23

REPEATED MEASURES DATA

OBJECTIVES

After reading this chapter, you should be able to:

1. Recognise a repeated measures data structure, and understand the unique characteristics of a repeated measures structure.

2. Use descriptive and graphical tools to quantify and visualise the repeated measures structure of a dataset.

3. Use simple univariate approaches to analyse repeated measures data.

4. Use mixed models to analyse repeated measures data, and understand the limitations of random-intercept mixed models for such data.

5. Choose among a variety of correlation structures that might be appropriate for repeated measures or spatial data.

6. Understand the fundamental differences between mixed model and generalised estimating equation (GEE) approaches for analysis of clustered data.

7. Use GEE procedures to analyse clustered data, in particular repeated measures data.

23.1 INTRODUCTION

In this chapter, we will describe methods for analysis of repeated measures data that, as discussed in Chapter 20, could be considered as a special type of clustered data. It is also one of the most commonly encountered data structures in veterinary epidemiology and the health sciences in general. A wide selection of methods and approaches exist for analysis of such data, and the choice between them depend on the characteristics of the data at hand as well as the objective of the analysis. We cannot, within a single chapter in the book, cover all methods, or cover the methods selected in full detail. Among the many excellent textbooks on repeated measures or longitudinal data, a standard (fairly theoretical) reference is Diggle *et al* (2002), and also Fitzmaurice *et al* (2004); Molenberghs & Verbeke (2005); Verbeke & Molenberghs (2001) provide extensive coverage in a blend of theory and practice.

To illustrate the methods, we will later revisit the somatic cell count dataset (-scc_40-) studied in Chapter 21 but first consider a dataset (-fish_trial-) on growth of salmon in a clinical field vaccine trial. Within a sea cage in commercial aquaculture, the fish were tagged electronically (with passive integrated transponder tags), randomly allocated to one of several vaccine groups and followed throughout the production cycle until harvest by repeated sampling events in which all fish in the cage were weighed and inspected. The dataset considered here, a small subset of the full dataset, includes weight measurements at 3 samplings after the initial vaccination as well at harvest, for a total of 5 measures over time on each fish. At each sampling, the fish were screened for health issues; one of these (jaw deformity) is included in the dataset (Table 23.1).

Table 23.1 Selected variables from the dataset fish_trial

Variable	Description
fish	fish identification
sample	sampling number (1-5), where 1~vaccination and 5~harvest
day	days since vaccination (0-900)
wt	weight in grams
vaccine	vaccine group (1-4)
jaw	presence of jaw deformity (0/1)

23.1.1 What is repeated measures data?

A **longitudinal** study can be characterised by having several measurements over time on the same **subjects** (individuals, or sometimes other experimental units such as sample plots in a field), as opposed to studies with only one measurement per subject. A longitudinal study certainly involves **repeated measures** (or measurements) on the same subjects, but the latter term is sometimes used in a slightly more general sense to denote consecutive measurements, *ie* measurements with a certain inherent ordering (*eg* along the line of cows in a tie-stall barn). If there is no ordering to multiple measures on the same subjects, we might think of these as clustered within the same subject instead, as discussed previously (Chapters 20-22).

From the clustering of measurements within subjects, we already know that it is usually unreasonable to consider the measurements as independent; by doing so we would ignore any

subject characteristics affecting our outcome. For example, a fish that is large relative to its fellow inmates at young age would tend to remain relatively large throughout the growth period. The (time) ordering of the measurements introduces another type of association between measurements because usually 2 measures on the same subject that are taken close in time will be stronger associated than measures taken further apart in time. In the fish trial, the initial weight should be stronger correlated with the weight at the first sampling event than the harvest weight. Such a pattern of correlations we broadly refer to as **autocorrelation** without stating specifically how the correlation is reduced by increasing time distance. It is because of this feature of repeated measures data that they cannot generally be treated as a hierarchical data structure. Specifically, a 2-level hierarchical structure (with measurements nested within subjects) does not take time ordering of the measurements into account in the random part of the model. Where animals within a herd can be interchanged without altering the meaning of the data, observations over time on the same subject cannot. Despite the intuitive logic of autocorrelation, some outcomes may not show any autocorrelation, so we will need to assess for each dataset individually whether autocorrelation or a simple clustering within subjects is present in the data, or perhaps no clustering at all.

As the range of methods that can be applied to a dataset depends on its structure, it is useful to introduce some terminology to describe repeated measures data. The most regular data type has the same number of measures taken for each subject (*ie* is **balanced** over time) with **uniform** (*ie* the time points are the same across subjects) and equally spaced (**equidistant**) time points. For example, the fish_trial data is balanced and has uniform but non-equidistant points because the sampling was not carried out at regular intervals. Protocols in clinical trials commonly require equidistant sampling or follow-up. Generally speaking, the most regular data types will not only allow a wider range of analytical approaches, but will also be easier to analyse. The presence of **missing data** will make designs unbalanced, but are difficult to avoid. A unique feature to repeated measures data is that observations may be missing because the subject exits prematurely from the study (**drop-outs**).

23.1.2 Descriptive statistics and graphical displays

As the choice of analytical methodology for repeated measures data will also depend, to some extent, on the characteristics of the data at hand, it is crucially important to familiarise oneself with the data before plunging into a complicated analysis. Two obvious approaches for that are suitably chosen descriptive statistics and visualisations of the data, and analysis by simple (possibly simplistic) procedures such as those described in Section 23.2. To begin with, one should assess the distribution of time points within each subject to determine how regular these are (*eg* balanced, equidistant). Next, one should compute suitable means across subjects at different time points to get an impression of how time affects the outcome; these can be plotted against time in a **mean plot**. If time points are uniform, it is often also useful to compute the crude correlations between measurements at different time points. This will often require to shift from **long data format** (each row corresponds to one measurement) to **wide data format** (each row corresponds to one subject, with the measurements distributed across several columns). Finally, it is recommended to construct one or multiple **profile plots** showing the series of observations over time on the subjects. If there are too many subjects to display them all in a single plot, one may construct plots for suitably chosen groups (formed by predictor values) and/or select some subjects for display. Examples 23.1-2 illustrate the approach for the fish_trial and scc_40 datasets.

23.1.3 Longitudinal versus cross-sectional study designs

Diggle *et al* contrasted the longitudinal and cross-sectional designs (Diggle *et al*, 2002, Chapter 2), and we'll briefly review the main points. A cross-sectional study can be used to give

Example 23.1 Graphs and descriptive statistics for fish trial
data = fish_trial

As the fish grow from approximately 50g at vaccination to several kilograms at harvest, it is impractical to analyse the weights on the original scale, because variances and possibly, correlations, among other things, will depend heavily on time. We work instead on a log-transformed scale where additive effects correspond to multiplicative effects and linear growth corresponds to exponential growth on the original scale. Our focus will be on effects of vaccine groups and jaw deformities on the logarithmic growth, so it is natural to compute descriptive statistics for all combinations of these 2 categorical predictors.

Mean log weight (SD)		Number of fish	Sampling				
vaccine	jaw		1	2	3	4	5
1	0	91	4.10 (.17)	4.37 (.17)	6.52 (.15)	7.55 (.24)	8.68 (.21)
	1	9	4.04 (.19)	4.32 (.24)	6.40 (.32)	7.59 (.44)	8.36 (.31)
2	0	86	4.10 (.20)	4.36 (.21)	6.53 (.19)	7.54 (.24)	8.65 (.26)
	1	14	4.04 (.18)	4.30 (.11)	6.43 (.19)	7.36 (.44)	8.33 (.38)
3	0	88	4.10 (.19)	4.37 (.19)	6.52 (.17)	7.58 (.24)	8.69 (.22)
	1	12	4.11 (.22)	4.34 (.24)	6.48 (.26)	7.47 (.44)	8.65 (.37)
4	0	88	4.07 (.18)	4.32 (.19)	6.47 (.18)	7.50 (.24)	8.61 (.25)
	1	12	3.98 (.18)	4.22 (.17)	6.45 (.20)	7.39 (.44)	8.45 (.22)

The table shows that differences between vaccine groups are small but apparently consistent over time, whereas patterns are more variable among fish with jaw deformities, possibly due to the lower sample size. Fig. 23.1 shows a profile plot for 15 selected fish and a mean plot comparing fish with and without jaw deformities.

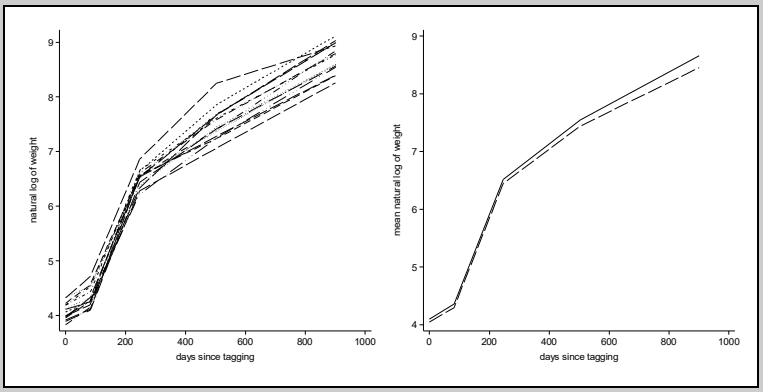

Fig. 23.1 Profile plot (left) and mean plot (right) for growth of salmon; fish with jaw deformities shown with dashed lines.

(*continued on next page*)

Example 23.1 (*continued*)

The profile plot shows fairly regular growth curves and some evidence of high within-subject correlation because subjects tend to remain high or low throughout the growth period (this visual phenomenon is sometimes called 'tracking'). The mean plot shows that on the average jaw-deformed fish grow slightly less than health fish, in agreement with our biological expectation. We finally also present in tabular form below the simple correlations (left), variances and covariances (right) of the 5 measures on the same fish (see Sections 20.1 and 23.2.3).

	Correlations					Variances/Covariances				
Sample	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
1	1					0.034				
2	0.878	1				0.031	0.036			
3	0.553	0.660	1			0.019	0.023	0.033		
4	0.439	0.507	0.681	1		0.021	0.025	0.032	0.068	
5	0.226	0.266	0.396	0.490	1	0.011	0.013	0.019	0.033	0.068

For example, the correlation between measurements at vaccination (sample=1) and harvest (sample=5) is 0.226, and the variance of measurements at vaccination is 0.034. We see that on log-scale the variances vary moderately over time, being larger in the last part of the growth period. As expected, the data show autocorrelation because correlations drop down as time between samplings increase. However, correlations between adjacent samplings also drop down over time, presumably in part due to the non-equidistant time points.

information about differences between subjects in different subpopulations; in addition, a longitudinal study can give information about changes in subjects over time. This is particularly important if we want to assess the impact of predictors that change over time. In a cross-sectional study, these can only be estimated from between-subject regressions, and in order to interpret them as changes within an individual, we would need to assume that the within-subject regression has the same slope, *ie* that the predictor has no contextual effect (Section 21.4).

In addition, longitudinal designs can be substantially more powerful statistically than cross-sectional designs for inference about within-subject predictors. This is analogous to the gain of a block design with treatments allocated within blocks, *eg* a cross-over design. For between-subject predictors, the cross-sectional design with its between-subject independence is the most powerful if the cost of sampling different subjects is not larger than the cost of repeatedly sampling the same subject, by the same reasoning as in Section 20.3.3.

23.2 UNIVARIATE AND MULTIVARIATE APPROACHES TO REPEATED MEASURES DATA

In this section, we will briefly review some relatively simple statistical procedures to deal with repeated measures data in regular between-subject designs. That is, we assume balanced and uniform series on all subjects, and consider inference about predictors at the subject level; the -fish_trial- data is an example of such a data structure. These methods are less commonly used than those in the following chapters, in part because of their demands on the data structure and because they may not fully use the information in the data. Nevertheless, they may serve as reference points for more complicated analyses, and could in some situations suffice to draw conclusions about the study hypotheses.

Example 23.2 Graphs and descriptive statistics for somatic cell count data
data = scc_40

The somatic cell count dataset is highly unbalanced with the number of observations per cow ranging from 2 to 11 (cows with only a single observation were excluded). Also, the actual time points (days in milk, -dim-) vary between cows and farms because -dim- is relative to the calving date and because farm visits were only roughly monthly. The variable -test- was constructed to give an approximate month of the lactation for each test, so as to make time points easier to compare between cows. Fig. 23.2 shows a profile plot for all the cows in one of the small herds (22 cows) in the dataset.

The plots shows considerable variation between cows, both in their actual values and in their pattern of time points. The shortest curves start from time 0 and extend only till time 1 or 2. Some of the curves are relatively stable over time, and general time effects are hard to discern. Missing values seem to occur either at the beginning or as drop-outs.

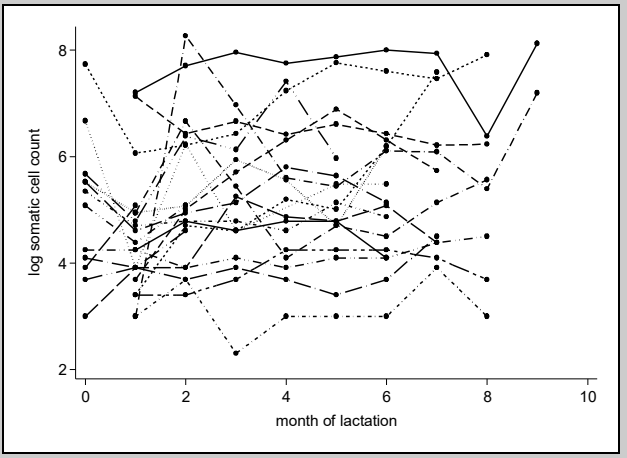

Fig. 23.2 Profile plot for log somatic cell counts of cows in a single herd

23.2.1 Univariate methods

We call these methods 'univariate' because they essentially avoid modelling the repeated measures structure by reducing the within-subject series of measurements to one (or several) statistics computed for each subject. With one observation (the computed statistic) per subject, a multivariable analysis involving any predictors and further hierarchical structure for the subjects follows the lines of previous chapters of the book.

The most basic procedure is analyses by **separate time points**. In the fish trial, one might argue that the most important measure from an economic point of view is the harvest weight and therefore analyse only the weights at the last time point. With the 2 categorical predictors vaccine and jaw, the analysis would be a linear model corresponding to a 2-way ANOVA (Chapter 14). The analysis would not be wrong, but it would be inefficient because all the preceding measurements are not used. If a similar analysis was carried out for each of the other time-points, it might become difficult to combine the conclusions from the different analyses.

As the data from different time points were treated separately, we would not know how strongly correlated they were, and then we could not tell whether a few significances at different time points strengthen the evidence of predictor effects, or whether we essentially just saw the same evidence several times. On the other hand, running the analysis at multiple time points could increase our Type II error, in particular if we fell victim to the temptation of selecting the time points with the 'best' effects. This discussion shows the need for a formal rule for managing any selection of time points for analysis; one general applicable rule is the **Bonferroni correction** for performing multiple analyses (Section 14.10.1) by dividing the significance level by the number of analyses performed or the number of time points considered for selection for analysis. This approach produces a valid analysis for separate time points but it is a weak analysis, in part due to the (conservative) Bonferroni correction. More importantly, the analysis does not use the longitudinal information in the data (the subjects at different time points could be different), and it does not describe or analyse the development over time. Example 23.3 demonstrates the approach applied to the fish trial data.

Analysis by a **summary statistic** (also, response feature or derived variable) is a refinement of the time point method, performed in 2 steps. In the first step, you choose a single quantity to calculate from each subject's profile, for example the gain from the first to last measurement. This again results in a single observation per subject on which we then carry out a between-subject analysis as above. The effectiveness of the approach depends on whether one can devise a good summary statistic that captures the relevant information inherent in the profiles; the choice is usually guided by inspection of profile plots. Some standard choices of summary statistics are: the subject mean or median, the within-subject slope, the gain, and the area under the curve (AUC). Summary statistics should generally be chosen to have interpretations of practical and/or scientific interest. They are not (primarily) based on statistical considerations; for example, the use of the within-subject slope from a regression does not require that the curve is modelled well by a straight line (or a statistical assessment of linearity). The slope can simply be used as a measure of average increase even if there is some curvature. In animal production, average daily gain (defined as the weight gain defined by the length of the growth period) is a standard growth measure, although the growth may show some non-linearity. We illustrate the use of summary statistics by the fish trial data in Example 23.3.

Advantages of the approach are its simplicity and flexibility, including its potential use for discrete data, and the direct access to features of interest that may be difficult to extract from complex models. Suitably chosen summary statistics can be both powerful and robust towards model assumptions and data irregularities (Everitt, 1995; Senn *et al*, 2000). Disadvantages are the subjective choice of the statistic, the loss of information by reducing each profile to a single statistic, and the limited information provided by the analysis (*eg* no correlations or predictions). Also, it is difficult to incorporate strong or key within-subject predictors into the approach.

23.2.2 Repeated measures ANOVA

Treating the repeated measures within subjects as a hierarchical structure leads to models with subject random effects. The simplest of such models, the random-intercept model, can in regular between-subjects designs be analysed with the ANOVA-based approach for mixed models (Section 21.5.1), and is sometimes termed the 'split-plot' approach to repeated measures data, referring to the link between hierarchically structured data and a split-plot design explained in Section 20.2.1. We saw in Chapter 21 that, in a random-intercept model, the

Example 23.3 Univariate methods for fish trial data
data = fish_trial

The measurements at each time point were analysed by a 2-way ANOVA linear model to test the effects of vaccines, jaw deformity as well as their interaction. The analysis at vaccination is mostly of interest to assess whether the randomisation of fish to vaccine groups created comparable groups. Therefore, the Bonferroni adjustment should involve the 4 subsequent time points; thus, P-values less than 0.05/4=0.0125 could be considered as significant. In addition, 2 summary statistics were explored: the gain (difference between log-weights at harvest and vaccination, therefore the log of the ratio between the corresponding weights) and the slope for periods 3-5 covering the long seawater growth period, and also motivated by the roughly linear profiles in Fig. 23.1. The table gives P-values for the effects of vaccines, jaw and the interaction.

P-values	Separate analysis at time point					Summary statistic	
Effect	1	2	3	4	5	gain	slope 3-5
Vaccine	0.261	0.147	0.667	0.107	0.004	0.081	0.025
Jaw	0.087	0.045	0.012	0.025	<0.001	<0.001	<0.001
Vaccine*Jaw	0.651	0.824	0.606	0.318	0.027	0.116	0.104

The analyses at separate time points showed significant effects only at harvest. The group means were given in Example 23.1. It is seen that weights at harvest are much lower for fish with jaw deformities, and that vaccine group 3 performs somewhat better than the other groups, in particular for fish with jaw deformities (but the interaction was not significant). Similar results were found in the summary statistics; the significant difference between vaccines was again caused by a steeper slope for vaccine 3.

correlations are the same (and positive) between all units within a cluster (*eg* Eq 21.4); in Section 23.3.1, we will call this a **compound symmetry** correlation structure. But we noted already in the introduction that we would expect autocorrelation to be present in repeated measures data, so a random-intercept model induces the wrong correlation structure! In essence, the hierarchical model fails because it does not take into account the time ordering of the repeated measures on each subject. For these reasons, the random-intercept model is by now considered inadequate for most repeated measures data analyses. It is, however, perfectly valid for data with only 2 repeated measures per subject, and may give a reasonable analysis for short series (with 3 or 4 time points) because compound symmetry may not be that far off in such cases. As the random-intercept model provides the simplest analysis of the full dataset, it is often used as a starting point (or reference) for further more complex models. It could also be used for first decisions about the modelling that are unlikely to require an accurate correlation structure, *eg* a choice of transformation of the outcome (although one would be advised to the reassess the transformation with the final model).

As the first of several approaches to assess (test) the assumed correlation structure, we describe the **repeated measures ANOVA** method for regular between-subjects designs. The aim of this method is to assess, and possibly adjust, the impact of the assumed correlation structure on the test statistics of the ANOVA table; thus, it is essentially a method for correcting test statistics but does little to adjust other features of the statistical inference such as standard errors on estimates. For this reason, and because of the design requirements (which *eg* imply that missing values cannot be managed in any easy way), the repeated measures ANOVA approach has largely been superseded by extensions of the mixed modelling (Section 23.3). We mention it here mainly because of the insights it offers into the impact of wrongly assuming a compound

symmetry structure. It can be shown (in regular designs) that a violation of the compound symmetry assumption affects only within-subjects effects (*ie* effects involving time) and makes the corresponding uncorrected test statistics of the ANOVA table too liberal (*ie* gives a too small P-value). Several correction factors exist to adjust (reduce) the degrees of freedom of the *F*-statistics to achieve approximately correct inference. We illustrate the procedure in Example 23.4. As already mentioned, the general consensus among statisticians is that the repeated-measures ANOVA approach offers no real advantages in return for its strong restrictions, and we therefore do not recommend it for general use.

23.2.3 Multivariate analysis

Multivariate statistical methods apply to data where the measurement on each subject, or experimental unit, consist of multiple records instead of a single record, as has been the case in previous chapters of the book. As an example, routine recordings in dairy production contain several records on each test day, including milk yield, fat percentage and somatic cell count. Such multiple records are usually compiled into a vector of observations on each subject (in the example, a 3-dimensional vector), so we can think of **multivariate data** as consisting of vectors of observations instead of single observations per measurement. A large body of theory and methods exist for multivariate data, but we concentrate on how they can be applied to repeated measures data (*eg* Davis (2002), Chapters 3-4). Repeated measures on the same subject may be considered as a single observation (vector), consisting of the entire set of values across time points. Let us introduce a bit of notation to support the idea: Y_{ij}=measurement for subject i at time j, where there are m time points j=1,...,m. Then, in a multivariate framework, the basic observation for subject i is the vector Y_i=(Y_{i1},...,Y_{im}). Multivariate linear models extend the usual linear models (Chapter 14) by modelling the observation vector in terms of its mean (vector) and variance (matrix). The mean vector consists of the mean outcome at different time points, and the (co)variance matrix consists of the variances at and the covariances between the different time points. It is more common to refer to the latter matrix as the **covariance matrix** (sometimes also variance-covariance matrix), so we'll use that term. Also, the covariances are more intuitive to interpret when rescaled as correlations (for the relation between covariance and correlation, see Eq 20.1). For a set of measurements (Y_1,...,Y_m) on the same subject (where we for simplicity suppress the subject indicator i), the covariance matrix cov(Y) and the correlation matrix corr(Y) are the (mxm)-matrices holding all the covariances, or correlations, between pairs of measurements:

$$\mathrm{cov}(Y)=\begin{pmatrix} \mathrm{var}(Y_1) & & & & \\ \mathrm{cov}(Y_1,Y_2) & \mathrm{var}(Y_2) & & & \\ \mathrm{cov}(Y_1,Y_3) & \mathrm{cov}(Y_2,Y_3) & \mathrm{var}(Y_3) & & \\ \vdots & \vdots & \vdots & & \vdots \\ \mathrm{cov}(Y_1,Y_m) & \mathrm{cov}(Y_2,Y_m) & \mathrm{cov}(Y_3,Y_m) & \cdots & \mathrm{var}(Y_m) \end{pmatrix} \qquad \textit{Eq 23.1}$$

$$\mathrm{corr}(Y)=\begin{pmatrix} 1 & & & & \\ \mathrm{corr}(Y_1,Y_2) & 1 & & & \\ \mathrm{corr}(Y_1,Y_3) & \mathrm{corr}(Y_2,Y_3) & 1 & & \\ \vdots & \vdots & \vdots & & \vdots \\ \mathrm{corr}(Y_1,Y_m) & \mathrm{corr}(Y_2,Y_m) & \mathrm{corr}(Y_3,Y_m) & \cdots & 1 \end{pmatrix} \qquad \textit{Eq 23.2}$$

The matrices are symmetric, so for clarity, the values above the diagonal have been left blank.

In Example 23.1, we displayed the covariance and correlation matrices for the 5 repeated measures of logarithmic weight in the fish-trial data.

Multivariate analysis of variance (MANOVA) assumes normally distributed multivariate outcomes with a mean modelled in terms of subject-level predictors and a covariance without any specific structure (although covariances may be assumed either constant or heterogeneous across predictor groups). Thus, the model defaults to heterogeneous variances across time points and makes no assumptions about the correlation structure. The analysis provides estimates of means (with SEs), variances and correlations, and these can be used to test hypotheses about the effects of between-subject predictors and time. Different test statistics exist for the same hypotheses (although for simple hypotheses they'll coincide); Wilk's lambda is a sensible overall choice, with reference F-distributions. The standard test statistics offered by MANOVA software do not include hypotheses related to time (because in general multivariate analysis, there is no structure among the multivariate responses), so these may need to be set up manually by specifying suitable contrasts; for details consult (*eg* Davis C (2002), Chapter 3-4). The multivariate analysis is similar to analysis by a mixed model with unstructured correlations (Section 23.3.2), which (with suitable statistical software) gives easier access to specific contrasts and tests. Example 23.4 gives results from both the multivariate and mixed-model analysis for the fish-trial data.

One advantage of the multivariate approach is that it avoids any problems with a wrongly specified correlation structure (as in the random-intercept model). However, this advantage also contains the potential drawback that estimation of many covariance parameters (all the variances and correlations) may be ineffective or outright impossible, especially with long series of measurements. Davis (2002, Chapter 6) cites simulation studies with small/moderate number of subjects that have shown the multivariate approach to provide exact and better statistical inference than analysis by the mixed model. The main drawbacks of the multivariate approach as presented here are its strong requirements: normally distributed, balanced data with uniform time points, no missing values and no within-subject predictors. However, one software implementation (MLwiN) relaxes all these conditions and also allows for additional hierarchical structure (Rasbash *et al*, 2008, Chapter 14).

23.3 LINEAR MIXED MODELS WITH CORRELATION STRUCTURE

Having noted in Section 23.2.2 the deficit of the simplest linear mixed model, the random-intercept model, for repeated measures data, we will discuss here 2 ways of extending the model to incorporate more realistic correlations for continuous repeated measures data. With both of these extensions (in Sections 23.3.2-3) the model would still be termed a linear mixed model, so the important details lie in the actual specification of the model. Also, the advantages of the linear mixed model, such as its flexibility to handle hierarchical structure and predictors at multiple levels as well as its likelihood-based inference and resulting robustness to missing values as long as these are missing at random (Section 15.5), will remain intact. These are some of the distinct advantages of the linear mixed model approach over the simpler approaches reviewed so far. In this section, we will not revisit the entire analysis of the linear mixed model from Chapter 21 but concentrate on describing the 2 extensions of the model and their impact on the analysis.

Example 23.4 ANOVA and MANOVA analysis of fish-trial data
data = fish_trial

We continue the analysis of the growth of salmon from Examples 23.1 and 23.3 by showing results from a random-intercept model, its repeated measures ANOVA adjustment, a multivariate (MANOVA) analysis and a mixed model with unstructured covariance matrix (Section 23.3.2). Due to the balancedness of the data, the parameter estimates for all models are the means for the vaccine*jaw*sample combinations shown in Example 23.1. The table below gives P-values for fixed-effects hypotheses. The estimated within- and between-subject variances in the random-intercept model were 0.0246 and 0.0216, respectively, corresponding to an *ICC* of 0.47. The repeated measures ANOVA gave an estimated Huyhn-Feldt correction factor of ε=0.66 (where ε=1 means that no adjustment to test statistics is required because of violations of the assumed correlation structure of the random-intercept model), and adjusted the F-distribution degrees of freedom by multiplication with this factor; for example, the adjusted degrees of freedom for sample*vaccine were (ε*12, ε*1568)=(8, 1035).

P-values	Model / Method			
Effect	Random Intercept	rep. meas. ANOVA	MANOVA	mixed, unstruct.
vaccine	0.075	0.075	0.003*	0.080
jaw	<0.001	<0.001	0.000*	<0.001
vaccine*jaw	0.575	0.575	0.011*	0.575
sample	<0.001	<0.001	0.000	<0.001
sample*vaccine	0.004	0.013	0.006	0.006
sample*jaw	<0.001	<0.001	0.006	0.006
sample*vaccine*jaw	0.006	0.017	0.014	0.014

*simultaneous test across all time points

The estimated error covariance matrix for the MANOVA and the mixed model were identical and very close to the values shown in Example 23.1 (not shown). Contrasting the correlations with the single estimated correlation (*ICC*) of the random-intercept model elucidates how far the random-intercept model is off the actual correlation structure. However, the impact on test statistics was marginal. Note that the 3 first MANOVA tests are for different hypotheses (involving all time points simultaneously) and cannot be compared to those of the other methods.

The conclusion about the effects of vaccine and jaw deformity is that significant differences do exist, and that these change over time. The overall effect of jaw deformity was already discussed in Example 23.1, and the analysis here only established its strong significance. The means in Example 23.1 also show how the impact of jaw deformity varies with vaccine groups over time: at samples 4 and 5, some but not all, vaccine groups show an effect of jaw deformity. We could use multiple comparisons to make statements about which differences are statistically significant, but we won't pursue that here.

23.3.1 Correlation structure

Before we proceed to the model extensions, we will distinguish more precisely than we have done so far between the different correlation structures of importance for the modelling. First, despite our usage of the term 'correlation structure', we really mean the structure of the covariance matrix, because the variances are equally part of the structure we're modelling

(however, 'covariance structure' seems less intuitive than correlation structure). Second, the correlation structure targeted by the modelling, and whose violation may affect the inference, is that of the errors, not the structure of the observed data. The difference between the 2 is that, in the former, the fixed and random effects of the model have been estimated and adjusted for (eliminated) from the correlations. We discuss the impact of fixed and random effects in turn.

If the fixed effects are strong, the crude and adjusted correlations can be appreciably different. We noted in Example 23.4 that the crude correlations were very similar to those estimated for the errors of the model; in these data, the fixed effects were indeed fairly small (within each time point). In order for fixed-effects predictors to account for some of the anticipated autocorrelation in repeated measures data, they must include a within-subject predictor that itself shows autocorrelation. A possible example is inclusion of milk yield (kg) as a predictor in a model for somatic cell counts throughout the lactation. Realistically, though, it is not common that the fixed effects eliminate substantial parts of the autocorrelation.

Random effects of random-intercept type can, as we have seen, only induce compound symmetry correlations. Random slopes however can induce autocorrelation if the predictor involved is correlated with time. One obvious candidate for a random slope with potential to induce autocorrelation is therefore ... time! Adding a random slope with time to a random-intercept model induces autocorrelation, and may therefore remove autocorrelation from the errors. Linear mixed models with random slopes for time are also called **trend** models, and constitute one of the 2 extensions of the random-intercept model to deal with autocorrelation (Section 23.3).

We will next describe a range of correlation structures that can exist either for repeated measures data or for the errors in a model of such data. Table 23.2 lists some of the more common correlation structures for repeated measures in the case of m=4 repeated measures on the same subject. For simplicity, we show only the correlation matrix in all cases except the last one but, if variances are assumed to be equal (σ^2), the covariances are simply the correlations multiplied by σ^2.

The first 2 correlation structures are well known and included mainly to familiarise the reader with the display. Recall that the correlation ρ in the **compound symmetry** structure induced by a random-intercept model can be expressed in terms of the variance components σ^2_h and σ^2 as $\rho=\sigma^2_h/(\sigma^2_h+\sigma^2)$ (Eq 21.4). The alternative name, an **exchangeable** structure, refers to the fact that since correlations are the same all over, the units (here the time points, but in our hierarchical models the animals within a cluster) can be interchanged (or exchanged) without affecting the structure.

The simplest structure showing the desired decay in correlation with increasing distance between observation is first order **autoregressive**, or **ar(1)**, in terminology originating from time series analysis (Section 14.11). It involves 2 assumptions: that all pairwise correlations a certain number of time steps (or points) apart are correlated to the same degree, and that the correlations decay as powers of the number of time steps that separate 2 observations. The first assumption implies for example that the correlation in the observation pairs (1,2), (2,3) and (3,4) are all the same (and equal to ρ). The assumption is sometimes called **homogeneous** or stationary correlations (discussed further below). **Note** The decay of ar(1) correlations is quite rapid; *eg* for ρ=0.5, observations 4 time steps are close to uncorrelated (0.5^4=0.063).

More complex correlation structures than ar(1) are often useful as well, in order to incorporate either a slower or less consistent decay of correlations with time distance. The **arma(1,1)**

structure, also originating from time series analysis, and the **Toeplitz** structures accomplish this with 1 and (m-2) additional parameters, respectively. The choice (and nomenclature) of correlation structures available for modelling depends on the statistical software; many additional and more complex structures exist, but the 3 homogeneous structures mentioned here are usually always available. **Homogeneous** structures are most meaningful if the time points are equidistant. In some situations, when processes occur at different speeds in different stages of the time period considered, one could perhaps argue non-equidistant time points to be 'biologically equidistant' (*ie* that they should have the same impact). For example, if one studies the impact of the injection of a pharmaceutical into animals, one may choose to measure the response at follow-up times 1, 2, 5, 10, and 30 minutes post-injection. If the biological processes happen much more quickly in the initial phase after injection, it may still be meaningful to assume homogeneous correlations. Clearly such reasoning may be difficult to justify rigorously.

An **unstructured correlation structure** will let the data speak for themselves; we already saw their use in the fish-trial data (Example 23.4). Their drawback is that, with a long series of repeated measures, the number of parameters involved grows so large that they become difficult to estimate and interpret. Another question is whether heterogeneous variances across the time points should be assumed. All correlation structures have a corresponding version with heterogeneous variances (but it may not be implemented in the software). There are often good biological reasons why variances should not be assumed constant over time; on the other hand, heterogeneous variance structures will also increase the number of parameters appreciably for a long series.

Table 23.2 Repeated measures correlation structures for four repeated measures/animal

Name	Correlation structure	Interpretation
uncorrelated or independent	$\mathrm{corr}(Y)=\begin{pmatrix}1&&&\\0&1&&\\0&0&1&\\0&0&0&1\end{pmatrix}$	uncorrelated (for normal data: independent) observations
compound symmetry, or exchangeable	$\mathrm{corr}(Y)=\begin{pmatrix}1&&&\\\rho&1&&\\\rho&\rho&1&\\\rho&\rho&\rho&1\end{pmatrix}$	hierarchical, mixed model (same correlation between all pairs of observations)
ar(1), or first order autoregressive	$\mathrm{corr}(Y)=\begin{pmatrix}1&&&\\\rho&1&&\\\rho^2&\rho&1&\\\rho^3&\rho^2&\rho&1\end{pmatrix}$	repeated measures or time-series model with power decay of correlations
arma(1,1), or first order autoregressive moving average	$\mathrm{corr}(Y)=\begin{pmatrix}1&&&\\\gamma&1&&\\\gamma\rho&\gamma&1&\\\gamma\rho^2&\gamma\rho&\gamma&1\end{pmatrix}$	extended repeated measures or time series model with power decay
Toeplitz, or stationary	$\mathrm{corr}(Y)=\begin{pmatrix}1&&&\\\rho_1&1&&\\\rho_2&\rho_1&1&\\\rho_3&\rho_2&\rho_1&1\end{pmatrix}$	repeated measures with unconstrained correlations at different spacings
unstructured	$\mathrm{corr}(Y)=\begin{pmatrix}1&&&\\\rho_{12}&1&&\\\rho_{13}&\rho_{23}&1&\\\rho_{14}&\rho_{24}&\rho_{34}&1\end{pmatrix}$	repeated measures with entirely unconstrained correlations
unstructured with inhomogeneous variances, or non-stationary	$\mathrm{cov}(Y)=\begin{pmatrix}\sigma_1^2&&&\\\sigma_{12}&\sigma_2^2&&\\\sigma_{13}&\sigma_{23}&\sigma_3^2&\\\sigma_{14}&\sigma_{24}&\sigma_{34}&\sigma_4^2\end{pmatrix}$	repeated measures, unconstrained variances and correlations

One note of caution about correlation structures: you need to ensure that the time points are properly understood by your software, in particular if the data contain incomplete series (due to missing values). For example, it makes a difference with most correlation structures whether recordings were taken at times (1,2,3,4), at times (3,4,5,6) or at times (1,2,5,6). If the time points are not uniform across subjects (with allowance for missing values), at least to a good approximation, the correlation structures will not be meaningful across the dataset, and modelling based on certain fixed correlation structures will be misleading. Such data structures therefore raise the need to incorporate into the matrices the actual recording times.

For non-equidistant repeated measures or spatial data, denote by $d_{jj'}$ the distance between observations j and j'. For longitudinal data where locations correspond to time points, the $d_{jj'}$ would be the (absolute) difference between the recording times of observations j and j', and for spatial data the distances would be actual physical distances (*eg* between herds). Table 23.3 lists

some examples of correlation structures defined from such distances. The structures are **isotropic** when only the distances, not the actual locations for observations j and j', are used. The power (or exponential) structure is the extension of the ar(1) structure to non-equidistant time points; the parameter ρ equals the correlation between 2 observations one unit apart.

Table 23.3 Spatial (or non-equidistant repeated measures) correlation structures

Name	Correlation structure	Interpretation
power, or exponential	$\text{corr}(Y_j, Y_{j'}) = \rho^{d_{jj'}} = \exp(-d_{jj'}/\theta)$	power decay with distance; note the relationship: ρ=exp(-1/θ)
power, or exponential, with nugget effect	$\text{corr}(Y_j, Y_{j'}) = \frac{\sigma^2}{\sigma^2+\sigma_0^2}\rho^{d_{jj'}}$	power decay with distance, close observations not fully correlated
Gaussian	$\text{corr}(Y_j, Y_{j'}) = \exp(-d_{jj'}^2/\theta)$	exponential-quadratic decay with distance
linear	$\text{corr}(Y_j, Y_{j'}) = \begin{cases} 1-\rho d_{jj'} & \text{if } \rho d_{jj'} < 1 \\ 0 & \text{if } \rho d_{jj'} \geq 1 \end{cases}$	linear decay with distance

23.3.2 Linear mixed models with complex correlation structure

Recall that, in the linear mixed model from Chapter 21 (Eq 21.8):

$$Y = X\beta + Zu + \varepsilon \qquad \textit{Eq 23.3}$$

we assumed the components of ε to be independent, and modelled the hierarchical structure using the random effects in the Zu part of the model. In order to enable complex correlation structure, in particular autocorrelation, we will now allow dependence corresponding to a particular correlation structure within some sets of ε-values. In the repeated measures context, each set contains all the repeated measures for a subject, and in the spatial context, each set contains a particular group of observations for which we want to model a spatial correlation (*eg* herds within a certain region).

In such mixed models with correlation structure, both the random part Zu and the error correlation structure contributes to total (co)variance (not explained by the fixed effects). If random effects are specified at the same level as the error correlation structure, *eg* subject random effects and within-subject correlation structure, the resulting model may be difficult to estimate and in worst cases even be overparameterised. To illustrate the problem, random effects (intercepts) for subjects cannot be fitted in a model with compound symmetry correlation structure for subjects. This is because both parts of the model will lead to the same correlation structure, so only one of them is needed. Random effects for subjects can however be combined with an ar(1) structure; this produces a structure with autocorrelations that does not decay to zero but instead to the *ICC* one could compute from the between- and within-subject variances. Similarly, a Toeplitz structure cannot be combined with subject random effects, and, if unstructured correlations are specified, it is pointless to include any random effects at the subject level (either a random intercept and random slopes). If the model in Eq 23.3 has no random effects, it is perhaps misleading to call it a mixed model; the name **covariance pattern model** is also used (Hedeker & Gibbons, 2006, Chapter 6).

The statistical analysis of the 'extended' mixed models evolves along the same lines as previously discussed, only with additional variance parameters to be estimated. For large

structures (with many time points), parsimonious models for cov(Y) are recommended unless the number of subjects is very large, to avoid overspecification of the model and unexpected impacts of the covariance structure on the fixed-effects parameters. The choice of correlation structure can be formalised by using likelihood-ratio statistics to test nested correlation-structure models against each other. For example, the compound symmetry and ar(1) models can be tested against both arma(1,1) or Toeplitz models, but they cannot be tested against each other. The test of a compound symmetry model against any of these models allowing for autocorrelation is one of the best ways to assess whether the compound symmetry model is inadequate due to unmodelled autocorrelation. The fit of models with the same number of parameters can be compared by their log-likelihood values (the higher log-likelihood model is generally preferred). Model selection criteria such as the AIC (Chapter 15) are also applicable here. A visual assessment of the residual autocorrelation function may also be helpful (employed *eg* in Vigre *et al* (2009)). Example 23.5 examines different correlation structures for the full SCC data.

Mixed models with complex correlation structure are currently only available in a few software packages: SAS (Littell *et al,* 2006) and R/S-Plus (Pinheiro & Bates, 2000), and the recently released version 11 of Stata includes some of these models as well.

23.3.3 Trend models

Let us first recap from Section 23.3.1 that trend models are characterised by having random slopes of time. We already argued in favour of these random slopes because they would introduce autocorrelation into the model. It could also be said that models assuming all subjects develop in the same way over time are unrealistic for most longitudinal data (Hedeker & Gibbons, 2006), Chapter 4). With the inclusion of subject-random slopes for time, the model includes terms representing the development over time at the population level (the fixed effects for time) as well as the random effects representing the development over time at the individual level.

We need to be a bit more specific when it comes to how time should be modelled. The simplest option is a linear effect of time but it is often too simplistic to assume a linear change across time; for example, linear trends may eventually level off towards a plateau or a minimum level. The choice of an appropriate form of the time effects follows the same principles as for other continuous predictors (Section 15.4). Ideally, for consistency, the same (non-linear) relation with time would be used for the fixed and random effects. As was noted in Section 21.3, with the need for some parsimony in our use of random slopes, it becomes attractive to consider models with time effects represented by only a few parameters. In some situations, a non-linear monotone transformation (*eg* log or square-root) of the time scale can help to achieve an approximately linear relation. If the effects of time need multiple parameters, it is helpful for estimation and interpretation of variance parameters if these are as unrelated ('independent') as possible. For example, if a polynomial model is used, it is recommended to parameterise it by orthogonal polynomials (Hedeker and Gibbons, Chapter 5). If it is not possible to model time effects in a simple fashion, one may choose a simpler form (*eg* linear) for the random slopes while retaining a more complex form for the fixed effects. An alternative random part modelling of time has independent fluctuations around the fixed effects for each time point and subject (essentially, a random interaction between subject and time). It avoids the assumption that subjects are on parallel trajectories over time, but it will not introduce autocorrelation into the model and is, therefore, usually less attractive.

Example 23.5 Linear mixed models with correlation structure for log somatic cell counts

data = scc_40

Several correlation structures were examined for the full 40-herd somatic cell count data, using the same fixed effects as in the examples of Chapter 21. The model is now a 3-level model with herd-random effects and within-cow correlation structure.

	Correlation	Estimated ρ			-2 ln
Correlation structure	parameters	1 month	2 months	3 months	likelihood
compound symmetry	1	0.541	0.541	0.541	39004.73
ar(1)	1	0.661	0.437	0.289	38563.57
non-equidistant power	1	0.673	0.453	0.305	38574.30
arma(1,1)	2	0.657	0.578	0.509	37802.21
Toeplitz	10	0.657	0.578	0.512	37795.72

The table above illustrates how the different structures adapt to the data. In terms of statistical significance, the Toeplitz model is no better than the arma(1,1) model (the likelihood-ratio test statistic equals 6.49 with 8 df), which in turn is clearly preferable to the structures with only one correlation parameter. The estimated correlations for tests one, 2 and 3 time steps (for the non-equidistant structure each considered equivalent to 30 days) apart demonstrates the deficiencies of the one-parameter models. The compound symmetry structure does not allow for autocorrelation, and the autoregressive-type structures, on the other hand, produce too rapidly decaying correlations.

For comparison with the results of the 2-level data and a subsequent analysis by GEE procedures, we also present a table of estimates for the fixed effects and random parameters from the arma(1,1) model. The fixed-effects estimates and SEs were very close for all the models considered above.

	Coef	SE	t	P	95% CI	
shsize	0.627	0.306	2.05	0.047	0.009	1.245
heifer	-0.777	0.040	-19.22	0.000	-0.857	-0.698
season = spring	0.034	0.022	1.54	0.125	-0.009	0.078
season = summer	0.039	0.027	1.57	0.117	-0.010	0.087
season = fall	-0.007	0.023	-0.32	0.752	-0.052	0.037
sdim	0.328	0.014	24.08	0.000	0.301	0.354
constant	5.283	0.060	-	-	5.163	5.402

In addition, the estimated correlation parameters and variance components (also with SEs) were:

$$\gamma = 0.657(0.008), \rho = 0.880(0.006), \text{ and } \sigma_h^2 = 0.104(0.028), \sigma^2 = 1.378(0.027).$$

The parameter ρ has the same interpretation as in the ar(1) structure as the factor by which the correlation drops for each additional month, and γ is the correlation for observations one month apart.

For further details about trend models, we refer to the general sections on random slopes (Section 21.3) and inference for mixed models (Section 21.5). We already showed an example of a model with a linear random slope for time in the 2-level somatic cell count data (Example 21.4), but as the data contained only a single test on each cow, there were no repeated

measures. In Example 23.6, we give the results of fitting random slopes for time at both the cow and herd levels. Models with more complex fixed and random effects have been fitted to the complete dataset from which the scc_40 data were extracted (Stryhn *et al*, 2001).

23.4 MIXED MODELS FOR DISCRETE REPEATED MEASURES DATA

From the relative ease with which the linear mixed model could be extended to incorporate autocorrelation one might expect things to be similar for discrete data, but that is not so. After explaining the challenges of adding correlation structure to a GLMM, we give pointers to some of the many different approaches that have been tried, describe in more detail the concept of a **transitional** model, and illustrate by the scc_40 data the impact this approach and the trend model (from the previous section) has on estimates from a random-intercept logistic regression model.

23.4.1 Adding correlation structure to a GLMM

The linear mixed model approach of incorporating correlation structures into the model's error component (ε) runs into the serious problem in GLMMs that the linear predictor (*eg* Eqs 22.1 and 22.5) does not contain an error component! The reason is that a GLM(M) models the mean and variance on different scales: the mean on the scale of the linear predictor given by the link

Example 23.6 Linear trend model for log somatic cell counts
data = scc_40

The random-intercept model (with compound symmetry correlation structure) of Example 23.5 was extended with linear random slopes for -sdim- at both the cow and herd levels. The table shows estimates for the random parameters (for comparison including also those of the random-intercept model); the fixed-effect estimates and SEs were similar to those of Example 23.5.

Level	Parameter	Random-intercept model Estimate (SE)	Random slopes for -sdim- Estimate(SE)
herd	variance (interc.)	0.101 (.028)	0.104 (.028)
	variance (slope)	-	0.006 (.003)
	covariance	-	-0.016 (.007)
cow	variance (interc.)	0.750 (.026)	0.760 (.027)
	variance (slope)	-	0.173 (.011)
	covariance	-	-0.030 (.012)
test	variance	0.637 (.008)	0.532 (.007)
	-2log likelihood	38986.31	38251.97

The random slopes are significant at both levels; this is obvious from the estimates at the cow level and the huge improvement in the log-likelihood, and omitting the herd-level random slope resulted in an about 7-unit increase in the log-likelihood. Still, the fit of the model is not as good as for several of the models with complex correlation structure in Example 23.5. One might consider combining the 2 approaches, but such models become more computationally difficult to fit so we won't go any further here.

function (in short, the link scale), but the variance on the observation scale. As the error term is on observation scale, it is subject to the restrictions related to the discrete outcome (see *eg* Section 16.1). A second problem following from this is the separation of correlation into parts explained on different scales; recall that in linear mixed-models correlation could be split between the random effects and the error correlation structure (Section 23.3.2), but the situation is more complex when these are on different scales. The third problem is that modelling clustering on the link scale yields parameters with a cluster-specific (SS) interpretation, whereas modelling clustering on observation scale (*eg* in a beta-binomial model, Section 22.4.5, or by the generalised estimating equations, Section 23.5) yields parameters with a population-average (PA) interpretation. If clustering is modelled on both scales, it is not clear which interpretation the parameters will have.

It is therefore more difficult to incorporate the correlation structures discussed for linear mixed models into a GLMM, and this is one of the reasons why no general GLMM-type class of models exists for repeated measures and spatial structures. Instead, models are, to a large extent, developed specifically for the most interesting data types: binary and count data. The literature in this field is large, technical and largely beyond the scope of this book. We introduce a few of the ideas that tie in with the GLMM framework.

The random-intercept model (Eq 22.1) includes a single random effect for each cluster. As this will not suffice to create a within-cluster correlation structure (Diggle *et al*, 2002, Chapter 11) expanded the model on link scale by including random effects for each time point (for each subject). In a binary model, with probabilities p_{ij} for subject i at time point j, the extension of Eq 22.1 therefore takes the form:

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 X_{1ij} + \ldots + \beta_k X_{kij} + u_{ij}, \quad \text{with} \quad u_{ij} \sim N(0, \sigma^2) \qquad \textit{Eq 23.4}$$

The idea is now to assume the set of random effects on each subject, (u_{i1},...,u_{im}), to be autocorrelated, *eg* according to the ar(1) structure with correlation ρ (note that ρ is the correlation between the random effects, not between the binary outcomes). In the special case ρ=1, the random effects will be perfectly correlated and thus identical, so that we're back in the random-intercept model with a single random effect (u_i). Unfortunately, the model is difficult to estimate (the MCMC methods of Chapter 24 were suggested as an option). The same random-effects structure has also been applied to autocorrelated count data, *eg* in a times series of counts (Davis *et al*, 2000).

Quasi-likelihood or pseudo-likelihood estimation software may allow specification of a repeated measures or spatial model for the adjusted variate computed in each step of the iteration (Section 22.5.2). This will lead to correlation structures of repeated measures or spatial type (Gotway & Wolfinger, 2003), although covariance parameters specified in this way may have no direct interpretation in the discrete model. Molenberghs and Verbeke (2005, Chapters 8 and 22) discussed the approach (as implemented in SAS, Proc Glimmix), and it was also included among the methods assessed in recent simulation studies (Masaoud & Stryhn, 2009). Some of the conclusions were: (i) if only a correlation structure is used, the procedure yields estimates with a PA interpretation and is comparable to GEE estimation; (ii) if both random effects and a correlation structure is used, the estimates will be intermediate between PA and SS parameters, and thus be biased for both interpretations.

A GLMM with a correlation modelled at the original scale (Barbosa & Goldstein, 2000) and a multivariate multilevel logistic model (Yang *et al*, 2000) have been developed, but both require specialised software (MLwiN macros), and do not appear to have been used much. A

multivariate model for discrete outcomes is also available in MLwiN and can be used for repeated measures data (Rasbash *et al*, 2008); without additional hierarchical structure, the parameters have a PA interpretation.

Repeated measures of counts have been modelled with random effects by a variety of approaches (Nelson & Leroux, 2006), including also the extensions of the Poisson regression reviewed in Chapter 18 (*ie* zero-inflation (Min & Agresti, 2005); overdispersion (Molenberghs *et al*, 2007) and the transitional models to be described next (Li *et al*, 2007).

23.4.2 Transition models

A generally accepted classification of modelling approaches for clustered data (including repeated measures) is into 3 types: subject-specific, marginal, and transitional (Diggle *et al*, 2002), Chapter 7; (Schukken *et al*, 2003). Our discussion has until now focused on the first 2, but we will here outline the third approach and explain how it can be used to incorporate autocorrelation into a GLMM. To focus on the basic idea, we consider the simplest case of a binary outcome.

In random-effects models, we accounted for the within-subject clustering by modelling the probability of the event for subject i at time point j conditionally on the (latent) subject random effect u_i, but it might seem more intuitive to model the probability conditionally on the previous event $Y_{i,j-1}$, and perhaps further events before that. A one-lag transition model (conditioning only on the previous event) could be expressed using the notation of Eq 22.7 as:

$$\text{logit}(p_{ij}) = (X\beta)_{ij} + (Zu)_{ij} + \gamma Y_{i,j-1}, \qquad \textbf{\textit{Eq 23.5}}$$

where only the transitional term $\gamma Y_{i,j-1}$,is new; note, it is no misprint that the outcome Y is present on the right hand side of the equation! The fixed-effect parameter γ equals the log *OR* for a comparison between subjects who at the previous time did and did not experience the event. The model in Eq 23.5 still includes subject random effects because the transitional term cannot be expected to account for all within-subject clustering. Conversely, even if we expect the transitional term to pick up autocorrelation, there may be still be some unmodelled autocorrelation left in the data. A transitional term is sometimes used informally in this way to capture autocorrelation in the data (Thurmond *et al*, 2005).

In Eq 23.5, the probability of an event at time j is different for a preceding non-event ($Y_{i,j-1}$=0) and a preceding event ($Y_{i,j-1}$=1), so essentially the model fits an equation for both of these 2 situations. If a disease event occurring after a non-event is interpreted as a new case, the former situation corresponds to **incidence**, and the probability of a disease event following an event would then is interpreted as 1 minus the **cure rate**. In other words, the model in Eq 23.5 is really for the 2 transitions: 0→1 (new case), and 1→0 (cure). With this interpretation, it may seem awkward that the impact of our predictors is assumed to be equal for both transitions; it means that the predictors have numerically exactly opposite effects for incidence and cure. In order to avoid this assumption (which should certainly not be considered the default), we can add interaction terms between $Y_{i,j-1}$ and the predictors. Similarly, we may want to include a random slope for $Y_{i,j-1}$. The regression parameters in models such as Eq 23.5 (with or without added interactions) are different than those in the marginal and cluster-specific equations (Eq 22.7), and no general conversion formula exists between them (Diggle *et al*, 2002, Chapter 7). Another issue that distinguishes a transitional from a usual random-effects model is that a special handling is needed for the first time point (j=1), where no previous outcome is available

as a predictor. The values for j=1 may either be omitted, or if they are included, the predictors should contain a dummy variable for this time point and Y_{i0} should be set to zero. In both cases, the model/data will not be quite the same as in the usual random-effects model which will lead to further differences in the parameters. We demonstrate the transitional model in Example 23.7 in the next section.

23.4.3 GLMMs without explicit correlation structure

Although our focus has been on alternatives to the random-intercept model with subject random effects, this simpler model may still be valid, provided one is willing to accept its lack of autocorrelation. It was indeed the type of model used for the conception data from Reunion Island in Chapter 22. To detect violations of compound symmetry may require much more data than in the continuous case because the information content is lower in discrete data—something that certainly is true for binary observations. (**Not**e We may think of the correlation structure as compound symmetry, although strictly speaking the within-subject correlations are only constant when the fixed effects are constant (because the variance is a function of the mean), and in a repeated measures model one would usually have time as a fixed effect.) However, a recent simulation study on binary data (Masaoud & Stryhn, 2009) concluded that even with a repeated measures series as short as m=4, biases may result from ignoring autocorrelation generated by *eg* the model in Eq 23.4. An earlier study cautioned against using the random-intercept model in the presence of autocorrelation (Heagerty & Kurland, 2001) based on a theoretical assessment of the bias in the estimates and a simulation study with m=5 from the same model.

In Example 23.7, we illustrate how the 2 main approaches for modelling correlation structure, random slopes for time (trend models, Section 23.3) and transitional models, affect the estimates of a large binary repeated measures dataset. We use again the scc_40 data, and define a binary outcome (-highscc-) by a threshold of 200,000 cells/ml for the somatic cell counts. Samples exceeding this cut-off may be considered as indicative of subclinical mastitis (Dohoo & Meek, 1982), although such a rule inevitably leads to some misclassification.

23.5 GENERALISED ESTIMATING EQUATIONS

The previous chapters have presented mixed models as an approach for dealing with the problem of clustering (lack of independence among observations) in a dataset. As noted, these mixed models are very flexible and can handle any number of levels of hierarchical clustering as well as more complex data structures. However, some unresolved issues remain. As discussed in Section 23.4, the mixed model approach is not as successful with repeated and spatial structures for discrete data as it is for continuous data. Also, its assumption of normally distributed random effects is perhaps a limitation; in practice, you will encounter data that clearly do not conform to that assumption. From a more philosophical point of view, one might argue that, in our analyses, we should only make the absolutely necessary distributional assumptions and for ‘nuisance effects’, rely on robust procedures that are less affected by the peculiarities of the data. This would follow the trend in modern statistics toward non- and semi-parametric procedures, as seen for example in survival analysis. Finally, complex mixed models are sometimes difficult to fit due to the size of the data or to numerical difficulties.

Example 23.7 Generalised linear mixed models for high somatic cell counts
data = scc_40

Out of the 14,357 test results, 5,653 (39.4%) exceeded the cut-off of 200,000 cells/ml. Among the 2,078 cows, 1,032 (49.7%) had constant values of -highscc- throughout the lactation, suggesting a strong within-cow clustering. In the next table, we compare estimates from a random-intercept model, a simple transitional model (without any interactions), and a linear-trend model with a random slope for -sdim-. The effective dataset for the transitional model excluded all (in total, 2,145) tests where the previous test was missing or did not exist (*ie* at the first time point).

Model Variable	Random interc. Estimate (SE)	Transitional Estimate (SE)	Linear trend Estimate (SE)
shsize	1.367 (.855)	1.012 (.565)	1.559 (1.005)
heifer	-2.227 (.137)	-1.409 (.105)	-2.458 (.160)
season=spring	0.016 (.076)	0.094 (.079)	0.078 (.091)
season=summer	0.071 (.079)	0.144 (.083)	0.099 (.010)
season=fall	0.042 (.079)	0.098 (.082)	0.036 (.093)
sdim	0.915 (.038)	0.652 (.047)	1.064 (.061)
constant	-0.060 (.169)	-0.945 (.125)	-0.087 (.199)
prev outcome	-	1.745 (.125)	-
herd variance	0.741 (.211)	0.317 (.093)	1.043 (.296)
cow variance	6.297 (.367)	2.141 (0.248)	8.092 (.511)
cow random slope (sdim)	-	-	2.147 (.233)
covariance	-	-	0.740 (.270)

The 2 extensions of the random-intercept model were both highly significant: the coefficient for $Y_{i,j\text{-}1}$ in the transitional model is much larger than its SE, and log-likelihood values for the random-intercept and trend models differed by some 144 units. In the transitional model, the odds ratio for having the same outcome as at the previous time point was high ($e^{1.745}$=5.7). The estimates are conspicuously different between the 3 models. One reason for this is very strong within-cow clustering, *eg* in the random-intercept model, the (latent variable) *ICC* equals (0.741+6.297)/(0.741+6.297+3.29)=0.68 (Section 22.2.3). The estimates from 2 random-effects models are not directly comparable when the variances are high because of the scaling caused by the random effects (Section 22.6). One possibility is to scale both estimates to PA scale; for example, scaling the estimates for -sdim- yields: $0.915/\sqrt{1+0.346*(0.741+6.297)}=0.494$, and $0.652/\sqrt{1+0.346*(0.317+2.141)}=0.479$,

so these estimates are in good agreement. The estimated intercept in the transitional model has a different interpretation and role in the model: it corresponds to tests whose predecessor was zero, whereas in the other models, it corresponds to any test. Due to non-constant variances created by the random slopes, it is more difficult to scale the estimates of the trend model. The conclusion from the example is that both extensions of the random-intercept model affect the model quite strongly, and it is not clear which (if any) of them is preferable. We continue the analysis of this model in Example 23.10.

Generalised estimating equations (GEE) were introduced in 2 papers by Liang and Zeger 1986); Zeger & Liang, (1986), as a set of estimating equations to obtain parameter estimates for discrete and continuous repeated measures data. The idea has proven not only durable but also

extendable to other data structures (*eg* hierarchically clustered and spatial data), statistical inference accompanying the estimates, as well as many variants of estimating equations (Hanley *et al*, 2003). A fairly recent (statistical) monograph (Hardin & Hilbe, 2003) is devoted entirely to GEE methods, which today are one of the most popular approaches in the health and biological sciences. We will confine ourselves here to describing the original (and probably still most popular) GEE method to obtain population-averaged estimates for clustered data. To illustrate the methods, we will use the repeated measures somatic cell count data (scc_40) and the clustered pig-pneumonia data (pig_adg) from Chapter 22.

23.5.1 Estimating equations

Let's initially explain the meaning of an 'estimating equation'. When using maximum likelihood (ML) estimation, the parameters are chosen to maximise the log-likelihood function. In practice, maximising a function involves computing the (partial) derivatives of the function with respect to its parameters and equating these to zero. These would be the estimating equations for ML estimation (and the derivatives of the log-likelihood function is called the score function). Except for very simple cases, the equations do not have an explicit solution and must be solved iteratively. The approach we are going to take here involves GLMs and a partially specified model, so that no likelihood function is available. Specifically, the GEE method requires only assumptions about the marginal mean and variance (and information about the subjects, or more generally clusters, of the data). Nevertheless, estimation is based on iterative solution of similar generalised estimating equations. These equations involve the mean of the outcome across clusters, therefore GEE yields estimates with a **PA interpretation**. Recall however from Section 22.4.1, that a distinction between SS and PA estimates is unnecessary for models with an identify link, such as a linear (mixed) regression models.

23.5.2 Statistical inference using GEE

The Liang and Zeger version of GEE is based on correlations in a working correlation matrix. Despite the fact that no assumptions about the form of the correlation of the data within the clusters are made, the estimating equations involve a **working correlation matrix** containing the estimated correlations among observations within a cluster, in each cycle of the iterations. This matrix can be given different forms (independent, compound symmetry, autoregressive, unstructured *etc* as in Section 23.3.1) to tailor the estimating algorithm toward one's perception of the data structure. Because the matrix is not part of the model, its form is not as crucial as in a fully parametric model. Theoretically, the GEE method gives asymptotically unbiased estimates even if the working correlation matrix is misspecified; that might, however, lead to loss in efficiency (Fitzmaurice, 1995). Estimation of variance (*ie* standard errors and correlations among estimates) can be either model-based or robust (or empirical) as described in Section 20.5.4. The latter method is also asymptotically unbiased, and is generally recommended because the GEE method loses its robustness to misspecification if model-based variance estimation is used. It is worth noting the general relationship that GEE with independence working correlation structure and robust variance is exactly the same as ordinary clustered robust variance estimation (Section 20.5.4).

As to the choice of working correlation structure, you should first and foremost be guided by your understanding of the data. For hierarchically clustered data (*eg* pigs in farms), anything but a compound symmetry (or exchangeable) correlation structure would seem unreasonable.

Particular caution should be exercised with negatively correlated binary data. In this case, an ordinary logistic model with robust standard errors has been recommended (Hanley *et al*, 2000). For repeated measures data, one would usually choose a structure that allows for autocorrelation. It might also be tempting to try an unstructured correlation to see what patterns the data show when not constrained by a particular structure. However, large correlation structures imply estimation of a large number of 'working parameters' and numerical problems might be encountered especially in unbalanced datasets. Recently a criterion (QIC), similar to Akaike's information criteria has been developed to guide the choice of correlation matrix (Pan, 2001) and implemented in standard software (Cui & Qian, 2007). We first illustrate the GEE method in Example 23.8 by applying it to the 2-level pig pneumonia data with a binary outcome.

A word of caution: the use of the GEE approach is appropriate when it comes to missing data (Section 15.5). It has long been recognised that GEE is not robust to missing data under the missing at random (MAR) assumption, but that addition of a weighting scheme to the procedure could resolve the problem (Robins *et al*, 1995); (Molenberghs *et al*, 2007, Chapter 27). The actual scheme depends on the structure of the missing values (*eg* whether these are drop-outs or intermediate missing values). An implementation of a weighting scheme for drop-outs has been published (Jansen *et al*, 2006), but such adjustments to GEE do not seem to be generally available in standard statistical software.

23.5.3 GEE for multilevel data structures

One apparent drawback of the GEE method is its limitation to a single level of clustering. Except for the alternating logistic regression (ALR) version of GEE discussed below, the problem of extending GEE algorithms to account for more complex data structures has received relatively little attention in the literature (Chao, 2006). The question of how to best set up a classical GEE analysis for binary repeated measures with an added hierarchical level (*eg* repeated measures on cows clustered in herds) was discussed on the basis of multiple simulation studies (Masaoud and Stryhn, 2010). The recommendations were that, with moderate- to-large numbers of highest level clusters, it is sufficient to cluster at the highest level to achieve approximately unbiased estimates and standard errors at all levels, and that other schemes such as ignoring the highest level clusters or modelling them by fixed effects were less successful. This finding agrees with the recommendation by Hardin and Hilbe (2003), Chapter 3 that for complete datasets with a number of clusters above 30, there is little gain in using more complicated correlation structures than independence (which de facto is ordinary logistic regression with robust standard errors), and it also agrees with the common approach to survey data to adjust for clustering by primary sampling units and pay less attention to

Example 23.8 Generalised estimating equations for pig-pneumonia data
data = pig_adg

For the model of Example 22.13, a GEE analysis with a compound symmetry structure for the working correlation matrix and robust standard errors gave a regression coefficient of 0.354 (0.216). For comparison with the previous random- effects estimate (0.437), we might compute its PA counterpart using Eq 22.2: $\beta^{PA} \approx 0.437/\sqrt{1+0.346*0.879} = 0.383$. Thus, the difference between the 2 estimates is not entirely due to their different interpretations; however, relative to the SEs, the difference is small. The working correlation matrix had a correlation of 0.18 (between pigs in the same farm), which is quite similar to ρ=0.21 computed in Example 22.2.

subsequent levels (Sections 2.10 and 20.5.5). As the herd-level working correlation structure cannot easily be set up to account for autocorrelation within cows, an exchangeable structure must be used and a possible loss of power by misspecification of the structure must be accepted. In Example 23.9, we compare different 3-level GEE approaches to analysis of a continuous outcome by reanalysing the somatic cell count data from Examples 23.5-6.

For binary outcomes, an alternative to the standard GEE algorithms was developed by Carey *et al* (1993) and termed **alternating logistic regression** (ALR) because the estimation algorithm in each step of the iterations employs 2 (very different) logistic regression models to update the parameters. As this approach was favoured for binary data in a comprehensive review of GEE methods (Hardin & Hilbe, 2003), and it also has the ability to deal with 2 levels of clustering, we briefly describe the idea and demonstrate in Example 23.10 its use (together with other GEE implementations) to the binary somatic cell count data. The standard GEE procedure describes within-subject clustering in terms of a working correlation matrix; however, correlation is not the most obvious measure of association for binary outcomes. The ALR method instead describes the clustering in terms of odds-ratios for 2 subjects within the same cluster, and offers estimates with SEs of such quantities. As the estimating equation for the fixed effects is the same as for standard GEE, the robustness properties of GEE are retained. One drawback of the approach is that it is only implemented in a few statistical packages (SAS and R/S-plus) and only with exchangeable correlation structures. That is, in the repeated measures context, the odds-ratio parameter gives the ratio between odds of disease when it is known that another observation on the same subject is disease-positive versus when it is disease-negative. A numerical illustration is given also in Example 23.10.

23.5.4 Summary remarks on GEE and discrete mixed models

We expand here a bit on the summary Table 20.4 to specifically address the choice between GEE and discrete mixed models. The advantage of the GEE method (and many of its generalisations) is that it has robust theoretical properties with few model assumptions. It is also computationally feasible for large datasets and can be fit with a wide range of working correlation structures. It is one of the few general methods for use with discrete repeated measures and spatial data; however, it does not provide much information about the random structure of the data, and it cannot be used to model random structure in terms of random slopes. Its lack of likelihood-based inference and standard errors for correlation parameters are perhaps less of an issue, but GEE estimation may require additional analysis for data with a large proportion of missing values that cannot be assumed missing completely at random.

A general GLM(M) class of random-effects models that allow inclusion of autocorrelation and other complex correlation structures does not exist (disregarding the quasi-likelihood approach discussed in Section 23.4.1), but a range of specific methods are available for binary and count data. The choice between methods may require a considerable effort to understand their theoretical basis, and can also be difficult in practice, *eg* in binary data with strong within-subject clustering, as demonstrated by our examples. It is recommended to try multiple approaches in order to assess the robustness of the results to the particular choice of method. Modelling of time by random slopes (trend models) should probably be included among the methods used, unless the time series is very short. The ability to include additional hierarchical structure remains one of the main advantages of mixed models.

Example 23.9 Generalised estimating equations for log somatic cell counts
data = scc_40

We analysed these data using linear mixed models for repeated measures in Examples 23.5-6. Because of the identity link function, the SS and PA parameters coincide. The difference of the GEE approach lies therefore, entirely in the estimation method. The table shows parameter estimates from GEE analyses clustered at the cow level with compound symmetry, autoregressive (ar(1)) and unstructured working correlation matrices. The table also gives values of the working correlations 1, 2 and 3 time steps apart; the values for the unstructured correlation were obtained by averaging the corresponding values in the matrix. Some software implementations of GEE (*eg* in SAS) will fit stationary (Toeplitz) structures without excluding incomplete sets of repeated measures; the results were close to those shown for the unstructured correlations.

	Cow-level working correlation matrix structure			Herd work corr.
	Comp. symm.	Autoregressive	Unstructured	Comp. symm.
Variable	β (SE)	β (SE)	β (SE)	β (SE)
shsize	0.826 (.123)	0.799 (.124)	0.755 (.121)	0.732 (.322)
heifer	-0.777 (.042)	-0.755 (.042)	-0.771 (.041)	-0.750 (.054)
season=spring	0.015 (.023)	0.054 (.024)	0.031 (.022)	0.086 (.036)
season=summer	0.026 (.026)	0.060 (.026)	0.033 (.024)	0.115 (.045)
season=fall	-0.022 (.025)	0.003 (.025)	-0.010 (.023)	0.033 (.039)
sdim	0.336 (.014)	0.315 (.014)	0.327 (.013)	0.312 (.017)
constant	5.290 (.033)	5.256 (.033)	5.285 (.032)	5.211 (.068)
ρ (1 month)	0.555	0.671	0.647	0.072*
ρ (2 months)	0.555	0.451	0.592	0.072*
ρ (3 months)	0.555	0.303	0.538	0.072*
QIC	21102.24	21084.62#	21101.43#	21199.31

*correlation among all values within a herd; #QIC for model ignoring gaps

These values should be compared with those of Example 23.5. The parameter estimates for -heifer- and -sdim- are in reasonable agreement between all models, including the uncorrected analysis. This may be said also for -shsize- when considering its large SE (from the linear mixed model). For -season-, the estimates from GEE clustered at the herd level are markedly off all the other estimates (including the mixed model in Example 23.5); the values are actually closer those of a simple linear regression (not shown). The best agreement with the mixed model is achieved by the unstructured correlations, and also the estimates obtained by the autoregressive structure differ slightly. These data demonstrate that choice of working correlation structure is not always of minor importance for the fixed effects, even in a large dataset. The standard errors are all fairly close except for -shsize- where only the analysis clustered at the herd level produces a sensible value (close to the mixed model); the cow-level GEE analyses do not account for herd and thus, cannot be expected to produce a valid SE for a herd-level predictor. The cow-level correlations show good agreement with the mixed model estimates, and still indicate both the compound symmetry and autoregressive structures to be inadequate. The QIC points, perhaps surprisingly, to the autoregressive structure as the preferable one, and strongly rejects the analysis clustered at the herd level. Thus, model choice by the QIC statistic leads to different conclusions than in the mixed model and the recommendations from simulation studies.

In summary, there is a fair agreement between the GEE and linear mixed models analysis, but some questions remain with respect to choice of working correlation structure.

Example 23.10 GEE and ALR estimation for high somatic cell counts
data = scc_40

Using the same dataset as for Example 23.7, we compare different implementations of GEE to account for the repeated measures and additional clustering in herds. For comparison, an ordinary logistic regression is included as well. The GEE analysis with an autoregressive working correlation structure did not account for clustering for herds.

Model	Ord. logist. regr.	GEE: ar(1) at cows	GEE: cs at herds	Altern. log. regr.
Variable	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
shsize	1.150 (.102)	1.034 (.194)	1.045 (.631)	0.748 (.484)
heifer	-1.100 (.038)	-1.125 (.072)	-1.111 (.081)	-1.165 (.074)
season=spring	0.162 (.051)	0.110 (.047)	0.126 (.062)	0.030 (.060)
season=summer	0.191 (.052)	0.119 (.051)	0.168 (.075)	0.047 (.056)
season=fall	0.081 (.054)	0.051 (.050)	0.093 (.066)	0.021 (.053)
sdim	0.440 (.023)	0.437 (.026)	0.439 (.028)	0.475 (.030)
constant	-0.224 (.041)	-0.143 (.056)	-0.226 (.106)	-0.062 (.092)

We see immediately that the estimates show less variation across different analyses than in Example 23.7. Two explanations can be offered: contrary to Example 23.7, the analyses correspond to the same models (fixed effects), and all the estimates are on the same (PA) scale. Some major differences remain, in particular between ALR and the GEE variants. These may be linked to the presence of a substantial portion of missing data (approximately 40% when compared to the full dataset with 11 observations per cow). The 2 ALR log-odds ratio parameters were estimated at:

$$\text{within-subject: } \phi = 2.229\ (.086), \quad \text{and} \quad \text{between-subject (within herd): } \phi = 0.218\ (.058)$$

The interpretation of these values is that the odds of a high somatic cell count at one test is $e^{2.229}=9.29$ times higher when it is known that another test of the same animal was positive than when that other test was negative; similarly, the odds is only $e^{0.218}=1.24$ times higher when the same information is given about another animal in the same herd. As we noted before, the within-subject clustering is very strong, and the between-subject and within-herd clustering is fairly weak in comparison. The random-intercept model estimates from Example 23.7 agree well with the ALR estimates after rescaling to PA scale, *eg* for the herd size (-shsize-):

$$1.367/\sqrt{1+0.346*(0.741+6.297)}=0.738$$

With discrepancies of this magnitude between the GEE-type procedures the choice of an appropriate procedure for these data is not quite yet resolved, but the procedures do seem to be more robust to the choice procedure than the random-effects models of Example 23.7.

REFERENCES

Barbosa M, Goldstein H. Discrete response multilevel models Quality and Quantity. 2000; 34: 323-30.

Carey V, Zeger S, Diggle P. Modelling multivariate binary data with alternating logistic regressions Biometrika. 1993; 80: 517-26.

Chao EC. Structured correlation in models for clustered data. Stat Med. 2006;25(14):2450-68.

Cui J, Qian G. Selection of working correlation structure and best model in GEE analyses of longitudinal data Communications in Statistics-Simulation and Computation. 2007; 36:987-96.

Davis C. Statistical Methods for the Analysis of Repeated Measurements. Springer; New York. 2002.

Davis R, Dunsmuir W, Wang Y. On autocorrelation in a Poisson regression model Biometrika. 2000; 87: 491-505.

Diggle P, Heagerty P, Liang K, Zeger S. Analysis of Longitudinal Data, 2nd Ed. Oxford University Press; Oxford. 2002.

Dohoo I, Meek A. Somatic cell counts in bovine milk Can Vet J. 1982; 23: 119-125..

Everitt B. The analysis of repeated measures: A practical review with examples J R Stat Soc D (The Statistician). 1995; 44: 113-35.

Fitzmaurice G. A caveat concerning independence estimating equations with multivariate binary data Biometrics. 1995; 51: 309-17.

Fitzmaurice G, Laird N, Ware J. Applied Longitudinal Analysis. Wiley; New York. 2004.

Gotway C, Wolfinger R. Spatial prediction of counts and rates Stat Med. 2003; 22: 1415-32.

Hanley J, Negassa A, Edwardes M. GEE: Analysis of negatively correlated binary responses: a caution Stat Med. 2000; 19: 715-22.

Hanley J, Negassa A, Edwardes M, Forrester J. Statistical analysis of correlated data using generalized estimating equations; an orientation Am J Epidemiol. 2003; 157: 364-75.

Hardin J, Hilbe J. Generalized estimating equations. Chapman & Hall/CRC; Boca Raton. FL. 2003.

Heagerty P, Kurland F. Misspecified maximum likelihood estimates and generalised linear mixed models Biometrika. 2001; 88: 973-85.

Hedeker D, Gibbons R. Longitudinal Data Analysis. Wiley; New York. 2006.

Jansen I, Beunckens C, Molenberghs G, Verbeke G, Mallinckrodft C. Analyzing incomplete discrete longitudinal clinical trial data Stat Sci. 2006; 21: 52-69.

Li J, Yang X, Wu Y, Shoptaw S. A random-effects Markov transition model for Poisson-distributed repeated measures with non-ignorable missing values. Stat Med. 2007;26(12):2519-32.

Liang K, Zeger S. Longitudinal data analysis using generalized linear models Biometrika. 1986;

73: 13-22.

Littell R, Milliken G, Stroup W, Wolfinger R, Schabenberger O. SAS for Mixed Models, 2nd ED. SAS Publishing; Cary, NC. 2006.

Masaoud E, Stryhn H. A simulation study to assess statistical methods for binary repeated measures data. Prev Vet Med. 2010;93(2-3):81-97.

Min Y, Agresti A. Random effect models for repeated measures of zero-inflated count data Statistical Modelling. 2005; 5: 1-19.

Molenberghs G, Verbeke G. Models for Discrete Longitudinal Data. Springer; New York. 2005.

Molenberghs G, Verbeke G, Demetrio C. An extended random-effects approach to modeling repeated, overdispersed count data Lifetime Data Analysis. 2007; 13: 513-31.

Nelson K, Leroux B. Statistical models for autocorrelated count data Stat Med. 2006; 25: 1413-30.

Pan W. Akaike's information criterion in generalized estimating equations Biometrics. 2001; 57: 120-5.

Pinheiro J, Bates D. Mixed-effects Models in S and S-Plus. Springer; New York. 2000.

Rasbash J, Steele F, Browne W, Goldstein H. A User's Guide to MLwiN. University of Bristol; Centre for Multilevel Modelling, Bristol. 2008.

Robins J, Rotnizsky A, Zhao L. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data J Amer Statist Assoc. 1995; 90: 106-21.

Schukken Y, et al. Analysis of correlated discrete observations: background, examples and solutions Pre Vet Med. 2003; 59: 223-40.

Senn S, Stevens L, Chaturvedi N. Tutorial in biostatistics: Repeated measures in clinical trials: simple strategies for analysis using summary measures Stat Med. 2000; 19: 861-77.

Stryhn H, Andersen J, Agger J. Milk production in cows studied by linear mixed models Symposium in Applied Statistics. 2001.

Thurmond MC, Branscum AJ, Johnson WO, Bedrick EJ, Hanson TE. Predicting the probability of abortion in dairy cows: a hierarchical Bayesian logistic-survival model using sequential pregnancy data. Prev Vet Med. 2005; 68: 223-39.

Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. Springer; New York. 2001.

Vigre H, Dohoo IR, Stryhn H, Jensen VF. Use of register data to assess the association between use of antimicrobials and outbreak of Postweaning Multisystemic Wasting Syndrome (PMWS) in Danish pig herds. Prev Vet Med. 2010;93(2-3):98-109.

Yang M, Heath A, Goldstein H. Multilevel models for binary outcomes: attitudes and vote over the electoral cycle J R Stat Soc A. 2000; 163: 49-62.

Zeger S, Liang K. Longitudinal data analysis for discrete and continuous outcomes Biometrics. 1986; 42: 121-30.