pairs to the BIB allocations (Approach 2). Both are compared to a BIB using the same number of subjects (Figure 8.14). As expected, the variances of the comparisons of secondary treatment pairs are larger than in the BIB design (Figure 8.15). Again, the benefits of lower variances for the primary comparisons would need to be weighed against the disadvantage of higher variances for the secondary comparisons. The number of extra subjects (m) included in Approach 2 may be varied to achieve the desired balance between the primary and secondary pair variances.

SAS Code

The SAS code is given to obtain Figures 8.4, 8.5 and 8.9. The variances for Figures 8.4 and 8.5 were obtained using formulae, and for Figure 8.9, `PROC MIXED` was used to compute variances for alternative designs. SAS code for the other figures is available on the book web pages at www.wiley.com/go/brown/applied_mixed.

Code for Figures 8.4 and 8.5

```
OPTIONS ORIENTATION=LANDSCAPE;
SYMBOL1 C=R L=1 V=DOT I=JOIN;
SYMBOL2 C=BLUE L=1 V=TRIANGLE I=JOIN;
SYMBOL3 C=ORANGE L=1 V=HASH I=JOIN;
SYMBOL4 C=PURPLE L=1 V=HASH I=JOIN;
SYMBOL5 C=GREEN L=1 V=HASH I=JOIN;
SYMBOL6 C=GREY L=1 V=HASH I=JOIN;
SYMBOL7 C=BROWN L=1 V=HASH I=JOIN;
```

```
* CALCULATE VARIANCE FOR DESIGNS WITH VARYING N AND T, AND
     DIFFERENT VALUES FOR GAMMA;
DATA ALL_DESIGNS;
NS = 100; * SET FIXED NUMBER OF SESSIONS TO USE FOR VARIANCE
     CALCULATION;
DO N=2 TO 8; * TOTAL NUMBER OF TREATMENTS IN STUDY;
DO GAMMA=0, 0.2, 0.5, 1.0, 2, 3, 4, 5;
DO T=2 TO N; * T=TREATMENTS PER SUBJECT;
NC=COMB(N,T); NP=COMB(N-2,T-2); NU = NC-NP-COMB(N-1,T);
IF NU=. THEN NU=0;
VAR = 2*(T*GAMMA+1)/(N*NP*GAMMA+NP+NU); * VARIANCE FOR
     STUDY USING ALL TREATMENT COMBINATIONS,
     SO NC*T TREATMENT SESSIONS;
VAR_PER_SESSIONS = VAR*NC*T/NS; * MULTIPLY VAR BY NC*T AND
     DIVIDE BY NS, TO GET VARIANCE FOR NS SESSIONS;
OUTPUT;
END; END; END;
PROC SORT; BY N GAMMA;

* OBTAIN VARIANCES FOR 2 TREATMENTS PER SUBJECT;
DATA T_TWO; SET ALL_DESIGNS;
IF T=2;
VAR2_PER_SESSIONS=VAR_PER_SESSIONS;  * RELABEL VARIANCE
     FOR T=2;
KEEP GAMMA N VAR2_PER_SESSIONS;
PROC SORT; BY N GAMMA;

DATA FIXEDN; MERGE ALL_DESIGNS T_TWO; BY N GAMMA;  * MERGE
     VARIANCES FOR ALL DESIGNS WITH VARIANCES FOR T=2;
IF N=8; * SELECT ONLY N=8 FOR PLOT;
RATIO_PER_SESSIONS=VAR_PER_SESSIONS/VAR2_PER_SESSIONS;

PROC GPLOT;
TITLE "8.18 VARIANCE OF TREATMENT DIFFERENCE FOR VARYING T
     AND N=8";
TITLE3 "AS RATIO TO VARIANCE FOR STUDY WITH T=2 AND
     FOR A FIXED NUMBER OF ASSESSENT SESSIONS";
LABEL T='TREATMENTS PER SUBJECT (T)'
     RATIO_PER_SESSIONS='RATIO' GAMMA='GAMMA = RATIO OF
     BETWEEN SUBJECT VARIANCE COMPONENT TO RESIDUAL';
PLOT RATIO_PER_SESSIONS*GAMMA=T; RUN;

DATA N_SET; SET ALL_DESIGNS;
IF N=T;
VARN_PER_SESSIONS=VAR_PER_SESSIONS;
KEEP GAMMA N VAR_PER_SUBJECTS3 VARN_PER_SESSIONS;
PROC SORT; BY N GAMMA;
```

```
DATA FIXEDT; MERGE ALL_DESIGNS N_SET; BY N GAMMA;
IF T=2;
RATIO_PER_SESSIONS_PER_TREAT = (VAR_PER_SESSIONS/N) /
     (VARN_PER_SESSIONS/2);
PROC SORT; BY N GAMMA;

AXIS1 LABEL=(F="ARIAL" "GAMMA = RATIO OF BETWEEN SUBJECT
     VARIANCE COMPONENT TO RESIDUAL");
PROC GPLOT;
TITLE "FIGURE 8.5 VARIANCE OF TREATMENT DIFFERENCE FOR
     VARYING N AND T=2";
TITLE3 'AS RATIO OF VARIANCE FOR STUDY WITH N=2 AND FOR
     TRIALS WITH A FIXED NUMBER OF ASSESSENT SESSIONS PER
     TREATMENT';
LABEL N='TOTAL TREATMENTS IN STUDY (N)'
     RATIO_PER_SESSIONS_PER_TREAT='RATIO';
PLOT RATIO_PER_SESSIONS_PER_TREAT*GAMMA=N/ HAXIS=AXIS1;
```

Code for Figure 8.9

```
* ALL COMBINATIONS OF 3 FORMING INCOMPLETE BLOCK DESIGN WHERE
     SUBJECTS RECEIVE 3 OF 6 TREATMENTS;
DATA IB; INPUT TREAT1 $ TREAT2 $ TREAT3 $;
CARDS;
A B C
A B D
A B E
A B F
A C D
A C E
A C F
A D E
A D F
A E F
B C D
B C E
B C F
B D E
B D F
B E F
C D E
C D F
C E F
D E F
;

* COMBINATIONS OF 3 OMITTING ALL THOSE WITHOUT TREATMENT A,
     FORMING FAVOURED TREATMENT DESIGN;
```

```
DATA SELECT; INPUT TREAT1 $ TREAT2 $ TREAT3 $;
CARDS;
A B C
A B D
A B E
A B F
A C D
A C E
A C F
A D E
A D F
A E F
;

* MACRO TO OBTAIN VARIANCES FOR COMPARISONS;
%MACRO GET_VARS(DESIGN,GAMMA);
  DATA A; SET &DESIGN;
  ID=_N_;
  PROC TRANSPOSE OUT=P2; BY ID; VAR TREAT1 TREAT2 TREAT3;

  PROC MEANS DATA=A; VAR ID; OUTPUT OUT=NUM_SUBJECTS
     N=NUM_SUBJECTS;

  * GENERATE VALUES FOR THE ANALYSIS (ACTUAL VALUES
     IRRELEVANT);
  DATA P3; SET P2;
  TREAT=COL1;
  Y=1; Y=RANNORM(1);

  PROC MIXED; CLASS TREAT ID;
  MODEL Y=TREAT;
  * FIX THE VALUES OF GAMMA AND RESIDUAL (ACTUAL VALUES OF
     Y IRRELEVANT);
  PARMS &GAMMA 1 / HOLD=1 2 NOITER NOPROFILE;
  RANDOM ID;
  LSMEANS TREAT/ DIFF;
  ESTIMATE 'A-B' TREAT 1 -1 0; * COMPARISON TO T*;
  ESTIMATE 'B-C' TREAT 0 1 -1; * COMPARISON BETWEEN 2
     OTHER TREATMENTS;
  ODS OUTPUT ESTIMATES=EST;

  * CREATE DATASET WITH VARIANCES OF A-B AND B-C,
     ALONG WITH NUMBER OF
  OBSERVATIONS IN DATASET;
  DATA VAR; SET EST; IF _N_ EQ 1 THEN DO; SET NUM_SUBJECTS;
  END;
  LENGTH DESIGN $ 20;
  DESIGN="&DESIGN";
```

```
  GAMMA="&GAMMA"*1;
  VAR=STDERR**2;

  * CONCATENATE WITH RESULTS FOR OTHER VALUES OF GAMMA;
  DATA &DESIGN._VAR; SET &DESIGN._VAR VAR;
  RUN;
%MEND;

%MACRO OVER_GAMMA(DESIGN); * RUN ABOVE MACRO OVER SEVERAL
     VALUES FOR GAMMA;
  %GET_VARS(&DESIGN,0.1);%GET_VARS(&DESIGN,0.2);
  %GET_VARS(&DESIGN,0.3);%GET_VARS(&DESIGN,0.4);
  %GET_VARS(&DESIGN,0);%GET_VARS(&DESIGN,0.5);
  %GET_VARS(&DESIGN,1);%GET_VARS(&DESIGN,2);
  %GET_VARS(&DESIGN,3);%GET_VARS(&DESIGN,4);
  %GET_VARS(&DESIGN,5);*%GET_VARS(&DESIGN,10);
%MEND;

DATA IB_VAR; SET _NULL_;
DATA SELECT_VAR; SET _NULL_;

%OVER_GAMMA(IB); %OVER_GAMMA(SELECT);

* MODIFY DATASET CONTAINING RESULTS FOR IB DESIGN;
DATA IB_VAR2; SET IB_VAR;
IB_SUBJECTS=NUM_SUBJECTS;
IB_VAR=VAR;
IF DESIGN='IB' AND LABEL='A-B' THEN OUTPUT; * OUTPUT
     RESULTS FOR ONE COMPARISON, DOESNT MATTER WHICH;
KEEP GAMMA IB_SUBJECTS IB_VAR;
PROC SORT; BY GAMMA;

PROC SORT DATA=SELECT_VAR; BY GAMMA;

* MERGE RESULTS FOR 2 DESIGNS BY GAMMA AND CALCULATE
     RATIOS OF VARIANCES;
DATA MERGED; MERGE SELECT_VAR IB_VAR2; BY GAMMA;
RATIO=VAR/IB_VAR;
SCALE=NUM_SUBJECTS/IB_SUBJECTS;
RATIO_SCALED=RATIO*SCALE;
PROC SORT; BY LABEL GAMMA;

OPTIONS ORIENTATION=LANDSCAPE;
SYMBOL1 C=R L=1 V=DOT I=JOIN;
SYMBOL2 C=BLUE L=1 V=TRIANGLE I=JOIN;
SYMBOL3 C=ORANGE L=1 V=HASH I=JOIN;
```

```
PROC FORMAT; VALUE $TDIFF 'A-B'='PRIMARY COMPARISONS
     INVOLVING T*'
 'B-C'='SECONDARY COMPARISONS NOT INVOLVING T*';

PROC GPLOT DATA=MERGED;
PLOT RATIO_SCALED*GAMMA=LABEL/ VREF=1 VAXIS=0 TO 1.5 BY 0.1;
TITLE 'FIGURE 8.9 VARIANCE OF PRIMARY AND SECONDARY
     COMPARISONS AS RATIO OF BIB VARIANCE, FOR T=3 AND N=6';
TITLE3 "COMPARING STUDIES WITH SAME NUMBER OF SUBJECTS";
LABEL LABEL='TREATMENT DIFFERENCE' GAMMA='GAMMA = RATIO OF
     BETWEEN SUBJECT VARIANCE COMPONENT TO RESIDUAL'
     RATIO_SCALED='RATIO';
FORMAT LABEL $TDIFF.;
```

9

Software for fitting mixed models

In this chapter, we will look at software for fitting mixed models, with a particular emphasis on the SAS package, which has been used to analyse the majority of our examples. In Section 9.1, we mention briefly some of other packages and programs that are available for fitting mixed models. Basic details on the use of the SAS procedure PROC MIXED for fitting normal mixed models are given in Section 9.2. In Section 9.3, we give details on using PROC GENMOD and PROC GLIMMIX for fitting GLMMs, and in Section 9.4, details on PROC MCMC for fitting models using a Bayesian approach.

9.1 Packages for fitting mixed models

There have been very many different programs and software packages developed to implement mixed models or multi-level models. These range from programs to implement a single type of model to comprehensive statistical packages such as SAS. Some are regularly updated to run on the latest versions of operating systems, while others are only available on relatively old operating systems. New products are constantly being introduced to the market, often for a very specific application. As an example, GEMMA is software 'implementing the Genome-wide Efficient Mixed Model Association algorithm for a standard linear mixed model and some of its close relatives for genome wide association studies (GWAS)'. Further details can be found at (home.uchicago.edu/xz7/software.html). It is impractical to list all of the software available for fitting mixed models, and we will not attempt to try it. We will, however, say a few words about some of the packages available. Inclusion within the following list is not an endorsement of these packages, and neither should the absence of any package be taken as a criticism.

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

License fees for many of the more comprehensive statistical software packages can be quite expensive, and so it is a wise precaution to ensure that the capabilities match all of a user's needs and not just those relating to mixed models. For those unable to afford license fees, it is worth mentioning R software. This is free software that is now used widely, but by no means exclusively, in the academic community. The authors of this book are not themselves R users, but some of our colleagues are enthusiastic, particularly about its graphics capabilities. It requires an initial investment of time to learn R but is well worth considering, especially when funding is an issue, though its mixed models capability is not as good as that provided by SAS.

SAS The MIXED and GLIMMIX procedures provide very versatile software for fitting mixed models. They can be used to fit all types of mixed models (random effects, random coefficients and covariance pattern models). In addition, the MCMC procedure may be used to fit mixed models using a Bayesian approach. More details are provided in Sections 9.2–9.4.

Genstat Mixed models for normal data can be fitted using the VCOMP and REML directives, and GLMMs can be fitted using the GLMM directive. In addition, a set of procedures to fit the GLMM models described by Lee and Nelder (1996) is available. These allow the random effects in GLMMs to assume alternative distributions to the normal distribution. The VSTRUCTURE directive can be used to specify covariance patterns. There is not a directive for analysing ordinal data; however, the Biometris GenStat Procedure Library written by members of the Centre for Biometry Wageningen (CBW) contains a suitable procedure, IRCLASS. This procedure library may be freely downloaded from http://www.wageningenur.nl/en/show/Biometris-GenStat-Procedure-Library-Edition-15.htm.

MLwiN MLwiN is a package originally developed by the Multilevel Modelling Project at the Institute of Education in London. It can be used to fit both normal mixed models, GLMMs, and mixed models for ordinal and categorical data. However, the choice of covariance patterns available is more limited than in PROC MIXED. An R command interface to the MLwiN software package is available via software called R2MLwiN. Further information on MLwiN can be found at http://www.bristol.ac.uk/cmm/software/mlwin/.

R R is a free software environment for statistical computing and graphics. Mixed models for normal data can be fitted in using 'lme4' and 'nlme4' functions. More details on these functions can be found in the textbook 'Mixed-effects models in S and S-PLUS' by Pinheiro and Bates (2000) (R functions were originally developed for use within S and S-PLUS). Covariance patterns can be specified using the 'correlation' argument. GLMMs can be fitted using the 'glmm' or 'glmmpql' functions. The 'glmmpql' function is discussed in the fourth edition of *Modern applied statistics with S*, Venables and Ripley (2002); the book also covers normal theory models; and there is online support for the book at www.stats.ox.ac.uk/pub/MASS4/. Interfaces have been written to allow use of

both MlwiN and WinBUGS software to be accessed from within R (R2MLwiN and R2WinBUGS). More information on R is available at www.r-project.org.

SPSS Mixed models for normal data can be fitted using the MIXED procedure. GLMMs may be fitted using the GLMM procedure, which is available within the Advanced Statistics software ad-on.

Stata Random effects models may be fitted using the 'xtmixed' command. Non-normal mixed models may be fitted within the GLLAMM add-on software (see www.gllamm.org/). The range of available covariance pattern models appears to be more restricted than is available in SAS.

WinBUGS This is a package dedicated to Bayesian analysis using the Gibbs sampler and has been developed by the Medical Research Council Biostatistics Unit in Cambridge. It can be used to fit random effects models to all types of data. The package can be obtained at www.mrc-bsu.cam.ac.uk/bugs. It is free of charge at the time of writing. WinBUGS software may be accessed from R using R2WinBUGS. This writes a data file, input file, and a script, runs the script in WinBUGS, and returns the output simulations to R.

9.2 PROC MIXED

PROC MIXED is a SAS procedure for fitting normal mixed models. It can be used to fit any type of mixed model (random effects, random coefficients, covariance pattern or a combination). It has great flexibility, and there are many options available for defining mixed models and their output. By default, the REML estimation method is applied (see Section 2.2.1). Alternatively, other likelihood-based methods can be applied, and a Bayesian analysis is also available for random effects and random coefficients models (see Section 2.3). Documentation for PROC MIXED is perhaps most easily accessed online during a SAS session. Details are also available from support.sas.com/documentation/onlinedoc. Another excellent text on using SAS to fit mixed models is *SAS System for Mixed Models, Second edition* by Littell *et al.* (2006).

Our aim in this section is to give a basic description of the most useful PROC MIXED statements and options to enable those not wishing to learn the procedure in depth to perform mixed models analyses. Details on the many other options available can be found in the SAS documentation. The presentation in this section will assume a working knowledge of SAS.

Syntax

The statements available in PROC MIXED in version 9.3 are as follows: PROC MIXED statement, BY, CLASS, CONTRAST, ESTIMATE, ID, LSMEANS, LSMESTIMATE, MODEL, PARMS, PRIOR, RANDOM, REPEATED, SLICE, STORE, and WEIGHT statements. In addition, ODS statements can be used to control what is printed when running PROC MIXED and what is output in SAS datasets

for handling by other procedures. The BY statement is the same as in other SAS procedures and will not be considered further in this chapter. We will mention all of the other statements, at least briefly, while concentrating on those statements we have used in this book. Those with older versions of SAS may find that some of the statements listed here are not implemented in their version.

Simple example

Data from the multi-centre hypertension trial introduced in Section 1.3 will be used to illustrate the use of PROC MIXED to fit a simple random effects model. The following code fits a random effects model with pre-treatment DBP (dbp1) and treatment effects fixed and centre and centre·treatment effects random.

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat;
RANDOM centre centre*treat;
```

```
                    The Mixed Procedure
                     Model Information
Data Set                              WORK.A
Dependent Variable                    dbp
Covariance Structure                  Variance Components
Estimation Method                     REML
Residual Variance Method              Profile
Fixed Effects SE Method               Model-Based
Degrees of Freedom Method             Containment

                  Class Level Information
Class        Levels       Values
centre           29       1 2 3 4 5 6 7 8 9 11 12 13 14
                          15 18 23 24 25 26 27 29 30 31
                          32 35 36 37 40 41
treat             3       A B C

                   Dimensions
       Covariance Parameters          3
       Columns in X                   5
       Columns in Z                 108
       Subjects                       1
       Max Obs Per Subject          288

              Number of Observations
       Number of Observations Read       288
       Number of Observations Used       288
       Number of Observations Not Used     0
```

```
                    Iteration History
Iteration      Evaluations      -2 Res Log Like      Criterion
        0                1        2072.30225900
        1                3        2055.64188178      0.00000322
        2                1        2055.63936685      0.00000000
                Convergence criteria met.
```

The Model Information gives basic information on the dataset that was used and some methodological information on how the model was fitted (in this case using the default options). The Class Level Information provides us with the values in our dataset for those categorical variables specified in the `CLASS` statement. Dimensions show the size of the matrices that `SAS` is working with. It is potentially confusing that the number of subjects is shown as one when we have 288 patients in the trial. This is because our dataset has not been 'blocked' (see Section 6.2). The Number of Observations allows us to see if any observations have been excluded from analysis, perhaps, because of missing values. The Iteration History table shows how quickly the algorithm has converged. The criterion is a measure of convergence and should be very close to zero. In this example, convergence has been reached quickly. The table gives the value of minus twice the REML log likelihood. If this is very large, then it is likely that the covariance matrix, $\mathbf{V}$, is singular, leading to an infinite likelihood. In this situation, the results would be invalid and the model would need to be respecified, probably refitting certain random effects as fixed or removing them altogether. The Evaluations column gives the number of evaluations of the objective function carried out at each iteration. Occasionally, a message stating that the G matrix is not positive definite will appear. Usually, this indicates that a negative variance component has been set to zero by `PROC MIXED` and is not a cause for concern. However, if this is not the case, it is possible that only a local maximum has been reached and a 'grid search' for other solutions might be advisable (see the `PARMS` statement).

```
 Covariance Parameter
      Estimates
Cov Parm       Estimate
centre           6.4628
centre*treat     4.0962
Residual        68.3677
```

This gives the variance parameter estimates. The variance component for centres is 6.46 and for the centre·treatment interaction 4.10. The residual estimate is 68.37.

```
          Fit Statistics
-2 Res Log Likelihood       2055.6
AIC (smaller is better)     2061.6
AICC (smaller is better)    2061.7
BIC (smaller is better)     2065.7
```

This table gives information about the model fit. It includes several statistics based on the likelihood value. `AIC` is Akaike's Information Criterion, `AICC` is a variation on `AIC` and `BIC` is Schwartz's Bayesian Information Criterion. Each of these criteria attempts to make adjustment for the number of parameters fitted in the model so that models can be compared directly. However, note that `SAS` presents minus twice the usual criteria. They are of greatest value for comparing models using different covariance patterns (see Chapter 6). More detail on their calculation is given in the `SAS` documentation.

```
          Type 3 Tests of Fixed Effects
            Num     Den
Effect       DF      DF     F Value     Pr > F
dbp1          1     208        6.31     0.0128
treat         2      48        2.28     0.1131
```

Results from *F* tests are given for all fixed effects. The 'Type 3' tests are the Wald tests described in Section 2.4.4. These tests arise naturally from mixed models and adjust for other effects fitted in the model. The denominator DF for the *F* tests are given by the residual DF or, if the fixed effect is contained, by the DF of the containing effect. Here `dbp1` is tested using the residual DF of 208 and `treat` using the centre·treatment DF of 48, since it is contained within this effect. Other methods for calculating the denominator DF are available by using the `DDFM` option in the `MODEL` statement (see the following section).

We will now give basic details on the use of each statement within the procedure.

The `PROC MIXED` statement

This statement initiates the procedure, and its options are concerned with the general aspects of fitting the procedure and specifying its output.

```
PROC MIXED options;
```

Some of the more useful PROC MIXED options are listed in the following table.

METHOD = <method>	Specifies estimation method.
= REML	REML (default).
= ML	Maximum likelihood.
= MINQUE	Minimum variance quadratic variance estimation. This is an estimation method that we have not covered. It is based on equating mean squares to their expected values, and it is less computationally expensive than maximum likelihood. Further information can be found in Rao (1971, 1972). Searle *et al.* (1992) show that the solution is equivalent to that obtained using just one iteration with REML.
CL	Prints Wald confidence intervals for covariance parameters. Default is 95% limits unless ALPHA=<p> option is used.
ASYCOV	Prints asymptotic covariance matrix of covariance parameters.
EMPIRICAL	Empirical estimator of variance matrix of fixed effects, $\text{var}(\widehat{\boldsymbol{\alpha}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\text{cov}(\mathbf{y})\mathbf{V}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, is used in place of $\mathbf{V}$ for all fixed effects variance estimates in covariance pattern models fitted using REPEATED statement (see Section 2.4.3).
IC	Displays additional information criteria to compare model fit. Criteria due to Akaike–AIC and AICC (1974), Hannan and Quinn–HQIC (1979), Schwarz–BIC (1978) and Bozdogan–CAIC (1987) are displayed.
NOCLPRINT	Suppresses printing of class levels. (This is useful when a CLASS effect has many categories.)
NOINFO	Suppresses printing of Model Information, Dimensions and Number of Observations tables.
NOITPRINT	Suppresses printing of Iteration History
PLOTS options	A variety of statistical graphics can be specified via the Output Delivery System (in more recent versions of SAS). See online help for details

The MODEL statement

This statement is used to specify the dependent variable and the fixed effects in the model. Options are available for specifying the model, significance tests and for requesting specific output to be printed.

```
MODEL <dependent> = <fixed effects>/options;
```

Options relating to the model specification

NOINT	Requests no intercept be included in the model.

Options relating to statistical tests

DDFM =<DF type>	Selects DF type for *F* test denominator DF for all *F* tests.
RESIDUAL	Uses residual DF.
CONTAIN	If the effect of interest is contained within another effect, then the DF of the containing effect are used. If the effect is not contained, then the residual DF are used. These are the default DF.
BETWITHIN	Assigns between- or within-subject DF to effects when the REPEATED statement is used. Effects nested within the SUBJECT effect take the subject effect DF, others take the residual DF.
SATTERTH	Uses Satterthwaite's approximation to the true DF (see Section 2.4.4).
KR <(FIRSTORDER)> or KENWARDROGER <(FIRSTORDER)>	Calculates DF suggested by Kenward and Roger (1997). This substitutes an improved estimate of the covariance matrix of the fixed and random effects into Satterthwaite's approximation for the DF (see Section 2.4.3). It also causes the improved estimates of fixed and random effects variances to be used throughout the whole procedure (e.g. for calculating standard errors and test statistics). Addition of the FIRSTORDER or LINEAR sub-option will sometimes lead to an improved adjustment in certain types of covariance pattern model (see Section 6.2.4).

Options relating to output

SOLUTION	Solution for fixed effects is printed.
CL	Requests *t*-type confidence intervals for fixed effects given by SOLUTION option.
ALPHA=<p>	Specifies size of confidence intervals (default 0.05).
E3	Prints design matrix for all fixed effects.
OUTP=<dataset>	For each observation gives predicted values ($\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}}$) and residuals ($\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}} - \mathbf{Z}\hat{\boldsymbol{\beta}}$) based on fixed and random effects. Approximate variances, standard errors and 95% confidence intervals for each predicted value are also listed.
OUTPM=<dataset>	Predicted values ($\mathbf{X}\hat{\boldsymbol{\alpha}}$) and residuals ($\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}$) based on fixed effects only are given. Approximate variances, standard errors and 95% confidence intervals for each predicted value are also listed.

Options relating to model checking

RESIDUAL	Produces residual plots when used with ODS graphics.
INFLUENCE	Influence diagnostic plots to be produced when used with ODS graphics.
VCIRY	Standardises the residuals in a covariance pattern model (see Section 6.2.4).

Example

```
PROC MIXED DATA=a; CLASS centre treat;
MODEL dbp = dbp1 treat/ E3 SOLUTION DDFM=KR OUTP=pred
   OUTPM=predm;
RANDOM centre centre*treat;
PROC PRINT NOOBS DATA=pred;
PROC PRINT NOOBS DATA=predm;
```

Use of the `SOLUTION` and `DDFM=KR` options

Solution for Fixed Effects

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		61.7638	11.2440	284	5.49	<.0001
dbp1		0.2689	0.1084	284	2.48	0.0137
treat	A	2.9274	1.4109	25.6	2.07	0.0482
treat	B	1.6415	1.4453	25.7	1.14	0.2666
treat	C	0	..	..	..	..

Note that use of `DDFM=KR` has caused the standard errors to be based on an improved estimate (as described by Kenward and Roger, 1997), which helps alleviate bias (see Section 2.4.3). These standard errors are also used to calculate the t statistics and their approximate DF. (It is somewhat confusing that the `DDFM=KR` option in fact relates to the standard error calculation. The DF are still calculated using Satterthwaite's approximation but based on these improved standard errors.)

Use of the `E3` option

Type 3 Coefficients for dbp1

Effect	treat	Row1
Intercept		
dbp1		1
treat	A	
treat	B	
treat	C	

Type 3 Coefficients for treat

Effect	treat	Row1	Row2
Intercept			
dbp1			
treat	A	1	
treat	B		1
treat	C	−1	−1

This shows that SAS has parameterised treatment effects by using differences from the last treatment (C).

Use of the `DDFM=KR` option

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dbp1	1	284	6.16	0.0137
treat	2	25	2.16	0.1364

Note that these tests differ from the earlier analysis where the DDFM=KR option was not used.

Use of the `OUTP=` option

patient	visit	centre	treat	dbp	dbp1	cf	cf1	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
1	6	29	C	86	97	.	1	84.9727	3.16871	18.4533	0.05	78.3272	91.618	1.0273
2	3	29	C	72	109	.	.	88.1995	3.15570	17.9756	0.05	81.5690	94.830	-16.1995
3	6	5	B	109	117	.	5	96.4061	3.03187	47.2421	0.05	90.3076	102.505	12.5939
4	6	5	A	87	100	3	1	93.7690	2.78800	16.4832	0.05	87.8727	99.665	-6.7690
5	6	29	A	85	105	3	3	90.6234	3.25199	11.7170	0.05	83.5189	97.728	-5.6234
7	6	3	A	100	114	1	2	99.9119	3.01901	20.6093	0.05	93.6263	106.198	0.0881
8	6													

etc.

Use of the OUTPM= option

```
                                    S
                                    t
                                    d
p                                   E
a     c                             r
t  v  e  t                          r              A      L       U        R
i  i  n  r       d         P        p              l      o       p        e
e  s  t  e   d   b    c    r        r              p      w       p        s
n  i  r  a   b   p  c f    e        e        D     h      e       e        i
t  t  e  t   p   1  f 1    d        d        F     a      r       r        d

1  6  29 C  86   97  . 1 87.8475 1.31873  95.212  0.05 85.2296 90.4655  -1.8475
2  3  29 C  72  109  . . 91.0744 1.29205  85.384  0.05 88.5056 93.6432 -19.0744
3  6   5 B 109  117  . 5 94.8671 1.95897 189.699  0.05 91.0029 98.7313  14.1329
4  6   5 A  87  100  3 1 91.5816 1.17842  58.874  0.05 89.2235 93.9397  -4.5816
5  6  29 A  85  105  3 3 92.9262 1.13687  55.620  0.05 90.6484 95.2039  -7.9262
7  6   3 A 100  114  1 2 95.3463 1.61199 153.490  0.05 92.1617 98.5309   4.6537
etc.
```

Options relating to model checking Use of these options is illustrated in Section 2.5 and Section 6.3.

The RANDOM statement

This statement can be used to specify random effects and/or random coefficients, β, and the form of their variance matrix, **G**. Several RANDOM statements may be specified, although usually only one will be needed. When no RANDOM statement is included, the results will be the same as those obtained using PROC GLM.

```
RANDOM <random effects>/options;
```

Output options

CL	Requests *t*-type confidence intervals for each random effects estimate.
ALPHA=<p>	Specifies size of confidence interval (default is 0.05 to give 95% confidence intervals).
SOLUTION	Solution for random effects is printed.

Options corresponding to the G and V matrices

G	Prints **G**.
GCORR	Prints correlation matrix corresponding to **G**.
V	Prints **V**.
VCORR	Prints correlation matrix corresponding to **V**.
GDATA=<dataset>	Specifies a fixed **G** matrix.

Example

```
PROC MIXED DATA=a; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR;
RANDOM centre centre*treat/ V SOLUTION;
```

Use of the `V` option

Estimated V Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	78.9267	10.5590			6.4628		
2	10.5590	78.9267			6.4628		
3			78.9267	6.4628			
4			6.4628	78.9267			
5	6.4628	6.4628			78.9267		
6						78.9267	6.4628

This matrix has a dimension equal to the number of observations in the dataset and is hence very large in this example. (If a `SUBJECT` option is used in the `RANDOM` statement, the matrix will relate to only the first subject, otherwise the full variance matrix for all observations is given.) The diagonal terms are equal to the sum of the three variance components. Off-diagonal terms are equal to the sum of the centre and centre·treatment variance components, 10.6, when patients are at the same centre and receive the same treatment; they are equal to the centre variance component, 6.5, when patients are at the same centre but receive different treatments; they and are zero when patients are at different centres (denoted by blank entries).

Use of the `SOLUTION` option

Solution for Random Effects

Effect	treat	centre	Estimate	Std Err Pred	DF	t Value	Pr > |t|
centre		1	0.7966	1.8325	20.7	0.43	0.6683
centre		2	-2.3441	2.1884	14.5	-1.07	0.3016
centre		3	1.9760	2.2547	12.1	0.88	0.3978
centre		4	-0.1429	2.1410	16.4	-0.07	0.9476
centre		5	1.1779	2.1764	15.4	0.54	0.5961
.							
.							
centre		31	-4.7913	1.8470	21.5	-2.59	0.0167
.							
.							

```
centre*treat   A   1   -0.4605   2.2025   2.36   -0.21   0.8511
centre*treat   B   1    1.6900   2.1965   2.38    0.77   0.5106
centre*treat   C   1   -0.7246   2.2093    2.3   -0.33   0.7705
centre*treat   A   2   -0.8330   2.2406   1.56   -0.37   0.7543
.
.
.
```

The `Estimate` and `Std Err Pred` columns give the random effects estimates and their standard errors. The Wald t tests help to determine whether any centre is outlying.

Options for specifying the G matrix structure

`GROUP=<effect>` This causes a separate **G** matrix to be estimated within each group. For example, if patient effects were fitted as random and the `GROUP=treat` option were used, then a separate variance patient component would be estimated for each treatment group. If patients 1 and 3 received treatment A and patients 2 and 4 received treatment B, then the **G** matrix for the first four patients would have the form

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{\mathrm{A}} & 0 & 0 & 0 \\ 0 & \sigma^2_{\mathrm{B}} & 0 & 0 \\ 0 & 0 & \sigma^2_{\mathrm{A}} & 0 \\ 0 & 0 & 0 & \sigma^2_{\mathrm{B}} \end{pmatrix}.$$

However, the statement should be used cautiously when there are many group levels, since a large number of parameters will then need to be estimated.

The **G** matrix is always diagonal in random effects models, and the following options given for the RANDOM statement are not required. However, a non-diagonal **G** matrix is necessary to fit a random coefficients model to allow the intercepts and slopes of the random coefficients to be correlated and also for fitting certain covariance structures in hierarchical repeated measures designs (see Section 8.1). The **G** matrix is blocked by the specified `SUBJECT` effect, and the covariance pattern is specified by the `TYPE` options in a very similar way to the `REPEATED` statement.

```
RANDOM <random effects or coefficients>/
   SUBJECT=<blocking effect> TYPE=<covariance pattern>;
```

Covariance patterns available for the TYPE= option will be listed under the REPEATED statement.

Random coefficients models In these models, the random coefficients on the same subject are assumed to be correlated (e.g. intercepts and slope effects are correlated for each subject in a repeated measures trial). This is achieved using a RANDOM statement of the following form:

```
RANDOM <random coefficients>/SUBJECT=<subject effect> TYPE=UN;
```

For example, to model random coefficients for the effect of time, the code might be

```
RANDOM INT time/SUBJECT=patient TYPE=UN;
```

This fits the random coefficients patient and patient·time. Patient effects are specified by the reserved SAS term INT, which fits an intercept for each patient.

Example

Examples of fitting random coefficients models are given in Section 6.6.

The LSMEANS statement

This statement calculates the least squares mean estimates of specified fixed effects.

```
LSMEANS <fixed effects>/options;
```

Options

CL	Requests *t*-type confidence intervals for least squares means.
ALPHA=<p>	Specifies size of confidence interval (default is 0.05 for a 95% confidence interval).
ADJUST=<type>	Requests adjusted *p*-values for multiple comparisons (see manual for further details).
DIFF	Prints differences between each pair of least squares means.
PDIFF	Prints *p*-values for comparisons between each pair of least squares means, by default. PDIFF=CONTROL compares difference with a control group (first level of the LSMEANS variables, by default).
AT<variables>= <values>	Modifies the covariate value(s) at which LSMEANS are computed. Otherwise they are evaluated at mean values.

Example

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR;
RANDOM centre centre*treat;
LSMEANS treat/ CL COV CORR DIFF PDIFF;
```

Least Squares Means

Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
treat	A	92.3491	1.1233	52	82.21	<.0001	0.05	90.0951	94.6032
treat	B	91.0632	1.1695	51.4	77.86	<.0001	0.05	88.7157	93.4107
treat	C	89.4217	1.1326	58.7	78.96	<.0001	0.05	87.1552	91.6882

Least Squares Means

Effect	treat	Cov1	Cov2	Cov3	Corr1	Corr2	Corr3
treat	A	1.2619	0.2923	0.2770	1.0000	0.2225	0.2177
treat	B	0.2923	1.3678	0.2807	0.2225	1.0000	0.2119
treat	C	0.2770	0.2807	1.2827	0.2177	0.2119	1.0000

The last six columns show covariance and correlation matrices for the least squares means.

Differences of Least Squares Means

Effect	treat	-treat	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
treat	A	B	1.2859	1.4300	23.8	0.90	0.3775	0.05
treat	A	C	2.9274	1.4109	25.6	2.07	0.0482	0.05
treat	B	C	1.6415	1.4453	25.7	1.14	0.2666	0.05

Differences of Least Squares Means

Effect	treat	-treat	Lower	Upper
treat	A	B	-1.6665	4.2384
treat	A	C	0.02511	5.8297
treat	B	C	-1.3314	4.6144

Note that the DF differ markedly between the least squares means and their differences because they are estimated using information from different combinations of error strata.

The `ESTIMATE` statement

This statement calculates a linear function of fixed and/or random effects estimates ($\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$).

```
ESTIMATE '<label>' <fixed effect1> <values>
              <fixed effect2><values>
              ... |
              <random effect 1> <values>
              <random effect 2> <values>
              ...                 / options;
```

Options

CL	Requests *t*-type confidence intervals for estimate.
ALPHA=<p>	Specifies size of confidence interval (default is 0.05 for a 95% CI).
DIVISOR=<n>	Value by which to divide all coefficients so that fractional coefficients can be entered as integer numerators.
E	Shows the effect coefficients used in the estimate (useful to check ordering if interaction effects are used).

When the effect of interest is contained within another fixed interaction effect, the coefficients for this effect are automatically included in the estimate by SAS. For example, the following code for a fixed effects analysis of the multi-centre hypertension data

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat centre centre*treat/ DDFM=KR;
ESTIMATE 'A-B' treat 1 -1 0/ E CL;
```

would cause the following coefficients to be used by SAS for constructing the estimate.

```
            Coefficients for A-B
Effect      treat      centre      Row1
centre                 4
centre                 5
centre                 6
centre                 7
centre                 8
centre                 9
centre                 11
.
.
.
```

```
centre*treat     A     1     0.0385
centre*treat     B     1      -0.04
centre*treat     C     1
centre*treat     A     2     0.0385
centre*treat     B     2      -0.04
centre*treat     C     2
centre*treat     A     3     0.0385
.
.
.
```

The centre·treatment coefficients are equal to $1/n_i$, where n_i is the number of centres at which treatment i is received (26 for treatment A and 25 for treatment B). Because treatments A and B are not received at every centre, this estimate is in fact 'non-estimable', and SAS gives the following output:

```
                          Estimates
                Standard
Label  Estimate     Error DF  t Value  Pr > |t|  Alpha  Lower  Upper
A-B    Non-est          .  .        .         .      .      .      .
```

However, if the containing effect is fitted as random, SAS does not include any coefficients for this containing effect in the estimate. For example, the following code for a random effects analysis of the multi-centre hypertension data

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat/DDFM=KR;
RANDOM centre centre*treat;
ESTIMATE 'A-B' treat 1 -1 0/ E CL;
```

would cause the following coefficients to be used by SAS for constructing the estimate:

```
   Coefficients for A-B
Effect       treat      Row1
Intercept
dbp1
treat        A             1
treat        B            -1
treat        C
```

```
                          Estimates
                Standard
Label Estimate    Error   DF t Value Pr > |t| Alpha   Lower  Upper
A-B     1.2859   1.4300 23.8    0.90   0.3775  0.05 -1.6665 4.2384
```

The ESTIMATE statement can also be used to estimate shrunken random effects by specifying the coefficients of the random effects. For example, to estimate the shrunken difference between the first two treatments within the first two centres, the following ESTIMATE statements can be added to the previous code:

```
ESTIMATE 'C1, A-B' treat 1 -1 0| centre*treat 1 - 1 0/ E CL;
ESTIMATE 'C2, A-B' treat 1 -1 0| centre*treat
          0 0 0 1 -1 0/ CL;
```

These cause the following coefficients to be used for the first estimate

```
          Coefficients for C1, A-B
Effect           treat     centre     Row1
Intercept
dbp1
treat            A                     1
treat            B                    -1
treat            C
centre                     1
centre                     2
centre                     3
.
.
.
centre                    41
centre*treat     A         1           1
centre*treat     B         1          -1
centre*treat     C         1
centre*treat     A         2
centre*treat     B         2
centre*treat     C         2
centre*treat     A         3
.
.
.
```

```
centre*treat          A          41
centre*treat          B          41
centre*treat          C          41
```

and give the following estimates:

```
                                    Estimates

                     Standard
Label     Estimate    Error    DF  t Value  Pr > |t|  Alpha     Lower    Upper
C1, A-B   -0.8646   2.8933  7.35    -0.30    0.7733   0.05   -7.6405   5.9113
C2, A-B    1.6405   3.4054  2.65     0.48    0.6669   0.05  -10.0418  13.3229
```

The CONTRAST statement

This statement can be used to define *F* tests for fixed and random effects. A single or multiple contrast, **C**, can be defined by

$$\mathbf{C} = \mathbf{L}' \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix}.$$

The test may involve either a single contrast (**L** has a single column), for example, to compare two treatments, or several contrasts (**L** has several columns), for example, to test the equality of several treatments. Details of how the *F* statistic is obtained from a contrast are given in Section 2.4.4. When only one row is specified, the *F* test results will give the same *p*-value as the *t* test in an equivalent ESTIMATE statement.

```
CONTRAST '<label>'  <effect1> <effect1 values>
                    <effect2> <effect2 values>
                    ... |
                    <random effect 1> <values>
                    <random effect 2> <values>
                    ...,
                    second row of L,
                    ...
                    / options;
```

As for the ESTIMATE statement, when the effect of interest is contained within another fixed interaction effect, the coefficients for this effect are automatically included in the estimate by SAS.

Options

CHISQ Requests Wald chi-squared test (see Section 2.4.4) in addition to the *F* test. This is an asymptotic test that makes no adjustment for the denominator DF and therefore will be inaccurate for small samples.

E Prints the **L** matrix.

Example

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR;
RANDOM centre centre*treat;
CONTRAST 'TREAT' treat 1 -1 0, treat 1 0 -1/ CHISQ E;
```

Contrasts

Label	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
TREAT	2	25	4.32	2.16	0.1154	0.1364

Use of the E option

Coefficients for TREAT

Effect	treat	Row1	Row2
Intercept			
dbp1			
treat	A	1	1
treat	B	-1	
treat	C		-1

This contrast gives an identical *F* test result to that given under 'Tests of Fixed Effects', seen earlier in this section. Printing the **L** matrix coefficients is particularly helpful when interactions are involved to check the ordering of effects used by SAS. This option may be most useful if identity of a subgroup of treatments requires testing.

The LSMESTIMATE statement

This statement has been introduced in recent versions of SAS. Its features and capabilities are an amalgam of the LSMEANS and ESTIMATE statements and has similar syntax. It facilitates a straightforward way of obtaining customized hypothesis tests among least squares means without actually producing the least squares means. Coefficients only need to be supplied for the least squares means, and these are automatically converted into estimable functions of the parameter estimates. This statement is likely to be very helpful with somewhat complicated experimental designs with a number of hypotheses of interest. It has not been applied to any examples in this book, and those interested should refer to the online SAS help.

The SLICE statement

Another more recent addition to PROC MIXED, this statement provides a method for performing a partitioned analysis of least squares means for interaction terms formed from classification variables. It has similarities to the LSMEANS statement.

The REPEATED statement

The REPEATED statement is used to specify a covariance pattern for the residual matrix, **R**. The repeated measurements should appear in a single variable with the time points specified by another variable. The blocking effect is specified by the SUBJECT variable, and the repeated effect (e.g. time) is the effect used to structure the **R** matrix.

```
REPEATED <repeated effect>/SUBJECT=<blocking effect> options;
```

Note that the blocking effect should have only one observation per repeated effect (otherwise, an infinite covariance matrix will occur). A repeated effect should always be used unless the covariance pattern does not depend on order (e.g. the compound symmetry structure).

Options relating to the R matrix structure

TYPE=<pattern>	This specifies the covariance pattern for the **R** matrix. A wide range of patterns is available, some of which are listed here. These patterns are defined in Section 6.2.
UN	General
AR(1)	First-order autoregressive
CS	Compound symmetry
TOEP	Toeplitz
UN(1)	Heterogeneous uncorrelated
CSH	Heterogeneous compound symmetry
ARH(1)	Heterogeneous first-order autoregressive
TOEPH	Heterogeneous Toeplitz
UN(n)	Banded general, *n* bands
TOEP(n)	Banded Toeplitz, *n* bands
UN@CS UN@AR(1) UN@UN	Direct product. Suitable for trials with blocked repeated measurements (e.g. repeated measurements within visits, see Section 8.1).

Other covariance patterns available are listed in the SAS manual.

GROUP=<effect>	This causes a separate covariance pattern to be estimated for each category of the group effect. This option should be used cautiously if there are a large number of group categories, since it can lead to many extra parameters, particularly if a complex pattern (e.g. general) is used.

Options to print the variance parameters

R=<values>	Prints the **R** matrix for subjects denoted by values (first subject is listed if <values> is omitted).
RCORR=<values>	Prints the correlation matrix corresponding to the **R** matrix for subjects denoted by values.

Other options

LOCAL	Expresses residual variance as $\mathbf{R} + \sigma_r^2\mathbf{I}$, allowing σ_r^2 to be fitted separately from the other parameters of **R**.

Example

Examples of using the REPEATED statement are given in Sections 6.3, 7.7, 8.1.2 and 8.2.

The PARMS statement

This statement can be used to:

- Fix the values of variance components. This can be useful when the covariances are known with a greater accuracy from previous studies than that likely to be obtained in the data analysed.
- Supply initial values for variance components for the iterative procedure.
- Carry out a grid search over a range of variance component values, and take those with the highest likelihood as initial values for the iterative procedure.
- Carry out a grid search over a range of variance component values, and take those with the highest likelihood as the final estimates.

The latter two options can be helpful in a situation where there is the possibility of local (rather than global) solutions. This is most likely to occur when a large number of variance parameters are fitted.

Since a PARMS statement is not needed for most types of mixed models analysis, we have not included details of its use in this section.

The PRIOR statement

A random effects model can be fitted using the Bayesian approach by including a PRIOR statement. Although the use of PROC MIXED for Bayesian analysis has largely been eclipsed by the introduction of PROC MCMC (see Section 9.4), there are situations where a Bayesian model would appear most easily fitted using PROC MIXED. For example, when different residual variances for treatment groups are required. The MIXED procedure uses an independence chain algorithm to sample the posterior distribution, rather than the Markov Chain Monte Carlo algorithm used by PROC MCMC (see SAS documentation for further detail).

Syntax

```
PRIOR <prior distribution>/ options;
```

If no prior distribution is specified in this statement, then the default Jeffreys prior is used (see Section 2.3.4). The only other non-informative prior available is a flat prior, which is specified by

```
PRIOR FLAT;
```

Options

`ALG=<algorithm>`	Specifies sampling algorithm, see Section 2.3.5.
`IC`	Independence chain (default).
`IS`	Importance sampling.
`RS`	Rejection sampling.
`RWC`	Random chain walk.
`NSAMPLE=<n>`	Specifies number of samples, default 1000.
`OUT=<dataset>`	Outputs sampled values to a dataset.

Fixed and random effects samples are produced when the `SOLUTION` option is used in the `MODEL` and `RANDOM` statements, respectively.

Example Use of the `PRIOR` statement is illustrated in Section 8.15.

The `ID` statement

Specifies which variables from the input dataset are to be included in the `OUTP=` and `OUTPM=` datasets from the `MODEL` statement. If you do not specify an `ID` statement, then all variables are included in these datasets. An example using this statement is given in Section 6.6.2.

ID <ID variables>;

The `WEIGHT` statement

Fits a weighted mixed model with weights given by specified variable.

`WEIGHT <weight variable>;`

The `STORE` statement

This statement saves the context and results of the analysis in a binary file format that cannot be modified.

The ODS OUTPUT statement

An ODS OUTPUT statement can be used to read a selected part of a PROC MIXED output into a SAS dataset, thus allowing it to be manipulated by other SAS procedures. Any number of ODS OUTPUT statements can be used in a PROC MIXED analysis.

ODS OUTPUT <table>=<dataset>;

Tables are available corresponding to all PROC MIXED output. Some of the more useful tables are as follows:

Table Name	Corresponding Statement/option
SOLUTIONF	MODEL/SOLUTION
SOLUTIONR	RANDOM/SOLUTION
LSMEANS	LSMEANS
ESTIMATES	ESTIMATE
CONTRASTS	CONTRAST
R<n>	REPEATED/R=<n>
RCORR<n>	REPEATED/RCORR=<n>

A PROC CONTENTS statement is suggested to find out the variable names created within each dataset (or alternatively they are listed in the SAS documentation). Other tables available are listed in the SAS documentation. The following statement can be used to suppress the printing of output from chosen tables; this is particularly useful for statements or options that can give lengthy output:

```
ODS LISTING EXCLUDE <table1> <table2> etc;
```

Example

```
PROC MIXED; CLASS centre treat;
MODEL dbp = dbp1 treat/ DDFM=KR;
RANDOM centre centre*treat;
LSMEANS treat/ CL;
ESTIMATE 'A-B' treat 1 -1 0/ CL;
ESTIMATE 'A-C' treat 1 0 -1/ CL;
ESTIMATE 'B-C' treat 0 1 -1/ CL;
CONTRAST 'TREAT' treat 1 -1 0, treat 1 0 -1;
ODS LISTING EXCLUDE LSMEANS ESTIMATES CONTRASTS;
ODS OUTPUT LSMEANS=lsmeans;
ODS OUTPUT ESTIMATES=estimate;
ODS OUTPUT CONTRASTS=contrast;
PROC PRINT DATA=lsmeans;
```

```
PROC PRINT DATA=estimate;
PROC PRINT DATA=contrast;

Obs Effect treat Estimate StdErr   DF tValue  Probt Alpha   Lower   Upper
1   treat    A    93.5545 0.8349 66.2 112.05 <.0001  0.05 91.8876 95.2214
2   treat    B    92.0060 0.8519 66.3 108.00 <.0001  0.05 90.3053 93.7067
3   treat    C    91.0582 0.8067   68 112.88 <.0001  0.05 89.4485 92.6679

Obs  Label  Estimate  StdErr    DF  tValue   Probt  Alpha    Lower   Upper
1      A-B    1.5485  1.0256  39.2    1.51  0.1391   0.05  -0.5256  3.6225
2      A-C    2.4962  0.9991  40.8    2.50  0.0166   0.05   0.4782  4.5143
3      B-C    0.9478  1.0127  40.8    0.94  0.3549   0.05  -1.0978  2.9933

                      Num       Den
Obs       Label       DF        DF        FValue       ProbF
1         TREAT        2      40.3         3.17        0.0528
```

The MAKE statement

This statement was available in earlier versions of SAS but has now been superseded by the ODS OUTPUT statement.

9.3 Using SAS to fit mixed models to non-normal data

A major advance in SAS in recent years has been the development of PROC GLIMMIX. This is a very flexible procedure that allows the fitting of all types of generalized linear mixed models (see Chapter 3). It will also fit random effects models to categorical data (see Chapter 4). In this section, we will only be introducing some of the basic syntax but it has additional attractive features. These include the capability to define variables with assignment statements within the procedure and to introduce smoothing effects to capture trends over time. Users can specify their own link functions and variance functions. Multivariate (rather than multivariable) models can be specified.

Most models that can be fitted using PROC MIXED can also be fitted in PROC GLIMMIX (through the use of the identity link function) so, as SAS themselves state, the MIXED procedure can be thought of as subsumed by the GLIMMIX procedure in some sense. Therefore, there is a strong impetus for regular users of mixed models to become more familiar with the GLIMMIX procedure than is possible from the information we supply in this section. The online help is very useful. PROC NLMIXED is also a possibility for random effects models but this has offered no advantages over GLIMMIX for the examples we have considered. PROC

GENMOD can also be used for covariance pattern models, using a different fitting method. If a covariance pattern model is required for use with categorical data, the trick of specifying the 'time' variable as random will generate a compound symmetry structure and allow the GLIMMIX procedure to be used. Otherwise, such an analysis cannot be performed at present using SAS. We have used, in this book, a macro written by Stuart Lipsitz (Lipsitz *et al.*, 1994) that works by iteratively calling PROC LOGISTIC. This is available through the website www.wiley.com/go/brown/applied_mixed.

In the following sections, we give the basic syntax required to fit GLMMs and some mixed categorical models using PROC GLIMMIX and PROC GENMOD.

9.3.1 PROC GLIMMIX

This procedure can be used to fit all types of GLMMs and additionally random effects models for categorical data. There is a range of optional fitting methods with the default method based on pseudo-likelihood (see Section 3.2). Many of the statements in GLIMMIX are in common with the MIXED procedure though a few extra statements are also available. A wide range of options is available as well, giving the expert user considerable control over the analysis. In this section, we will only provide syntax required to produce basic analyses. Examples using PROC GLIMMIX can be found in Sections 3.4, 4.5, 5.7, 6.4, 7.9, 8.5, 8.10, 8.11, 8.12, 8.13 and 8.16.

Syntax

The BY, CLASS, CONTRAST, ESTIMATE, ID, LSMEANS, LSMESTIMATE, PARMS, SLICE, STORE and WEIGHT statements are all similar to the corresponding statements in PROC MIXED and will not be considered further in this section. We will look briefly at the PROC GLIMMIX and MODEL statements and in slightly more detail at the RANDOM statement, which functions quite differently from the equivalent statement in the MIXED procedure.

The PROC GLIMMIX statement

There are many options available relating mainly to methods used to fit the model and the display of additional results and plots. In this section, we only list the EMPIRICAL option, which may help to overcome downward biases in the fixed effects variance estimates. Although the standard empirical variance is known to be biased, the latter four choices attempt to adjust for this bias. More detail on their calculation and appropriateness for different situations is given in the PROC GLIMMIX documentation.

EMPIRICAL= <type of empirical variance>	Requests variances and standard errors to be calculated using the empirical variance.
CLASSICAL	Standard empirical variance (default if no type specified).
DF	Bias-adjusted empirical variance.
ROOT	Bias-adjusted empirical variance.
FIRORES	Bias-adjusted empirical variance.
FIROREEQ	Bias-adjusted empirical variance.

The MODEL statement

This statement is used to specify the dependent variable and the fixed effects in the model. The dependent variable is specified as a single value for most data types or as a numerator with denominator for binary data (e.g. r/n). Bernoulli and binomial distributions are available for modelling binary data. We have used the binomial distribution for all examples involving binary data so that zero is used for the reference category. (If the data are recorded as zeros and ones and the BINARY option is used, then one is used as the reference category, and the signs of the fixed effects are reversed.) We suggest the use of DDFM=KR or DDFM=KR2 (available in SAS/STAT 12.1 or later) in most of the examples, as it helps to adjust for any bias in fixed effects standard errors; however, we have found occasionally that the use of these options has caused the procedure to fail. The DF2 option is based on a revised adjustment by Kenward and Roger (Kenward and Roger, 2009), which is expected to further reduce bias in fixed effects standard errors in certain types of covariance pattern model (see Section 6.2.4).

Ordinal data can be analysed using the DIST=MULT option. This option can also be used to analyse unordered categorical data; however, the response variable then needs to be specified as a GROUP in the RANDOM statement and the BYCAT option should be included in any ESTIMATE or CONTRAST statements used.

Options

DIST=<distribution>	This option specifies the distribution:
BINARY	Bernoulli (No denominator required for dependent variable. The highest category is taken as the reference category.)
B(INOMIAL)	Binomial (Denominator can be given with dependent variable. If omitted, it is assumed to be 1.)
P	Poisson
MULT	Multinomial (By default data are assumed ordinal.)
LINK=<link function>	The default link is often adequate, but with unordered multinomial data, the generalized logit link is needed.
GLOGIT	
OFFSET=<variable>	Sets the offset for a Poisson distribution.
CHISQ	Carries out chi-squared tests in addition to the F tests of fixed effects.

SOLUTION	Prints solution for fixed effects.
CL	Requests *t*-type confidence intervals for fixed effects given by SOLUTION option
ALPHA=<p>	Specifies size of confidence intervals (default 0.05)
ODDSRATIO	Displays odds ratios and confidence limits
DDFM=<DF type>	Selects DF type for *t* tests and for the denominator DF in *F* tests of fixed effects. All of the options given in Section 9.2 for the MIXED procedure also apply to GLIMMIX. In addition, the KR2 option applies the revised Kenward-Roger adjustment (Kenward and Roger, 2009) in SAS/STAT 12.1 or later.

The RANDOM statement

RANDOM statements can be used to specify random effects, covariance patterns and a dispersion parameter. Thus, the REPEATED statement used in PROC MIXED and PROC GENMOD is now subsumed into the RANDOM statement. The SAS documentation refers to random effects as 'G-side' effects and covariance patterns as 'R-side' effects. Several RANDOM statements may be used within the same analysis.

Dispersion parameters By default, the dispersion parameter is fixed at one. It can be fitted in the model by including the SAS fixed variable _RESIDUAL_ as a random effect.

```
RANDOM _RESIDUAL_;
```

Random effects Random effects are fitted just as in PROC MIXED by simply listing the effects in the main part of the statement.

```
RANDOM <random effect variables>/ options;
```

Covariance pattern models For covariance pattern models, the RANDOM statement has a similar syntax to the REPEATED statement in PROC MIXED except that it is necessary to include the RSIDE or RESIDUAL option:

```
RANDOM <time effect>/SUBJECT=<subject effect> RSIDE options;
```

Options corresponding to output

CL	Requests *t*-type confidence intervals for each random effect estimate.
ALPHA=<p>	Specifies size of confidence interval (default is 0.05).
G	Prints **G**.

`SOLUTION`	Prints solution for random effects.
`V=<subjects>`	Prints variances in **V** for subjects listed (given for subject, one if no subjects listed).
`VCORR=<subjects>`	Prints correlation matrix corresponding to **V** for subjects listed (for subject, one if no subjects listed).

Options corresponding to model specification

`GROUP=<effect>`	A different set of covariance parameters is used for each category of the group effect.
`RSIDE or RESIDUAL`	Causes a covariance pattern model to be fitted.
`TYPE=<pattern>`	This specifies the covariance pattern. A wide range of patterns is available, some of which are listed here. These patterns are defined in Section 6.2
`AR(1)`	First-order autoregressive,
`CS`	Compound symmetry,
`TOEP`	Toeplitz,
`TOEP(n)`	Banded Toeplitz, *n* bands,
`UN or UNR`	General,
`UN(1) or UNR(1)`	Heterogeneous uncorrelated,
`UN(n) or UNR(n)`	Banded general, *n* bands,

The last three covariance patterns can be achieved using either the `UN` or `UNR` options. The `UN` option parameterises the model directly in terms of variance and covariance parameters, while the `UNR` option uses variances and correlation parameters providing a clearer picture of the correlations between time points. Many more covariance patterns for the `TYPE` option are detailed in the procedure documentation. An example using `PROC GLIMMIX` to fit covariance patterns is given in Section 6.4.

The `ODS OUTPUT` statement

Tables are available corresponding to all `PROC GLIMMIX` output. Several model diagnostic graphs are also available using `ODS` to check residuals; however, for many examples where data are in Bernoulli form, this is not necessary (see Section 3.3.10). No graphs are available for checking random effects, and it is still necessary to construct these manually (see `SAS` code in Section 3.4).

9.3.2 `PROC GENMOD`

This procedure can be used to fit fixed effects GLMs, ordinal logistic regression models and GLMMs with covariance patterns. Covariance pattern models are

fitted using a method known as generalised estimating equations (GEEs, see Liang and Zeger, 1986); however, these are not available for ordinal data. The empirical estimator is produced by default for estimating fixed effects standard errors (see Section 2.4.3), but the model-based estimator can additionally be requested. Compared with PROC GLIMMIX (see Section 9.3.1), PROC GENMOD has the disadvantage that the tests of significance it produces for GLMMs are asymptotic, whereas the Kenward–Roger option can be used in PROC GLIMMIX. We give basic details of the SAS code required to fit covariance pattern models to binomial and Poisson data. More detail can be found in the *online* documentation.

Basic syntax

Many statements that are common to the MIXED and GLIMMIX procedures also appear in GENMOD. These include BY, CLASS, CONTRAST, ESTIMATE, LSMEANS, LSMESTIMATE, SLICE, STORE and WEIGHT. In this section, we will only highlight key statements that differ noticeably in PROC GENMOD.

PROC GENMOD statement

This has the structure:

PROC GENMOD options;

It invokes the GENMOD procedure. Useful options include plots for a variety of residual and influence statistics.

MODEL statement and options

```
MODEL <y variable(s)> = <fixed effects>/ <options>;
```

The y variable is specified as a single value for most data types or as a numerator with denominator for binomial data (e.g. r/n).

Useful Options

DIST=<distribution>	This option specifies the distribution:
B	Binomial.
P	Poisson.
MULT	Multinomial (but note that the REPEATED statement cannot be used to fit covariance pattern models with this data type).
OFFSET=<variable>	Sets the offset for a Poisson distribution.
TYPE3, WALD	Used together these options produce composite chi-squared tests of fixed effects.

`REPEATED` statement and options

```
REPEATED SUBJECT=<subject effect>/ <options>;
```

The subject effect is the blocking effect for the covariance pattern (see Section 6.2) and has the same role as the subject effect in the `REPEATED` statement of `PROC MIXED`.

Useful Options

`TYPE=<covariance pattern>`	This option specifies the covariance pattern:
`CS`	Compound symmetry.
`AR(1)`	First-order autoregressive.
`MDEP(n)`	Toeplitz (Notice that unlike the TOEP option in PROC MIXED, *n*, the number of bands of parameters, needs to be specified otherwise only one band is fitted.)
`UN`	General.
`WITHIN=<fixed effect>`	This defines the effect to be used for structuring the covariance pattern. It has the same role as the main effect used in the `REPEATED` statement in `PROC MIXED`. In repeated measures analyses, it is usually time or visit.
`CORRW`	Prints the correlation parameters in the **P** matrix ($\mathbf{R} = \phi \mathbf{A}^{1/2}\mathbf{B}^{1/2}\mathbf{P}\mathbf{B}^{1/2}\mathbf{A}^{1/2}$, see Section 3.2.1).
`MODELSE`	Prints fixed effects estimates with model-based standard errors in addition to the (default) empirical estimators.

The `OUTPUT` statement

```
OUTPUT OUT=<dataset> P=<predicted values variable>
       RESCHI=<Pearson residuals variable>;
```

This statement outputs the residuals and prediction values to a `SAS` dataset. There are also a large number of other keywords that can be specified, as well as `P` and `RESCHI` to output residual and influence statistics.

The `ODS OUTPUT` statement

```
ODS OUTPUT <TABLE>=<dataset>;
```

Tables are available corresponding to all `PROC GENMOD` output. Some of the more useful tables are now listed:

Table name	Corresponding statement/option
CONTRASTS	CONTRAST
ESTIMATES	ESTIMATE
GEEMODPEST	REPEATED
LSMEANS	LSMEANS
LSMEANDIFFS	LSMEANS/ DIFF
PARAMETERESTIMATES	MODEL

Use of the GENMOD procedure is illustrated in Sections 3.4, 7.8 and 8.16.2.

9.4 PROC MCMC

This procedure was introduced into SAS Version 9.3 and was further updated in SAS/STAT 12.1. It may be used to fit mixed models to both normal and non-normal data using the Bayesian approach. It uses the Metropolis algorithm to simulate the posterior distribution. In this section information is provided on its basic use for fitting mixed models. However, it is a flexible procedure that may be used to fit a wide variety of models to data with different underlying distributions. Use of the procedure including SAS output is illustrated in the examples in Sections 2.5, 3.4 and 4.5.

Syntax

```
PROC MCMC options;
ODS SELECT <output>;
PARMS <parameters followed by starting values>;
PRIOR <parameter> ~ <distribution>;
RANDOM <random parameter> ~ <distribution>
   SUBJECT=<random effect>;
<Model defined as an equation>;
```

The syntax is more easily understood with the aid of an example. The following syntax was used in Section 2.5 to analyse the multicentre hypertension example and fit a model with baseline and treatment effects as fixed and centre and centre·treatment effects as random (Model 5).

```
PROC MCMC DATA=C OUTPOST=post2 NMC=100000 THIN=5 SEED=7893;
ODS SELECT PARAMETERS REPARAMETERS POSTSUMMARIES
    POSTINTERVALS;
PARMS alpha0 alpha1 alpha2 alpha3 v1 v2 v3;
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
```

```
RANDOM b_centre ~ NORMAL(0, VAR=v2) SUBJECT=centre
  MONITOR=(CENTRE);
RANDOM b_ct ~ NORMAL(0, VAR=v3) SUBJECT=centre_treat
  MONITOR=(CENTRE);
mu = alpha0 + alpha1*dbp1 + alpha2*treata + alpha3*treatb
   + b_centre + b_ct;
MODEL dbp ~ NORMAL(mu, VAR = v1);
```

Each of these statements is now described. Although this is a model for normal data, we will indicate how the syntax may be adapted to fit a GLMM or a mixed model for ordinal data.

PROC MCMC statement

```
PROC MCMC DATA=C OUTPOST=post NMC=100000 THIN=5 NBI=1000
   SEED=7893;
```

The OUTPUT option specifies the SAS dataset for output of the simulated samples of parameters, here as 'post'.

The NMC option specifies the number of samples to be taken. The THIN option specifies what proportion of these are used to calculate parameter estimates. Here every one in five samples will be used. Higher thinning makes adjacent samples less correlated but means more samples need to be taken for a given size of sample. The NBI option specifies the number of initial samples to be discounted and is used when the samples need a while to settle down. These may be called 'burn in' samples. The eventual number of samples used to form the posterior distribution for the parameters will be (NMC-NBI)/THIN. The use of some of these options is considered in Sections 2.5 and 3.5.

The SEED option gives the seed for the random process. It can sometimes be helpful to repeat an analysis with different seeds and check whether the results are similar. Large odd numbers are recommended.

ODS SELECT statement

This statement selects the output to be provided. Note that use of this statement will cause some of the default output (obtained without use of the statement) to be omitted.

```
ODS SELECT POSTSUMMARIES POSTINTERVALS;
```

Options

POSTSUMMARIES – Gives the mean, SD, 25% 50% (median) and 75% centiles for each parameter

POSTINTERVALS – Lists the probability intervals for each parameter for specified alpha (default 0.05), as an equal tail interval and an HPD interval

Statement to define the model

The following statement specifies the model equation:

```
mu = alpha0 + alpha1*dbp1 + alpha2*treata + alpha3*treatb
   + b_centre + b_ct;
```

Unlike PROC MIXED and PROC GLIMMIX, the model equation needs to be specified in full with variable names for the model parameters to be sampled, as well as including variables for the observed values corresponding to the parameters. alpha0, alpha1, alpha2 and alpha3 will be the sampled parameters for the fixed effects: intercept, baseline and two treatment effects. Together they form the $\boldsymbol{\alpha}$ vector of fixed effects parameters used to define the generic mixed model defined in Section 2.1. The variables dpb1, treata and treatb are the corresponding observed fixed effects. These are the values that form the $\mathbf{X}$ matrix in the generic mixed model defined in Section 2.1. Note that it is not possible to specify the three treatments using a CLASS statement in this procedure, and treata and treatb are binary variables indicating the presence of treatments A and B. Treatment C is indicated by the absence of treatments A and B in treata and treatb. Convenient SAS code to set up such binary variables is given at the end of the section. The variables b_centre and b_ct are the random effect parameters and together form the $\boldsymbol{\beta}$ vector in the generic mixed model defined in Section 2.1. Note that, in contrast to the fixed categorical effect (treatment), we do not need to set up dummy variables for the levels of our random effect variables.

MODEL statement

This statement specifies the distribution of the outcome y variable, post-treatment dbp. Following our assumption for the mixed model for normally distributed data, dbp is assumed to have a distribution with mean mu and the residual variance, v1.

```
MODEL dbp ~ NORMAL(mu, VAR = v1);
```

For GLMMs, an alternative distribution needs to be used. For example, a binary variable cf (cold feet, Y/N) also recorded in this study is analysed in Section 3.4 using the following two statements:

```
mu  = alpha0 + alpha1*cf1 + alpha2*treata + alpha3*treatb
    + b_centre + b_ct;
expected = LOGISTIC(mu);
MODEL cf ~ BINARY(expected);
```

To fit a mixed model to ordinal data, multiple statements defining the model and the multinomial probabilities are required before defining the multinomial distribution. In Section 4.5, an ordinal variable `cfm` (cold feet on a scale of 1–5) is analysed by including the following statements:

```
mu1 = alpha01 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
mu2 = alpha02 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
mu3 = alpha03 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
mu4 = alpha04 + alpha11*cf11 + alpha12*cf12 + alpha13*cf13
      + alpha14*cf14 + alpha2*treata + alpha3*treatb
      + b_centre + b_ct;
p1  = LOGISTIC(mu1);
p2  = LOGISTIC(mu2)-LOGISTIC(mu1);
p3  = LOGISTIC(mu3)-LOGISTIC(mu2);
p4  = LOGISTIC(mu4)-LOGISTIC(mu3);
p5  = 1-LOGISTIC(mu4);
ARRAY p[5] p1 p2 p3 p4 p5;
MODEL cfm ~ MULTINOM(p);
```

`PARMS` statement

This statement provides variable names for parameters for the fixed effects and the variance components. Thus, unlike `PROC MIXED`, it is necessary to provide separate variable names for the parameters. Values for the first sample can optionally be specified after each parameter name. When this is not done, values will be obtained from their prior distributions, either by drawing a random sample or by taking the mode. In the following `PARMS` statement initial parameter values are provided.

```
PARMS alpha0 0 alpha1 0 alpha2 0 alpha3 0 v1 1 v2 1 v3 1;
```

`PRIOR` statements

The first prior statement defines non-informative priors for the fixed effects parameters using normal distributions with a very large variance. Use of ':' causes all parameters with names starting with 'alpha' to be given the specified normal distribution.

```
PRIOR alpha: ~ NORMAL(0, VAR = 10000);
```

The second prior statement defines non-informative priors for the variance component and residual parameters with inverse gamma distributions with small parameters. All parameters with names starting with 'v' are given the specified inverse gamma distribution.

```
PRIOR v: ~ IGAMMA(0.01, SCALE = 0.01);
```

RANDOM statement

Random statements specify the distributions for the random effects (centre and centre·treatment) with zero means and variance equal to the corresponding variance components. The MONITOR option causes sampled values for individual centres (and each centre·treatment combination) to be included in the parameter summary tables in addition to the fixed effects and variance component parameters.

```
RANDOM b_centre ~ NORMAL(0, VAR=v2) SUBJECT=centre
   MONITOR=(b_centre);

RANDOM b_ct ~ NORMAL(0, VAR=v3) SUBJECT=centre_treat
   MONITOR=(b_ct);
```

Defining categorical variables and interactions

As we noted previously, the MCMC procedure has no CLASS statement and cannot fit categorical fixed effects directly. If a categorical effect has c categories, then c-1 binary variables need to be specified. In addition, it is not possible to specify interactions between variables using the * symbol.

The following code first defines a variable, centre_treat, denoting the interaction between treatment and centre effects. Then, the TRANSREG procedure is used to define two dummy binary variables denoting the presence or absence of treatments A and B. The ZERO=LAST option specifies that the last treatment category (C) is not parameterised.

```
DATA b; SET a; BY patient;
IF LAST.patient;
centre_treat=COMPRESS(centre||treat);

PROC TRANSREG DATA=b DESIGN;
MODEL CLASS(treat / ZERO=LAST);
ID dbp dbp1 centre centre_treat;
OUTPUT OUT=c(DROP=_: INT:); RUN;
```

The first few observations from the resulting dataset ('c') are printed as follows showing the new variables, treatA and treatB.

treat A	treat B	treat	dbp	dbp1	centre	centre_ treat
0	0	C	86	97	29	29C
0	0	C	72	109	29	29C
0	1	B	109	117	5	5B
1	0	A	87	100	5	5A
1	0	A	85	105	29	29A
1	0	A	100	114	3	3A
0	1	B	80	105	3	3B
0	1	B	90	100	3	3B
1	0	A	100	102	3	3A
0	0	C	94	105	3	3C

Glossary

Terms referred to frequently within the book are defined below. Because mixed models have been developed for use in several areas there is sometimes ambiguity in the meanings of the same terms. Thus, the definitions given here may not always agree with those in other sources. They will, however, be adhered to within this book.

Balance. See Section 1.6.

Bernoulli form. Binary data specified as observations of zero and one. This form must be used if covariates at the residual level are modelled; for example, if a 0/1 variable is recorded pre- and post-treatment.

Binomial form. Binary data specified as frequencies with denominators.

Blocking effect. An effect used to block the variance matrix. A covariance pattern is specified for data within the categories of the blocking effect. For example, in a repeated measures trial the covariance matrix is blocked by patient effects (see Section 6.2).

Containment. An effect (A) is described as contained within another effect (B) either if (B) is nested within (A), or if (B) is a random interaction term containing (A) (e.g. $B = A \cdot C$) (see Section 1.6).

Containment level. This is defined as the residual error level unless the fixed effect is contained within another effect (see Section 1.6).

Contrast. A contrast defines a specific linear comparison of fixed effects categories. The objective is usually to determine whether the categories differ significantly by defining an appropriate test statistic from the contrast. Pairwise

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

differences between two categories (e.g. treatments) are a common type of contrast. Multiple contrasts can be used to test the overall equality of a set of contrasts; for example, to test the overall equality of a group of treatments. Contrasts are defined in more detail in Section 2.4.4.

Crossed effects. An effect (A) is described as crossed with another effect (B) when different categories of (A) occur within each category of (B). For example, in a cross-over trial, different treatment categories occur within patients, so treatment effects are crossed with patient effects.

Effect1·effect2. Interaction effect, for example centre·treatment denotes the interaction effect between centres and treatments.

Error stratum. See Section 1.6.

Fixed effects model. A model fitting only fixed effects.

Fixed or random effects category. An effect level; for example, if three treatments are fitted, then the treatment effect has three categories.

Full residuals. Residuals calculated by deducting only the fixed effects, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}$ (ordinary residuals are defined when both the fixed and random effects are deducted, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}} - \mathbf{Z}\hat{\boldsymbol{\beta}}$).

Generalised linear mixed model (GLMM). A mixed model for non-normal data that assumes residuals have variances proportional to those specified by a chosen distribution from the exponential family.

Least squares mean. Mean estimate for an effect category, adjusted for other effects in the model.

Mean predicted values. Values predicted for individual observations based on fixed effects only as $\mathbf{X}\hat{\boldsymbol{\alpha}}$.

Mixed model. The description 'mixed' was originally used to describe a model fitting both fixed effects and random effects. Here, we use it more widely to encompass random effects models, random coefficients models and covariance pattern models (see Section 1.1).

Nested. An effect (A) is defined as nested within another effect (B) if B takes the same value for all observations within every category of A. For example, in repeated measures trials each patient receives only one treatment and patients are nested within treatments. However, in cross-over trials treatments vary between periods and patients are not nested within treatments. Note that nesting is the reverse of containment: if A is nested within B, then B can be described as contained within A.

Normal mixed model. A mixed model for normally distributed data.

Predicted values. Values predicted for individual observations based on both fixed and random effects as $\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}}$.

Random effects model. A model fitting fixed and random effects. The residuals are assumed independent.

Random effects predictions. These are effectively random effects estimates. However, since they are not formally estimated within the model they are sometimes referred to as predictions. They are shrunken compared with their fixed effects counterparts (see Section 1.3).

Residuals. Residuals from the residual error strata, equal to $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}} - \mathbf{Z}\hat{\boldsymbol{\beta}}$.

SAS® Statistical Analysis System. The most commonly used statistical analysis package within the pharmaceutical industry.

Uniform fixed/random effects category. See Sections 3.3.3 and 3.3.4.

Variance component. Additional variation due specifically to a random effect.

Variance matrix. Matrix of variance and covariance terms.

References

Akaike, H (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Aldhous, MC, Prescott, RJ, Roberts, S, Samuel, K, Waterfall, M and Satsangi, J (2008) Does nicotine influence cytokine profile and subsequent cell cycling/apoptotic responses in inflammatory bowel disease? *Inflammatory Bowel Disease*, 14, 1469–1482.

Alexander, FE, Anderson, TJ, Brown, HK, Forrest, APM, Hepburn, W, Kirkpatrick, AE, Muir, BB, Prescott, RJ and Smith, A (1999) 14 years of follow-up from the Edinburgh randomized trial of breast-cancer screening, *Lancet*, 353, 1903–1908.

Box, GEP and Tiao, GC (1973) *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, Inc., New York.

Bozdogan, H (1987) Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, 52, 345–370.

Breslow, NE and Clayton, DG (1993) Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, 88, 9–25.

Brooke, H, Gibson, A, Tappin, D and Brown, H (1997) Case–control study of sudden infant death syndrome in Scotland, 1992–1995, *British Medical Journal*, 314, 1516–1520.

Brown, HK and Kempton, RA (1994) The application of REML in clinical trials, *Statistics in Medicine*, 16, 1601–1617.

Browne, WJ, Lahi ML and Parker RMA (2009) *A guide to sample size calculation for random effect models via simulation and the MLPowSim Software Package* (ISBN 0-903024-96-9) www.bristol.ac.uk/cmm/software/mlpowsim/mlpowsim-manual.pdf, Bristol.

Carlin, BP and Louis, TA (2008) *Bayesian Methods for Data Analysis* (third edition), Chapman and Hall/CRC Press, Boca Raton, FL.

Carstairs, V and Morris, R (1991) *Deprivation and Health in Scotland*, Aberdeen University Press, Aberdeen.

Clayton, D and Hills, M (1993) *Statistical Models in Epidemiology*, Oxford University Press, Oxford.

Collett, D (1991) *Modelling Binary Data*, Chapman and Hall, London.

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

Cunningham, S, Logan, C, Lockerbie, L, Dunn, MJ, McMurray, A and Prescott, RJ (2008) Effect of an integrated care pathway on acute asthma/wheeze in children attending hospital: cluster randomized trial, *Journal of Pediatrics*, 152, 315–320.

Demidenko, E (2013) *Mixed Models: Theory and Applications with R* (second edition), John Wiley & Sons, Inc., Hoboken, NJ.

Diggle, PJ (1989) Testing for random dropouts in repeated measurement data, *Biometrics*, 43, 1255–1258.

Diggle, P and Kenward, MG (1994) Informative drop-out in longitudinal data analysis, *Journal of the Royal Statistical Society, Series C*, 43, 49–93.

Diggle, PJ, Liang, K-Y and Zeger, SL (1994) *Analysis of Longitudinal Data*, Oxford University Press, Oxford.

Dixon, JM, Ravisekar, O, Cunningham, M, Anderson, EDC, Anderson, TJ and Brown, HK (1996) Factors affecting outcome of patients with impalpable breast cancers detected by breast screening, *British Journal of Surgery*, 83, 997–1001.

Elston, DA (1998) Estimation of denominator degrees of freedom of *F*-distributions for assessing Wald statistics for fixed-effect factors in unbalanced mixed models, *Biometrics*, 54, 1085–1096.

Fisher, RA (1925) *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.

Giesbrecht, FG and Burns, JC (1985) Two-stage analysis based on a mixed model: large-sample asymptotic theory and small-sample simulation results, *Biometrics*, 41, 477–486.

Gilks, WR, Richardson, S and Spiegelhalter, DJ (1995) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Goldstein, H (1989) Restricted unbiased iterative generalized least-squares, *Biometrika*, 76(3), 622–623.

Goldstein, H (2003) *Multilevel Statistical Models* (third edition), Edward Arnold, London.

Goldstein, H (2010) *Multilevel Statistical Models* (fourth edition), Wiley, Chichester.

Greenhouse, SW and Geisser, S (1959) On methods in analysis of profile data, *Psychometrika*, 32, 95–112.

Hall, S, Prescott, RJ, Hallam, RJ, Dixon, S, Harvey, RE and Ball, SG (1991) A comparative study of Carvedilol, slow release Nifedipine and Atenolol in the management of essential hypertension, *Journal of Pharmacology*, 18(4), S36–S38.

Hannan, EJ and Quinn, BG (1979) The determination of an autoregression, *Journal of the Royal Statistical Society, Series B*, 41, 190–195.

Hartley, HO and Rao, JNK (1967) Maximum likelihood estimation for the mixed analysis of variance model, *Biometrika*, 54, 93–108.

Hay, CRM, Ludlam, CA, Lowe, GDO, Mayne, EE, Lee, RJ, Prescott, RJ and Lee, CA (1998) The effect of monoclonal or ion-exchange purified factor VIII concentrate on HIV disease progression: A prospective cohort comparison, *British Journal of Haematology*, 101, 632–637.

Healy, MJR (2000) *Matrices for Statistics* (second edition), Oxford University Press, Oxford.

Hedeker, D and Gibbons, RD (1994) A random effects ordinal regression model for multilevel analysis, *Biometrics*, 50, 933–944.

Henderson, CR (1953) Estimation of variance and covariance components, *Biometrics*, 9, 226–252.

Huber, P and Ronchetti, EM (2009) *Robust Statistics* (second edition), John Wiley & Sons, Inc., New York.

Hunter, KR, Stern, GM, Laurence, DT and Armitage, P (1970) Amantadine in Parkinsonism, *Lancet*, i, 1127–1129.

Jeffreys, H (1961) *Theory of Probability* (third edition), Clarendon Press, Oxford.

Jennrich, RI and Schluchter, MD (1986) Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, 42, 805–820.

Jones, B and Kenward, MG (1989) *The Analysis of Crossover Trials*, Chapman and Hall, London.

Kacker, RN and Harville, DA (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models, *Journal of the American Statistical Association*, 79, 388, 853–862.

Kempthorne, O (1952) *The Design and Analysis of Experiments*, John Wiley & Sons, Inc., New York.

Kenward, MG and Roger, JH (1997) Small sample inference for fixed effects from restricted maximum likelihood, *Biometrics*, 53, 983–997.

Kenward, MG and Roger, JH (2009) An improved approximation to the precision of fixed effects from restricted maximum likelihood, *Computational Statistics & Data Analysis*, 53, 2583–2595.

Kenward, MG, Lesaffre, E and Molenberghs, G (1994) An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random, *Biometrics*, 50, 945–953.

Laird, NM and Ware, JH (1982) Random effects models for longitudinal data, *Biometrics*, 38, 963–974.

Lee, Y and Nelder, JA (1996) Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society, Series B*, 58, 619–678.

Liang, KY and Zeger, SL (1986) Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13–22.

Liang, KL, Zeger, SL and Qaqish, B (1992) Multivariate regression analyses for categorical data, *Journal of the Royal Statistical Society, Series B*, 54, 3–40.

Lipsitz, SR, Kyungmann, K and Zhao, L (1994) Analysis of repeated categorical data using generalized estimating equations, *Statistics in Medicine*, 13, 1149–1163.

Littell, RC, Milliken, GA, Stroup, WW and Wolfinger, RD (2006) *SAS System for Mixed Models* (second edition), SAS Institute, Cary, NC.

Little, RJA and Rubin, DB (2002) *Statistical Analysis with Missing Data* (second edition), John Wiley & Sons, Inc., New York.

Long, JS and Ervin, LH (2000) Using heteroscedasticity consistent standard errors in the linear regression model, *The American Statistician*, 54, 217–224.

Longford, NT (1993) *Random Coefficients Models*, Oxford University Press, Oxford.

McCulloch, CE, Searle, SR and Neuhaus, JM (2008) *Generalized, Linear, and Mixed Model*, 2nd edition. Wiley Series in Probability and Statistics.

McCullagh, P (1980) Regression models for ordinal data, *Journal of the Royal Statistical Society, Series B*, 42, 109–142.

McCullagh, P and Nelder, JA (1989) *Generalised Linear Models* (second edition), Chapman and Hall, London.

McLean, RA and Sanders, WL (1988) Approximating degrees of freedom for standard errors in mixed linear models, *Proceedings of the Statistical Computing Section, American Statistical Association*, New Orleans, 50–59.

Mead, R (1988) *The Design of Experiments*, Cambridge University Press, Cambridge.

Meyer, K (1986) Restricted maximum likelihood to estimate genetic parameters in practice. In GE Dickerson and RK Johnson (eds), *Proceedings of the 3rd World Congress on Genetics applied to Animal Production Vol XII*, University of Nebraska, 454–459.

Molenberghs, G and Kenward, M (2007) *Missing Data in Clinical Studies*, Wiley and Sons Ltd, Statistics in Practice Series.

Nabugoomu, F and Allen, OB (1994) The estimation of fixed effects in a mixed linear model, *Proceedings of the 1993 Kansas State University Conference on Applied Statistics in Agriculture*, Kansas State University.

Patterson, HD and Thompson, R (1971) Recovery of inter-block information when block sizes are unequal, *Biometrika*, 58, 545–554.

Patterson, SD and Jones, B (2002) Simulation assessments of statistical aspects of bioequivalence in the pharmaceutical industry, *Pharmaceutical Statistics*, 3(1), 13–23.

Pawitan, Y (2001) *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press, Oxford.

Pinheiro, C and Bates, DM (2000) *Mixed-effects models in S and S-PLUS*, Springer, New York.

Prescott, RJ (1981) The comparison of success rates in cross-over trials in the presence of an order effect, *Journal of the Royal Statistical Society, Series C*, 30, 9–15.

Rao, CR (1971) Estimation of variance and covariance components–MINQUE theory, *Journal of Multivariate Analysis*, 1, 257–275.

Rao, CR (1972) Estimation of variance and covariance components in linear models, *Journal of the American Statistical Association*, 67, 112–115.

Raudenbush, SW and Bryk, AS (2002) *Hierarchical Linear Models: Applications and data analysis methods* (second edition), Sage, Thousand Oaks, CA.

Ridout, M (1991) Testing for random dropouts in repeated measurement data, *Biometrics*, 47(4), 1617–1619.

Ripley, BD (1987) *Stochastic Simulation*, John Wiley & Sons, Inc., New York.

Rowell, JG and Walters, RE (1976) Analysing data with repeated observations on each unit, *Journal of Agricultural Science, Cambridge*, 87, 423–432.

SAS/STAT documentation SAS/STAT 9.1 User's Guide, http://support.sas.com/documentation/onlinedoc/sas9doc.html.

SAS PROC GLIMMIX documentation, http://support.sas.com/rnd/app/papers/glimmix.pdf.

Satterthwaite, FE (1946) An approximate distribution of estimates of variance components, *Biometrics Bulletin*, 2, 110–114.

Schwarz, G (1978) Estimating the dimension of a model, *Annals of Statistics*, 6, 461–464.

Searle, SR, Casella, G and McCulloch, CE (1992) *Variance Components*, John Wiley & Sons, Inc., New York.

Senn, S (2002) *Cross-over Trials in Clinical Research* (second edition), John Wiley & Sons, Ltd, Chichester.

Smith, DM, Roberson, B and Diggle, PJ (1996) Object-orientated software for the analysis of longitudinal data in S, *Technical Report MA 96/192*, Dept of Mathematics and Statistics, University of Lancaster.

Smyth, JF, Bowman, A, Perren, T, Wilkinson, P, Prescott, RJ, Quinn, KJ and Tedeschi, M (1997) Glutathione reduces the toxicity and improves quality of life of women diagnosed with ovarian cancer treated with cisplatin: Results of a double blind, randomised trial, *Annals of Oncology*, 8(6), 569–573.

Snedecor, GW and Cochran, WG (1989) *Statistical Methods* (eighth edition), Iowa State University Press, Ames, IA.

Snijders TAB (2005) Power and sample size in multilevel linear models in Everitt BS and Howell DC (eds) *Encyclopaedia of Statistics in Behavioral Science Vol 3*, 1570–1573, Wiley, Chichester.

Spiegelhalter, DJ, Abrams, KR and Myles JP (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, John Wiley & Sons, Ltd, Chichester.

Talbot, M (1984) Yield variability of crop varieties in the UK, *Journal of Agricultural Science, Cambridge*, 102, 315–321.

Tanner, MA (1996) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (third edition), Springer, New York.

Thall, PF and Vail, SC (1990) Some covariance models for longitudinal count data with overdispersion, *Biometrics*, 46, 657–671.

Thompson, R (1977) The estimation of heritability with unbalanced data. I. Observations available on parents and offspring, *Biometrics*, 33, 485–495.

Thompson, SG and Pocock, SJ (1991) Can meta-analyses be trusted?, *Lancet*, 338, 1127–1130.

Thrusfield, M (1995) *Veterinary Epidemiology*, Blackwell Science, Oxford.

Tierney, L (1994) Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, 22, 1701–1762.

Venables, WN and Ripley, BD (2002) *Modern Applied Statistics with S* (fourth edition), Springer, New York.

Verbeke, G (1995) The *linear mixed model. A critical investigation in the context of longitudinal data analysis*. PhD thesis, Faculty of Science, Department of Mathematics, Catholic University of Leuven.

Verbeke, G and Lesaffre, E (1996), A linear mixed effects model with heterogeneity in the random effects population, *Journal of the American Statistical Association*, 91, 217–221.

Verbeke, G and Molenberghs, G (2000) *Linear Mixed Models for Longitudinal Data*, Springer, New York.

Verbeke, G and Molenberghs, G (2013) The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models, *Biostatistics*, 14, 477–490.

West, BT, Welch, KB and Galecki, AT (2006) *Linear Mixed Models: A Practical Guide Using Statistical Software*, Chapman and Hall.

Whitehead, A (2002) *Meta-Analysis of Controlled Clinical Trials*, John Wiley & Sons, Ltd, Chichester.

Williams, DA (1982) Extra-binomial variation in logistic linear models, *Journal of the Royal Statistical Society, Series C*, 31, 144–148.

Wolfinger, R and O'Connell, M (1993) Generalized linear mixed models: A pseudo-likelihood approach, *Journal of Statistical Computation and Simulation*, 48, 233–243.

Yates, F (1940) The recovery of inter-block information in balanced incomplete block designs, *Annals of Eugenics*, 10, 317–325.

Index

ADJUST option, 465
agriculture, 26
Akaike's information criterion, 240, 457
  PROC MIXED, 457
amantadine, 305–308
analgesic trial, 303–305
analysis of covariance, 15, 16, 18
animal breeding, 26
animal disease, cluster sample, 384–385
animal feeding, 72–73
animal physiology trial (breathing), 374–379
ANOVA, 5, 8, 9, 26
  repeated measures, 233
anti-anxiety agent, 402
apoptosis, 394, 395
approximation, and bias, 137
area under curve (AUC), 232, 335
asthma, 329, 391, 412
ASYCOV option, 458
autocorrelation plot, 83, 96
autoregression: see first-order autoregressive correlation
average bioequivalence, 397, 401

Balaam's design, 305–308
balance, 27, 29–31
  continuous effect, 31
  least squares, 51
  multi-centre, 211
  unbalanced design, 327
balanced incomplete block designs, 425–429, 431, 433
banded covariance, 239
baseline covariate, 143
  cross-over, 294
  fixed effects, 294
  hypertension study, 15–16
  linearity assumption, 15–16
  pre-treatment, 264
Bayesian approach
  classical statistics comparison, 57
  'cold feet' analysis, 147
  exact statistics, 59
  GLMM significance testing, 141
  historical, 27
  hypertension trial analysis, 84–86
  negative variance components, 72
  parameter estimation, 61
  pre-eclampsia, 220, 221
  software (PROC MCMC), 483
  software (PROC MIXED), 473
  standard error bias, 74
  see also posterior density; posterior distribution; prior distribution
Bernoulli distribution, 115
  general exponential form, 123
  variance matrix, 120
Bernoulli form, 133, 480
  dispersion parameter, 139
  PROC GLIMMIX, 220
  pre-eclampsia, 218–230

Applied Mixed Models in Medicine, Third Edition. Helen Brown and Robin Prescott.

Companion Website: www.wiley.com/go/brown/applied_mixed

beta distribution, 67
between-subject design
  animal physiology, 375
  different variances, 370–373
bias
  of approximations, 137
  Bayesian models, 60
  cross-over, 327
  dispersion parameter, 139–140
  empirical estimator, 243
  fixed effects, 50
  GLMM, 139
  mixed model, 26
  ML estimates, 48
  pseudo-likelihood, 131, 132
  random effects, 133, 137
  random effects prediction, 53
  shrinkage, 136–137, 139
  see also error; standard error
bilateral data, 418–422
binary data
  cross-over, 317–321
  dispersion parameter, 120
  extended binary form, 170, 177
  matched sets, 357
  modelling, 115–117
  odds ratio, 318, 326
  sample size estimation, 215, 288
  see also Bernoulli form; binomial form; parameterisation
binomial distribution, 116, 120
  exponential form, 123–124
  variance matrix, 120
binomial form, 133, 480
  covariance pattern model, 477
  dispersion parameter, 139
  mortality estimate, 388
bioavailability, 397
bioequivalence, 329, 397
bioequivalence trials, 329
bivariate normal distribution, 22
block diagonal matrix, 41, 44
blocking effect, 31, 234, 482
  covariance pattern, 31, 44–46
  error stratum, 32
  G matrix, 46, 279, 464
  R matrix, 44–45
  random coefficients model, 31
  `REPEATED` statement, 372
blood pressure: *see* hypertension
breast screening, 414–415
  Edinburgh randomised trial of, 414–415
breathing trial: *see* animal physiology trial

cancer
  children's (herpes virus study), 270–286
  ovarian, 262
canonical link function: *see under* link function
cardiac output trial
  complete block design, 297–299
  covariance pattern model, 309–314
cardiology: *see* heart failure; hypertension
cardiology trial, dogs, 381–383
carry-over
  Balaam's design, 305–308
  critiqued, 305
  four-period, four-treatment, 298–299
  information recovery, 303
  Koch's design, 303–305, 310
  model choice, 312
  standard error, 308
case control studies: *see* matched case–control studies
categorical data
  cross-over trial, 321–323
  sample size, 217, 288
  unordered, 180
  see also mixed ordinal logistic regression; ordinal logistic regression; unordered categorical data
categorical effects, 36
  X matrix, 36
categorical mixed model, 168–196
  'cold feet' analysis, 183, 196
  epilepsy trial, 258
  GLMM comparison, 240
  model checking, 182, 360–362
CD4 count, 267–269
centre effects
  model analysis, 198–202
  model checking, 93
  random, 372–373
  see also centre·treatment effects
centre·treatment effects, 199–201
  binary data, 198–202
  'cold feet' analysis, 150
  DBP analysis, 90, 95
  different variances, 372–373
  model analysis, 198–202

negative variance components, 211
random effects model, 18–19
see also multi-centre trial; treatment effect estimates
centre·treatment interaction
normal mixed models, 86–88
plausibility, 209
centres
number of, 210
size, 210
cerebrovascular insufficiency, 317–321
change score, 317, 322
contingency table, 323
chemotherapy: *see* herpes virus cancer study
chi-squared test, 78, 224
`CHISQ` option, 470, 478
GLM, 140–141
Pearson, 224
test for trend (contingency table), 115, 168
Wald statistic, 220, 470
childhood cancer study, 270–286
`CL` option, 458, 459, 462, 467
cluster randomised trials, 411–415
asthma treatment, 412
'cold feet' analysis: *see under* hypertension study
cluster sample survey, 384–385
communicating results, 24–25, 58
complete block designs, 297–299
compound symmetry pattern, 46
conditional distribution, 68
conditional logistic regression, 125
confidence intervals
'cold feet' analysis, 149–150
`ESTIMATE` statement, 467
GLM(M), 141–142
`LSMEANS` statement, 465
`MODEL` statement, 457, 458
normal mixed model, 78
conjugacy, 67
containment, 28–29
`ESTIMATE` statement, 466–470
containment stratum, 31–33
contingency table analysis
dysmenorrhoea trial, 321–323
GLM superiority, 115
logistic regression, 168
contrast, 75–77, 470
`CONTRAST` statement, 470–471
convergence
Bayesian methods, 59
GEE, 130
GLM, 123
GLMM, 134
iterative methods, 54–55
Newton–Raphson, 54
`PROC MIXED`, 271, 456
pseudo-likelihood, 132
shrinkage, 176
uniform effect categories, 133–134, 139
`CORR` option, 462, 482
correlation
compound symmetry, 46
negative, 72–73
sample size, 286–288
cot death study, 357–370
binary variables, 362
count data, 116, 254
categorical mixed model, 260
GLMM, 260
see also under covariance pattern model; Poisson distribution
covariance matrix: *see* V matrix
covariance parameters
categorical mixed model, 259
exponential decay, 245
fixed/mixed comparison, 19–20
nested, 240: *see also* nesting
R matrix, 182
simple model, 19–20
covariance pattern
compound symmetry model, 236–237
count data, 254–262
first-order autoregressive, 235–237, 241
general (unstructured), 235
heterogeneous, 237
`SAS`, 472, 479
simple, 235–237
Toeplitz, 236, 247
covariance pattern choice
categorical mixed models, 182
repeated measures, 239–241, 335
strategy for, 241, 335
and trial size, 247
covariance pattern model, 19–22, 175–177, 362
blocking variable, 31, 234–235
compound symmetry structure, 313
cross-over trial, 140, 290, 308–313

covariance pattern model (*continued*)
dispersion parameter, 140
error stratum, 32
first-order autoregressive, 333
four-way cross-over trial, 309–314
GLMM **V** matrix, 127
historical, 26
mixed ordinal logistic regression, 175
model fitting, 245–247
multi-centre trial with repeated measures, 350–352
normal data, 244–254
model checking, 248–254
pattern selection, 245–247
number of parameters, 181
quasi-likelihood, 128
random coefficients model comparison, 343
repeated measures, 234–244
`REPEATED` statement, 472–473
repeated within visits, 330–335
see also reparameterisation
covariance structure
covariance pattern models, 44–46
likelihood ratio test, 335
random coefficients model, 42–44
random effects model, 39–42
`SAS` options, 464
covariate effects
parameters, 36
X matrix, 36
Crohn's disease, 394
cross-over trial
AB/BA, 290–296, 317–321
Balaam's design, 305–308
balance, 30
binary data analysis, 317–321
categorical data, 321–323
complete block, 297–299
covariance pattern models, 140, 290, 308–313
defined, 289
error stratum, 31–33
fixed effects model, 35
four-way, 309–314
higher order complete block designs, 297–299
incomplete block designs, 302–305
Koch's design, 303–305
matched studies similarity, 357
mixed models advantage, 1–3, 24, 290
multi-centre, 355
non-informative prior, 66, 76
optimal designs, 305–308
parallel group comparison, 294, 295, 325, 375
random effects model advantage, 326–328
simple model, 3–12
structured covariance pattern model, 308–313
by time, 309
by treatment, 309
trial design, 325
two-period, 303–305
cubic random coefficients model, 271, 273
cumulative probability, 169, 170

data transformation, 79
`DDFM` option, 457, 459
degrees of freedom (DF)
confidence intervals, 78
F tests, 76–77
negative variance, 70, 211
REML adjustment, 26
`SAS` default, 459
variance parameter accuracy, 49, 50, 75, 148
see also Satterthwaite DF
density function, 123, 124
canonical link function, 118
general exponential form, 117
likelihood function, 47
deprivation score, 359–360
design matrix: *see* X matrix; Z matrix
`DIFF` option, 465
dispersion parameter, 120–121
and bias, 139
'cold feet' analysis, 148–149
covariance pattern model, 140
fitting, 139–140
GLM, 139
GLMM, 127, 139–140
diuretic treatment
heart failure, 292–294
pre-eclampsia, 218–230
dose response studies: *see* repeated measures within visits
dropout, 1–3, 91
cot death study, 363

DBP analysis, 95
hypertension trial, 13
repeated measures, 3, 242
drug registration, 313
dysmenorrhoea, 321–323

`E` option, 467
`E3` option, 459, 460
eczema treatment, 391–393
effective sample size, 83, 100, 101, 152
efficiency, 292
electrocardiogram assessment, 317–321
`EMPIRICAL` option, 75, 244, 458
empirical variance, 75, 137, 243
`PROC GENMOD`, 318, 479
epilepsy trial, 114, 254, 258
categorical mixed model, 258
GLMM analysis, 260
error
model choice, 7, 312
residual variance, 70
sample size, 327
variance parameters, 73
see also bias; residual; standard error
error stratum, 31
higher level, 32
mixed models information recovery, 32
multi-centre trial, 198
hypertension trial, REML, 85
`ESTIMATE` statement, 204, 345, 466–470
non-estimable, 468
non-estimable effects, 468
ethics
medical studies, 25
sample size calculation, 286
trial design, 375
event history analysis, 391–393
exact logistic regression, 134
experiment design, sample size, 375–379
explaining results, 24–25, 59
exponential family, 113, 117–118 *see also* GLM
exponential form, general, 117–118, 123–125
extended binary form, 177

F distribution, 70
F test, 5
`CONTRAST` statement, 470
contrasts, 76
denominator DF, 76–77
GLM(M), 140–141
`PROC MIXED`, 457
repeated measures, 242
factor effects: *see* categorical effects
farm, disease prevalence, 384
FDA, 397, 398
first-order autoregressive correlation, 235–237
Fisher scoring, 55
fixed effects estimates, 50–51
categorical mixed model, 260
covariance pattern models, 23
fixed effects matrix: *see* X matrix
fixed effects model, 35–37
assessing fixed effects, 247
assumptions, 7
balance, 29
cerebrovascular insufficiency, 317–321
complete block designs, 297–299
GEE, 130
incomplete data, 7, 290
matched sets, 357
mixed models comparison, 7
multi-centre trial, 198–199, 202–204
outliers, multi-centre trial, 198
random coefficients comparison, 263
random effects model, 7
repeated measures, 232–233
choice, 241–242
significance testing, 75–78
uniform categories, 133–134, 139
see also contingency table analysis; GLM
foetal scans, 380
4 × 4 factorial design
laboratory study of, 394–397

G matrix
covariance pattern model, 46
mixed ordinal logistic regression, 174
random coefficients model, 43–44
random effects model, 39–40
`SAS`, 462, 464
unordered categorical data, 180
gait, 418
general autoregressive model: *see* Toeplitz covariance pattern model
generalisation of results, 210
generalised estimating equations (GEEs), 130–131, 481–483

generalised least squares: *see* iterative generalised least squares
generalised linear mixed models: *see* GLMM
generalised linear models: *see* GLM
generalised logit, 178
`Genstat`, 453
Gewerke test, 111, 152
`GLIMMIX` macro, 131, 158–480
GLM (generalised linear model), 113–125
  V matrix, 119–120
  confidence intervals, 141–142
  contingency table, 114
  defined, 118–121
  dispersion parameter, 139
  distributions, 105–107
  fitting, 121–123
  meta-analysis, 219–220
  `PROC GENMOD`, 155, 481–483
  significance testing, 140–141
GLMM (generalised linear mixed model), 125–132, 480
  categorical model comparison, 240
  confidence intervals, 141–142
  defined, 126–127
  dispersion parameter, 139–140
  epilepsy trial, 260
  fitting, 129–132
    Bayesian approach, 132
    GEE, 130–131, 481–483
    pseudo-likelihood, 131–132
  likelihood function, 127
  meta-analysis, 218–230
  model checking, 142
  quasi-likelihood, 128–129
  significance testing, 141
  see also `GLIMMIX` macro; mixed ordinal logistic regression; unordered categorical data
global estimates, 198, 355
  centre and centre·treatment effects random, 198
  centre number, 198, 355
  centre·treatment interaction, 198, 355
  cluster sample surveys, 384
  double hierarchical, 356–357
  effect on standard error, 217
Goldstein, random effects fitting software, 177
grid search, 473
`GROUP` option, 372, 464

haemophilia study: *see* HIV
heart failure (diuretics) study, 292–294
Hedeker and Gibbons software, 177
herpes virus cancer study, 270–286
  antibody level, 270–286
Hessian matrix, 54
hierarchical data, 23–24, 28–29
  containment, 28–29
  mixed model, 23–24
hierarchical model, 18
hierarchical structure
  double, 356–357
  multi-centre trial, 217, 356
highest posterior density interval (HPD), 61, 108
HIV
  linear random coefficients model, 267–269
  model fitting, 268
hypertension study
  analysis models, 84–86
  baseline covariate, 15–16
  between-subject analysis, 371
  categorical 'cold feet' analysis, 183
    correlation parameter interpretation, 186
    odds ratio, 185, 186
  centre effects, 16–17
  centre·treatment interaction, 17–18
  'cold feet' analysis, 143–147, 162–163
  data modelling, 13–14
  fixed effects models, 157
  Hedeker & Gibbons software, 183, 184
  incomplete data, 90–95
  introduction, 12–13
  model checking, 90–95, 162–163
  models, 143–144
  multi-centre trial, 350–352
  normal data, 244–254
  random effects model, 147
  repeated measures, 19–22, 350–352
  results analysis, 86–88
  sample size, 287
  `SAS` code, 101–112
  treatment·centre, 213–217

ID statement, 474
identity function, as link function, 115, 125
incomplete block design, 302–305
incomplete data, 7
  balance, 31
  cross-over trial, 290–291
  DBP analysis, 91
  mixed model advantage, 1–3, 23–24
  not at random, 234
  randomness requirement, 81
  repeated measures, 3, 231–234, 329–330
  unbalanced design, 327
individual bioequivalence, 401–402
inference
  fixed effects model, 202
  model choice, 198
  random effects model, 200, 209
  see also generalisation of results
information matrix, 55
information recovery, 10, 32, 303, 304 *see also* incomplete data
integrated care pathways, 412
intention to treat, 84
inter/intra-observer variability
  foetal scan, 380
  heart wall thickness, 381–383
invariance to time origin, 44, 278
inverse gamma distribution, 63, 66, 148
  `PRIOR` statement, 473
iterative generalised least squares, 49–50
  fixed effects estimation, 51
  incomplete block design, 303
  random effects prediction, 53
  restricted, 49–50, 57
  variance parameter estimation, 55–57
iterative methods, 53–55
  GEE, 132
  GLM fitting, 121–123
  Newton–Raphson, 54–55, 121–123
  pseudo-likelihood maximisation, 131–132

Jeffreys' method, 66, 474

Kenward–Roger adjustment, 74–76, 88, 89, 148, 243, 265, 272, 312, 313, 327, 404, 459
knee angles, 420
Koch's design, 303–305

L matrix, 470
last value carried forward, 14
least squares: *see* IGLS; ordinary least squares
least squares mean, 471, 490
leukaemia: *see* childhood cancer study
likelihood function
  GLM, 121
  GLMM, 127
  infinite, 456
  information criteria measures, 240
  log likelihood, 47–50
    GLM fitting, 121
    GLMM, 128
    ordinary/REML, 48–49
  mixed ordinal logistic regression, 173
  model fitting, 47–50
  non-informative prior similarity, 62
  ordinal logistic regression, 169
  ordinary/REML, 48–49
  standardised, 60
  true likelihood, 47–50
  see also maximum likelihood; pseudo-likelihood; quasi-likelihood; REML
likelihood ratio test, 240
  covariance structure determination, 335–336
  variance components significance, 77, 265
linear dependencies, **X** matrix, 37
linearised pseudo variable, 122, 131–132, 142
  link function, 114–115
  canonical, 115, 118–119, 123, 124
  GLMM, 126
  mixed ordinal logistic regression, 169
  non-canonical, 119
  see also log likelihood function; logit link function
Lipsitz macro
  parameterisation, 177
  source, 477
local estimates, 198, 355
  centre effects random, 355
  grid search, 456
  meta-analysis, 217

local maximum, PROC MIXED, 456
location parameter, 61, 117
log likelihood: *see under* likelihood function
log link function, 135–136
logistic regression, 125
  conditional, 125
  see also conditional logistic regression; logit link function
logit link function, 123
  Bernoulli, 123
  binomial, 123–124
  cumulative probability, 169
  generalised, 178
  see also link function; log link function
LSMEANS statement, 465–466
LSMESTIMATE statement, 454, 477, 481
lung function trial, 374–379

MAKE statement, 476
marginal methods, 49 *see also* REML
marginal posterior distribution, 61
Markov chain Monte Carlo methods, 68
matched case–control studies, 357
  cot death study, 357–370
  see also matched sets
matched sets
  cot death study, 362–364
  fixed effects model, 358
Matlab, 425, 426
matrix notation, 34
matrix, positive semi-definite, 264
maximum likelihood
  Bayesian comparison, 58
  fixed effects estimation, 50
  historical context, 25
  model fitting, 47
  random effects prediction, 51–53
  see also likelihood function; REML
mean response, 232–233
median, 61, 86
meta-analysis, 217–218, 356
  data inclusion criteria, 221
  example, 218–230
  mixed model, 2
method of scoring: *see* Fisher scoring
METHOD option, 458
Metropolis algorithm 68–69, 86, 483
micturition frequency, 293
missing data: *see* incomplete data

mixed model, 476
  advantages, 1–3, 23–24
  defined, 2, 22, 37–39
  disadvantages, 24
  historical, 25–27
  incomplete data, 234
  multi-centre trial, 198
  see also covariance pattern model; normal mixed model; random coefficients model; random effects model
mixed ordinal logistic regression
  binary modelling, 169
  covariance pattern model, 175
  defined, 173–174
  as GLMM, 169, 173–177
  model fitting, 177, 183, 184
  R matrix, 174–176
MLwiN, 453–454
mode, 61
model-based approach, covariance pattern models, cross-over trials, 312
model building, 3–12
  childhood cancer study, 270
  non-linear temporal, 270
  repeated measures, 335
model checking, 79–81
  categorical mixed, 182
  covariance pattern, 244, 248–254
  DBP analysis
    centre effects, 93, 95
    centre·treatment effects, 95, 95
    GLMM, 142
    outliers, 90
    residuals, 90–95
    SAS code, 101
  matched case–control, 360–362
  Pearson residuals, 256
  polynomial random coefficients, 270
  random coefficients, 265, 273
  random effects 'cold feet' analysis, 145, 162–163
  see also residual plots
model choice
  assumptions, 312
  binary data, 326
  covariance pattern model, 312
  multi-centre trial, 198–202
  non-normal data, 326
  small samples, 410
  standard error, 295, 312

model comparison, covariance pattern models, 240
model fitting, 46–57
  Bayesian, 57–69
  covariance pattern model
    repeated measures, 240, 245–247
    systolic blood pressure trial, 335–340
  GLM, 121–123
  GLMM problems, 139–140
  likelihood function, 47–50
  linear random coefficients, 267
  manually, 11
  measures of, 224, 240
  mixed ordinal logistic regression, 177, 183, 184
  multi-centre trial with repeated measures, 350–352
  overfitting, 239–240
  parameter estimation, 108–110
  `PROC MIXED`, 457
  statistical comparison, 240
  unordered categorical data, 180
model selection: *see* model choice
`MODEL` statement, 458–462, 481
mortality estimates, 386–388
mouthwash trial, 295, 325
multi-centre trial
  analysis considerations, 209–211
  balance, 211
  centre effects, 209–211
  cross-over, 355
  defined, 197
  fixed effects model
    analysis implications, 198–199
    centre·treatment effects omitted, 201–202
    outliers, 198
  hierarchical, 217, 356
  meta-analysis, 198–202, 217–218
    data inclusion criteria, 221
    example, 218–230
    outliers, 217–218
  mixed model, 198
  number of centres, 210
  random effects model, 220
  repeated measurements, 349–352
  sample size estimation, 211–217
multinomial correlation, **R** matrix, 172–173
multinomial distribution, mixed ordinal logistic regression, 172
multinomial probability, unordered categorical, 178
multiple contrast, 75–77
multivariate normal distribution
  checking, 248–254, 266, 273
  density function, likelihood, 47
  `SAS` coding, 279

negative correlation, 72–73
nesting, 240–241, 247, 351
Newton–Raphson iteration, 54–55, 121–123
nicotine, 394, 395
`NOBOUND` option, 264, 272, 409
`NOCLPRINT` option, 458
`NOINFO` option, 458
`NOITPRINT` option, 458
non-comparative data, 265
non-estimable effects, 158, 468
non-normal data
  distribution, 113, 114
  likelihood, 128
  model selection, 326
  see also GLM
normal data
  covariance pattern model, 244–254
  sample size estimation, 211–217, 286
normal distribution
  assumption checking, 79
  bivariate, 22
  link function, 125
  residual, 4
  transformation to, 267
  see also multivariate normal distribution
normal mixed model, 34, 37–39
  Bayesian approach, 57–69
  fixed effects estimation, 50–51
  likelihood function, 47–50
  model fitting, 46–57
  random effects estimation, 51–53
  relation to GLM, 113
  variance parameter estimation, 53–57
  see also covariance pattern model; mixed model; random coefficients model; random effects model
normal probability plots, 79, 90, 142, 163, 360, *361*

null hypothesis, 62, 76
numerical methods, 53–55, 67

observer variation
  binary data, 28, 325, 326
  categorical data, 265, 321
  'cold feet' analysis, 150
  dog cardiology, 381–382
  epilepsy, 260
  foetal scans, 380
  logit link function, 123
odds ratio, 135, 166, 185, 186, 318, 326
ODS OUTPUT statement, 475, 480, 482
ODS SELECT statement, 110, 164, 484
oedema status, 293, 296
offset, Poisson distribution, 116
one-parameter distributions, 115–117
order effects, 290
ordered categorical data: *see under* ordinal logistic regression
ordinal logistic regression, 168–173
  binary data, 319
  categorical data, 323, 357, 359, 363
  see also mixed ordinary logistic regression
ordinary least squares, 51, 84
outliers, 79–81
  GLMM model checking, 142
  meta-analysis, 217–218
  multi-centre analysis, 84, 87
  Pearson residual plots, 257
  polynomial random coefficients model, 244, 266
  random coefficients model, 267
  random effects model, 198
  repeated measures analysis, 248–254
  SOLUTION option, 464
ovarian cancer trial, 262
overfitting, 239–240
overinterpretation of data, 273
overparameterisation, 37
  p-values, Bayesian, 59, 78

parallel group design, 294, 325, 375
parameter estimation, Bayesian, 61
parameterisation, 175, 181
Parkinsonism drug trial, 305–308
PARMS statement, 473
partitioning
  categorical data, 217
  extended binary form, 169
  proportional odds assumption, 181
patella tracking, 419, 420
patient: *see* subject
PDIFF option, 465
Pearson residuals, 156, 257
period effect, 293: *see also* time effect
pharmaceutical industry
  empirical variance, 244
  model choice, 313
physiological response, 232
plaque score, 295
PM option, 459, 462
Poisson data
  covariance pattern model fitting, SAS, 481
  dispersion parameter, 139
  shrinkage, 136–137
Poisson distribution
  general exponential form, 124
  model checking, 257
  offset, 116
  variance matrix, 120
Poisson regression: *see* log link function
polynomial model building, 270–272
pooled comparison, 291
population bioequivalence, 397, 401, 402
population averaged method: *see* marginal quasi-likelihood
post-operative complications, 389
posterior density, 59–60
  determination, 59–60
posterior distribution, 58, 61
  evaluation, 67–69
  marginal, 61
  SAS, 110
  simulated, 68
power: *see* sample size estimation
pre-eclampsia, 218–230
pre-treatment: *see under* baseline covariate
predicted values, 482
Prescott's test, 318, 322
prior distribution, 58
  conjugate, 67
  flat, 63, 66, 474
  informative, 74
  non-informative, 58, 66
    'cold feet' analysis, 143, 149
    prior similarity, 63
    specification, 62–67

PRIOR statement, 473–474, 486–487
proper, 66–67
properties, 66–67
probability
cumulative, 169, 170
multinomial, 178
see also p-values
probability intervals, 60
probit function, 119
PROC GENMOD, 155, 480–483
PROC GLIMMIX, 137–138, 150, 185, 260, 313, 476
PROC MCMC, 483–488
PROC MIXED, 454–476
default fitting, 55
empirical variance, 75
negative variance, 70
statement options, 478: *see also* specific statement
variance parameter fixing, 74
proportional odds assumption, 181
protocol checking, multi-centre trial, 198
pseudo-likelihood, 131–132
'cold feet' analysis, 147
GLIMMIX macro, 476
pseudo variable, 122, 131–132

quadratic function, 53
quadratic random coefficients model, 273–278
qualitative variables: *see* categorical effects
quasi-likelihood, 128–129
categorical, 182
GLMM, 129
maximisation, 129–132

R matrix, 40
banded covariance, 239
compound symmetry structure, 238
covariance pattern model, 44–46
general structure, 238
GLMM, 126
mixed ordinal logistic regression, 174–176
random coefficients model, 43
random effects model, 40
repeated measures, 235, 238
REPEATED statement, 472–473
submatrices, 45
uncorrelated, 238
radiologist reliability, 380
random coefficients model, 23, 267
covariance pattern model comparison, 343
covariance structure, 42–44
cubic, 273
error stratum, 33
examples, 267–286
fitting by GEE, 132
GLMM **V** matrix, 127
likelihood function, 127
linear, 263
model fitting, 263
repeated measures (HIV), 267–269
model checking, 273
non-linear, 270
polynomial (childhood cancer), 270, 281
quadratic, 273–278
random effects, 21
RANDOM statement, 279, 464
repeated measures, 21–22, 262
repeated measures within visits, 341–343
shrinkage of estimates, 265
significance testing, 75–78, 265
systolic blood pressure trial, 341–347
see also count data; covariance pattern model; event history analysis; random coefficients model
random effects coefficient, 37, 51–53
random effects estimation, 12
random effects model, 22
assumptions, 7
balance, 30–31
'cold feet' analysis, 147
complete block designs, 297–299
compound symmetry, 236
covariance pattern comparison, 309–314
covariance structure, 39–42
cross-over trial, 143, 326–328
fitting by GEE, 132
fitting software, 181, 452–454
fixed models comparison, 7–9
GLMM, 126, 127
hierarchical, 18
likelihood, 127
model checking, 79–81
multi-centre trial, 200–202
negative variance, 69
pre-eclampsia trial, 218–230
R matrix, 176
random coefficient models, 21

random effects model (*continued*)
significance testing, 75–78
standard error bias, 148
uniform effects, 133–134, 139
V matrix, 127
see also reparameterisation
RANDOM statement, 279, 336–338, 462–465
random vs. fixed effects modelling, 7
randomisation, cross-over trial, 290
ranking, 322, 387–388
reciprocal distribution, 63
recovery of information
binary data cross-over, 317–321
cross-over, 292, 294
incomplete block, 303
Koch's design, 303
matched sets, 357
two-subject period cross-over, 303
reference category, 158
relative rate, 136, 257
relative risk, 136
see also log link function
REML (residual maximum likelihood)
DBP analysis model, 85
definition, 48
fixed effects estimation, 50
historical, 26
log likelihood, PROC MIXED, 454
model fitting, 47
random effects prediction, 51–53
standard error, 87
reparameterisation
case control, 362
GLMM, 176
random effects models as covariance pattern model, 175–176
repeated measures, 231–234
analyses by time point, 233
banded covariance, 239
containment stratum, 32
covariance pattern model, 234–244
count data, 254
normal data, 244–254
different covariances, 238
different variances, 237
error stratum, 32
event history analysis, 391
fixed effects model, 232–233
hypertension trial, 19–22, 244–254
linear, 262, 267–269
mixed model, 3, 234
model choice, 241–242, 266
non-comparative datasets, 265
polynomial, 270–272
sample size estimation, 287
see also outliers
repeated measures within visits, 329–347
covariance structure, 332
cross-over trial (SBP)
covariance pattern choice, 335–340, 343–347
random coefficients, 341–347
reps cross-over, 335–340, 343–347
defined, 329
multi-centre trial, 349–352
random coefficients model, 341–347
treatment·reps interaction, 329–330
REPEATED statement, 336–338, 472–473, 482
replicate cross-over designs, bioequivalence studies with, 397–411
residual, 4, 482
error strata, 31
full, 48
homogeneity, 79
normal distribution, 4
Pearson, 142
standardised, 79
see also R matrix; residual plots
residual maximum likelihood: *see* REML
residual plots
covariance pattern model, 244
GLMM, 142
matched case–control, 359
normal data, 79–81, 90–92
polynomial random coefficients model, 257, 262
repeated measures, 248–254, 262
SAS code, 155
residual variance, 4–6
response profile, 232
restricted IGLS, 49–50, 57
restricted maximum likelihood: *see* REML
resurface, 420
robust methods, 80
run-in period, 290, 310

safety trials, 263
sample size

model selection, 389
shrinkage, 386–390
small, 326
sample size estimation
animal physiology trial, 375–379
categorical data, 288
multi-centre trial, 211–217
precision of, 217
repeated measures, 287
sampling cost
binary data, 215
normal data, 214
SAS
empirical standard error, 352
GLIMMIX macro, 480
Lipsitz macro source, 453
ordinal logistic regression macro, 323
PROC GENMOD, 158, 480–483
PROC MIXED, 454–476
base density, 75
default fitting, 55
empirical variance, 75
negative variance, 75
statement options, 478: *see also* specific statement
residual plots, 155
syntax, 454
variance parameter fixing, 75
uniform reference category, 155
V matrix, 462
SATTERTH option, 459, 460
Satterthwaite DF, 76–77
GLMM, 141
repeated measures data, 242
Schwarz's information criterion, 241, 457
selected comparison design, 431–433
sensitivity analysis, 377
shrinkage, 12, 24
bias, 132, 136, 139
dispersion parameter as metric, 139
ESTIMATE statement, 469–470
GLMM, 136–137, 139
hierarchical multi-centre trial, 356
meta-analysis, 217–218
random coefficients estimates, 265
random effects, 52, 218, 220, 221
ranking, 386–388
raw data comparison, 386–388
small area mortality, 386, *387*
surgeon performance, 390
treatment by centre, 203
side effects, 80, 263
significance testing, 75–78
Bayesian: *see* p-values
'cold feet' analysis, 143, 147
GLM, 140–141
GLMM, 141
negative variance components, 211
random coefficients model, 265
repeated measures, 242
simulation, 59, 74, 243
SLICE statement, 471
SLICEDIFF option,
small area estimates, 386–388
small sample bias, 312, 326
shrunken estimates, 386
social deprivation score, 359–360
SOLUTION option, 163, 459, 463–464, 480
sparse data, 141, 313
sphericity, test of, 236
S-plus, 453
SPSS, 453
standard error, 50, 72, 87–88, 134, 149, 243–244
Bayesian, 61, 86–88
carry-over, 298, 303, 308
compound symmetry structure, 313
covariance pattern models, 23, 312
cross-over, 326
different variances, 370, 371
empirical
'cold feet' analysis, 187
model-based comparison, 247, 257
fixed effects, 50, 243–244
heart failure cross-over trial, 294
hypertension study, 86–88
increase by mixed model, 198
linear random coefficients model, 267
model-based, compound symmetry structure, 312
model-based/empirical
covariance pattern model (cross-over), 312, 337
epilepsy trial, 260
model choice, 10, 295, 313
multi-centre trial, 202–204, 217
random effects, 137
repeated measures fixed effects, 233

standard error (*continued*)
simple model, 6
treatment·centre interaction, 86, 87
treatment, fixed effects model, 199
`Stata`, 454
statistics, explaining results, 24–25
`STORE` statement, 474
subject effects, 3
`SUBJECT` option, **G** matrix blocking, 464
subject·time effects, 21
submatrices, **R** matrix, 45
sudden infant death syndrome: *see* cot death study
summary statistics, 329
surgical audit, 2
syntax: *see* specific procedure or option

t test, 6, 76, 322, 388
Taylor series, 122, 131
test for trend: *see* chi-squared test; t-test
thinning, 83, 111, 167
time effect, 293
covariance patterns, 237
model choice, 271
non-linear models, 263
random coefficients model, 21–22, 265, 341–343
virus antibodies, 19, 270–278
visit time–response model, 341–342
time interval, 239
time slope, 21, 341
Toeplitz covariance pattern model, 236, 245
repeated measures, 332
`SAS`, 480, 482
toxicology experiments, 119
trace plot, 69, 83, 96
treatment effect estimates, 198–199
binary data cross-over trial, 317–321
DBP study, 89
fixed effects, 198–199
local/global, multi-centre trial, 217
multi-centre trial with repeated measures, 349, 352
random effects model, 197, 203
relative rate, 257
separate analyses by time, 233
variance, 292
treatment·centre: *see* centre·treatment
treatment·rep effects, 344
treatment·time effect
hypertension trial, 19–20
linear random coefficients model, 265
omitted, 241–242
random coefficients models, 263
repeated measures, 247
weighting, 247
see also time effect
trial design
cross-over trial, 325
inclusion criteria, 221
optimisation, 375–379
sensitivity analysis, 377
variance components, 374–379
see also sample size estimation
trial·treatment, pre-eclampsia study, 218–221
type III tests: *see* Wald statistic
`TYPE` option, 279

ulcerative colitis, 394
ultrasound scans
dog cardiology, 381–382
foetal, 380
under-dispersion, 139
uniform distribution, non-informative prior, 63
uniform effect categories, 144
'cold feet' analysis, 143
GLMM, 133–134
reparameterisation, 362
uniform reference category, standard error, 144
unordered categorical data, 180

V matrix, 37, 39, 462
Bernoulli distribution, 120
between-subject trial, 371
binomial distribution, 120
fitting methods, 126
GLM, 119–120
GLMM, 126
Poisson distribution, 120
random coefficients model, 44
random effects model, 40–42
repeated measures within visits, 330–335
`SAS`, 456, 462
standard error bias, 87–88
time origin invariance, 44, 278

V option, 463
variance
  canonical link function, 119
  different variances, 237, 370–373
  dispersion parameter, 120–121, 127
  empirical, 75, 137, 244
  empirical/model based (PROC GENMOD), 480–483
  general exponential form, 117–118
  negative, 363
  non-convergence, 264
  randomisation, cross-over trial, 291
  sample size estimation, 215–216, 286
variance components, 7, 473
  animal physiology trial, 374
  ANOVA, 8
  Bayesian, 85
  bias, 149
  DBP, 89
  estimation, 61
  negative, 69–73, 134, 143, 211
    random coefficients, 272
  negative:Bayesian, 72
  number of centres, 210
variance matrix: *see* V matrix
variance parameters
  accuracy, 73, 148
  bias, 50
    pseudo-likelihood, 131
    shrinkage, 136–137
  covariance, 55
  defined, 46–47
  GEE, 131
  IGLS, 54–57
  maximum likelihood methods, 53
  significance testing, 77, 141

Wald statistic
  GLM(M), 141
  PROC MIXED (Type III test), 457
  t test, 76
washout, 290
WEIGHT statement, 474
weighting, treatment·time effect, 247
Williams modification, dispersion parameter, 121
WinBUGS, 454
withdrawal: *see* dropout
within-subject design
  animal physiology, 375–378
  correlation, 20
  cross-over trial, 291
  see also cross-over trial
within-visit observations, correlation matrix, 330

X matrix, 36–37
  mixed ordinal logistic regression, 171
  unordered categorical data, 179

Z matrix, 38–39
  random coefficients, 43
  unordered categorical data, 179

Statistics in Practice

Human and Biological Sciences

Berger – Selection Bias and Covariate Imbalances in Randomized Clinical Trials

Berger and Wong – An Introduction to Optimal Designs for Social and Biomedical Research

Brown, Gregory, Twelves and Brown – A Practical Guide to Designing Phase II Trials in Oncology

Brown and Prescott – Applied Mixed Models in Medicine, Third Edition

Campbell and Walters – How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research

Carpenter and Kenward – Multiple Imputation and its Application

Carstensen – Comparing Clinical Measurement Methods

Chevret (Ed.) – Statistical Methods for Dose-Finding Experiments

Cooke – Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity

Eldridge – A Practical Guide to Cluster Randomised Trials in Health Services Research

Ellenberg, Fleming and DeMets – Data Monitoring Committees in Clinical Trials: A Practical Perspective

Gould (Ed) – Statistical Methods for Evaluating Safety in Medical Product Development

Hauschke, Steinijans and Pigeot – Bioequivalence Studies in Drug Development: Methods and Applications

Källén – Understanding Biostatistics

Lawson, Browne and Vidal Rodeiro – Disease Mapping with Win-BUGS and MLwiN

Lesaffre, Feine, Leroux and Declerck – Statistical and Methodological Aspects of Oral Health Research

Lesaffre and Lawson – Bayesian Biostatistics

Lui – Binary Data Analysis of Randomized Clinical Trials with Noncompliance

Lui – Statistical Estimation of Epidemiological Risk

Marubini and Valsecchi – Analysing Survival Data from Clinical Trials and Observation Studies

Millar – Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB

Molenberghs and Kenward – Missing Data in Clinical Studies

Morton, Mengersen, Playford and Whitby – Statistical Methods for Hospital Monitoring with R

O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley and Rakow – Uncertain Judgements: Eliciting Expert's Probabilities

O'Kelly and Ratitch – Clinical Trials with Missing Data: A Guide for Practitioners

Parmigiani – Modeling in Medical Decision Making: A Bayesian Approach

Pintilie – Competing Risks: A Practical Perspective

Senn – Cross-over Trials in Clinical Research, Second Edition

Senn – Statistical Issues in Drug Development, Second Edition

Spiegelhalter, Abrams and Myles – Bayesian Approaches to Clinical Trials and Health-Care Evaluation

Walters – Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation

Welton, Sutton, Cooper and Ades – Evidence Synthesis for Decision Making in Healthcare

Whitehead – Design and Analysis of Sequential Clinical Trials, Revised Second Edition

Whitehead – Meta-Analysis of Controlled Clinical Trials

Willan and Briggs – Statistical Analysis of Cost Effectiveness Data

Winkel and Zhang – Statistical Development of Quality in Medicine

Zhou, Zhou, Lui and Ding – Applied Missing Data Analysis in the Health Sciences

Earth and Environmental Sciences

Buck, Cavanagh and Litton – Bayesian Approach to Interpreting Archaeological Data

Chandler and Scott – Statistical Methods for Trend Detection and Analysis in the Environmental Statistics

Christie, Cliffe, Dawid and Senn (Eds.) – Simplicity, Complexity and Modelling

Gibbons, Bhaumik and Aryal – Statistical Methods for Groundwater Monitoring, 2nd Edition

Haas – Improving Natural Resource Management: Ecological and Political Models

Haas – Introduction to Probability and Statistics for Ecosystem Managers

Helsel – Nondetects and Data Analysis: Statistics for Censored Environmental Data

Illian, Penttinen, Stoyan and Stoyan – Statistical Analysis and Modelling of Spatial Point Patterns

Mateu and Muller (Eds) – Spatio-Temporal Design: Advances in Efficient Data Acquisition

McBride – Using Statistical Methods for Water Quality Management

Ofungwu – Statistical Applications for Environmental Analysis and Risk Assessment

Okabe and Sugihara – Spatial Analysis Along Networks: Statistical and Computational Methods

Webster and Oliver – Geostatistics for Environmental Scientists, Second Edition

Wymer (Ed.) – Statistical Framework for Recreational Water Quality Criteria and Monitoring

Industry, Commerce and Finance

Aitken – Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition

Balding and Steele – Weight-of-evidence for Forensic DNA Profiles, Second Edition

Brandimarte – Numerical Methods in Finance and Economics: A MATLAB-Based Introduction, Second Edition

Brandimarte and Zotteri – Introduction to Distribution Logistics

Chan – Simulation Techniques in Financial Risk Management

Coleman, Greenfield, Stewardson and Montgomery (Eds) – Statistical Practice in Business and Industry

Frisen (Ed.) – Financial Surveillance

Fung and Hu – Statistical DNA Forensics

Jank and Shmueli – Modeling Online Auctions

Jank and Shmueli (Ed.) – Statistical Methods in e-Commerce Research

Lloyd – Data Driven Business Decisions

Kenett (Ed.) – Operational Risk Management: A Practical Approach to Intelligent Data Analysis

Kenett (Ed.) – Modern Analysis of Customer Surveys: With Applications using R

Kenett and Zacks – Modern Industrial Statistics: With Applications in R, MINITAB and JMP, Second Edition

Kruger and Xie – Statistical Monitoring of Complex Multivariate Processes: With Applications in Industrial Process Control

Lehtonen and Pahkinen – Practical Methods for Design and Analysis of Complex Surveys, Second Edition

Mallick, Gold, and Baladandayuthapani – Bayesian Analysis of Gene Expression Data

Ohser and Mücklich – Statistical Analysis of Microstructures in Materials Science

Pasiouras (Ed.) – Efficiency and Productivity Growth: Modelling in the Financial Services Industry

Pawlowsky-Glahn, Egozcue and Tolosana-Delgado – Modelling and Analysis of Compositional Data

Pfaff – Financial Risk Modelling and Portfolio Optimization with R

Pourret, Naim and Marcot (Eds) – Bayesian Networks: A Practical Guide to Applications

Rausand – Risk Assessment: Theory, Methods, and Applications

Taroni, Biedermann, Bozza, Garbolino and Aitken – Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science, Second Edition

Taroni, Bozza, Biedermann, Garbolino and Aitken – Data Analysis in Forensic Science

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.