22

MIXED MODELS FOR DISCRETE DATA

OBJECTIVES

After reading this chapter, you should be able to:

1. Understand the differences between linear mixed models (continuous data) and generalised linear mixed models (GLMMs) (discrete and continuous data) and the role of the link function in the latter.

2. Fit random effects logistic and Poisson models.

3. Understand the differences between population-averaged and subject-specific modelling.

4. Use a latent variable approach to compute the intracluster correlation (*ICC*) for binary outcomes.

5. Use either quasi-likelihood or maximum likelihood methods for fitting GLMMs.

6. Assess the statistical significance of both fixed and random effects in GLMMs.

7. Evaluate residuals to assess the adequacy of a GLMM that you have fit.

22.1 INTRODUCTION

In both theory and practice, it has proven more difficult than one might have anticipated to generalise the mixed models approach from continuous to discrete data. One effect of these difficulties is the existence of a wide variety of generalisations of mixed models to discrete data, some of them only for a particular type of discrete data (usually binary) and some of them within wider frameworks. In this chapter, we review the model class most analogous to linear mixed models: the **generalised linear mixed models** (GLMM). In order to fully appreciate this analogy, the reader is encouraged to review linear mixed models first (Chapter 21).

Our main focus here will be on binary data (logistic regression with random effects, Section 22.2) and on count data (Poisson regression with random effects, Section 22.3), but the random effects extension applies to a flexible class of discrete models which *eg* includes multinomial regression. As in Chapter 21, our mixed models will reflect a hierarchical structure but it is also possible to build models for other data structures. However, (and this goes generally for mixed models for discrete data) the statistical analysis is more difficult than for continuous data, and requires more care and choices by the researcher (of which the choice of software is an important one). The field is still growing and advancing but we attempt to give the applied researcher a snapshot of its present state.

We will use 2 binary data examples to illustrate the methods: one on pneumonia in pigs and another on first service conception risks in cows. The first dataset, pig_adg, stems from a 2-level hierarchy (pigs within farms) and we will consider only a single pig-level, dichotomous predictor. The second dataset, reu_cfs from the Reunion Island reproduction study (introduced in Example 20.1), contains a 3-level hierarchy (lactations within cows within herds) and 2 lactation-level predictors. Poisson regression models will be illustrated by the TB data from Chapter 18 (tb_real).

22.2 LOGISTIC REGRESSION WITH RANDOM EFFECTS

We consider again the example of animal disease observed in several herds (*eg* the pig-pneumonia data). The logistic regression analogue of Eq 21.2 for the probability p_i of the i^{th} animal being diseased is:

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + u_{\text{herd}(i)} \qquad \textit{Eq 22.1}$$

where $u_{\text{herd}(i)}$ is the random effect of the herd containing animal i, assumed to be $u_{\text{herd}(i)} \sim N(0, \sigma^2_h)$, the X_is are the predictor values for the i^{th} animal, and the relationship between the probability p_i and the binary outcome Y_i is unchanged: $p(Y_i=1)=p_i$. The only change from the ordinary logistic regression model is the herd random-effects term. Example 22.1 shows that adding random effects can have an appreciable impact on the model.

22.2.1 Analogies and differences to a linear mixed model

We have seen that a mixed logistic regression model adds the random effects to the fixed effects, both on a logistic scale. So, bearing the logistic scale in mind, we build the models in a similar way to linear mixed models and they might include multiple random effects and possibly random slopes as well. The statistical analysis also has strong similarities in the way confidence intervals and tests are computed.

Example 22.1 Random effects logistic model for pig-pneumonia data
data = pig_adg

Data on both atrophic rhinitis and enzootic pneumonia were recorded on 341 pigs at slaughter. Causally, it was assumed that atrophic rhinitis might increase the risk of pneumonia, through destruction of the pig's air-filtering mechanism (nasal turbinates). The atrophic rhinitis score was converted to a dichotomous variable (-ar_g1-) indicating the presence/absence of an atrophic rhinitis score greater than 1. Similarly, the lung score was converted to a dichotomous variable (-pn-) representing the presence/absence of pneumonia.

The unconditional association between -pn- and -ar_g1- was:

		ar_g1		
		1	**0**	**Total**
pn	1	109	77	186
	0	66	89	155
Total		175	166	341

Odds ratio=1.909
95% CI=(1.212,3.010)
Chi-square=8.69
P-value=0.003

These statistics indicate a moderate but clearly significant association between -pn- and -ar_g1-. However, we have ignored the fact that the pigs came from 15 farms, and the prevalence of -pn- actually varied from 17% to 95% across farms. Consequently, it appears that we should be concerned about farm effects. The logistic regression with random effects (Eq 22.1) gave the estimates:

	Coef	**SE**	**Z**	**P**	**95% CI**	
ar_g1	0.437	0.258	1.69	0.091	-0.069	0.943
constant	0.020	0.301	0.07	0.948	-0.57	0.61

In addition, the estimated variance of farm random effects (with SE) was:

$$\sigma_h^2 = 0.877\,(0.433)$$

We shall later see how to compute the significance of the random effect (it is highly significant). The regression coefficient for -ar_g1- should be compared with the log of the simple odds ratio (ln(1.909)=0.647). Accounting for the herd effects reduced the association considerably, and it was no longer significant. In other words, the farms had both a clustering and confounding effect.

The 2-level model (Eq 22.1) induces correlations between the observations in a similar way as its linear mixed model analogue: equal correlations between animals within the same herd and independence between herds. However, we have to be careful here: the correlations within a herd are the same only for animals with the same fixed effects. In our example, all -ar_g1- positive animals within a herd are equally correlated, and the same for all -ar_g1- negative animals. This difference between animals with different predictor values may seem strange and is usually small in practice (unless the predictor has a very strong effect). It is one of the many consequences of modelling the fixed and random effects on the logit scale. Nevertheless, the model is perfectly valid as a method to account for correlation (or clustering) between animals in the same herd.

Strictly speaking, the model in Eq 22.1 has a 2-step interpretation which is perhaps best understood by imagining how data would be generated by the model. For an animal i in the j^{th} herd, we would first select the herd random effect (u_j) according to its N(0, σ^2_h) distribution and compute p_i from the fixed effects and the selected u_j-value. We would then select the outcome

Y_i as positive with probability p_i or negative with probability 1-p_i. A common shorthand for this 2-step interpretation is that Eq 22.1 is 'conditional on' the random effects. In linear mixed models we modelled the outcome directly, so there was no need for a conditional interpretation.

In the next 2 sections, we describe how the interpretation of fixed and random effects parameters change from the logistic regression model and the linear mixed model.

22.2.2 Interpretation of fixed effects parameters

The interpretation of fixed effects in a linear mixed model was essentially unaffected by the added random effects. Again, the modelling on the logit scale complicates the interpretation of models such as Eq 22.1. The conditional interpretation of the model means that when exponentiating a regression coefficient (for -ar_g1- in the example) to obtain the odds-ratio (*ie* *OR*=exp(0.437)=1.55), the odds ratio refers to comparing pigs with and without atrophic rhinitis **in a particular herd** (corresponding to a selected herd random effect, no matter the actual u_j-value). Frequently this is called a **subject-specific** or **cluster-specific** (in our example, a herd-specific) estimate, as opposed to a **population-averaged** estimate, which would refer to the *OR* for comparing pigs with and without atrophic rhinitis **from any herds** in the population of herds (*ie* the 2 pigs can be from different herds). Therefore, if we think of the *OR* as the answer to questions such as 'how much is the risk increased?' (in our example, the risk of pneumonia for an 'ar'-pig versus a healthy pig), the cluster-specific estimate answers the farmer's question and the population-averaged estimate answers the slaughterhouse's question (where pigs are submitted from many different herds). That these 2 questions have different answers challenges our intuition, but is an incontestable fact.

Two alternatives exist to the cluster-specific *OR*. One is to convert from cluster-specific to population-averaged parameters (on the logit scale) using the following formula.

$$\beta^{PA} \approx \beta^{SS} / \sqrt{1 + 0.346\,\sigma_h^2} \qquad \textit{Eq 22.2}$$

Example 22.2 illustrates the procedure (see also Section 22.4 for further discussion of cluster-specific and population-averaged estimates). The second alternative to the cluster-specific *OR* is a re-interpretation of this value as a median odds ratio (*MOR*) across the population of clusters (farms in our case). The rationale behind this idea, introduced by Larsen K *et al* (2000), is that when comparing 2 pigs, with and without atrophic rhinitis, from different farms, the *OR* is really a random quantity because its value depends on the farm effects for the 2 selected farms. Just as any other random variable, it has a distribution, and hence it makes sense to look for a central value in this distribution. The mean in this distribution is the population-averaged *OR*, and the median in this distribution (*MOR)* is equal to the cluster-specific *OR* (computed at 1.55 above). We can now say that when comparing pigs with and without atrophic rhinitis from the population, the odds ratio will, with probabilities 0.5, take values above and below 1.55. An associated range within which the *OR* will lie with a given probability, *eg* 80%, can also be computed; the details are shown in Example 22.2.

22.2.3 Interpretation of variance parameter(s)

In Eq 22.1, the herd random effect variance σ^2_h has no direct interpretation in terms of the probabilities of disease. The equation shows that it refers to the variation between herds of the disease probabilities on a logit scale. We can still interpret it qualitatively: a value of zero

means no variation between herds (and therefore no clustering) and a large positive value means a high degree of clustering. However, the (correct) statement that the logits of probabilities vary within $\pm 1.96\sigma_h$ across herds with a probability of 95%, is not very intuitive. The variance σ^2_h can, without too much extra work, be interpreted in terms of either variance components or median odds ratios; we discuss these in turn.

In linear mixed models, the variance parameters could be interpreted as **variance components**, but in models of discrete data, we have a problem with this interpretation. If we compare Eq 22.1 with the linear mixed model (Eq 21.2), the error term (ε_i) is missing in the logistic equation. This is because the distribution assumption is on the original scale—in our example $Y_i \sim \text{bin}(1, p_i)$, so that the errors in the model stem from the binomial (binary) distribution instead of a normal distribution. Recall that in this binary distribution the variance equals $p_i(1\text{-}p_i)$. Now the total variance in the data, var(Y_i), is no longer just the sum of the error variance and the random effects variance, as they refer to different scales. Even worse, the total variance is not constant because the binomial variance varies with p, so a single decomposition of the variance does not exist. A few years ago, several papers reviewed the computation of variance components and intraclass correlation coefficients (*ICC*s, see Section 20.3.3; sometimes also denoted variance partition coefficients (*VPC*s) in acknowledgement of the non-constant variances and correlations, as explained above) in random effects logistic regression, and a number of different methods were suggested (Browne W *et al*, 2005; Goldstein *et al*, 2002; Rodriguez & Elo, 2003; Vigre *et al*, 2004). We confine ourselves to explaining a simple approximation method based on **latent response variables** to represent the logistic model as a threshold model (latent variables were introduced in Chapter 17; see also Snijders & Bosker (1999), Section 14.3, and Rabe-Hesketh & Skrondal (2008), Section 6.2.2.

The simplest approach to getting both the individual and herd variances on the same (logistic) scale is to associate with every animal i a latent continuous measure, Z_i, of the 'degree' of sickness. The observed binary outcome Y_i is then obtained simply as whether the degree of sickness exceeds a certain threshold. In formulae, if we denote the threshold by t, then Y_i=1 if Z_i>t, and Y_i=0 when $Z_i \leq t$. Sometimes this may seem a plausible theoretical construct, and sometimes less so. For the pig_adg data, the pig pneumonia scores indeed quantified the amount of damage observed in the lungs; for the reu_cfs data on the other hand, success in fertilization seems a truly binary event. Mathematically speaking, any model for Z_i is then translated into a model for the binary outcomes. In particular, Eq 22.1 is obtained exactly when t=0 and

$$Z_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + u_{\text{herd}(i)} + \varepsilon_i \qquad \textit{Eq 22.3}$$

where the fixed effects and the herd effects are exactly as before, and where the error terms ε_i are assumed to follow a logistic distribution with mean zero and variance $\pi^2/3$=3.29. (The logistic distribution is similar in shape to the normal distribution, and for most practical purposes, it is equivalent to assume either of these distributions.) Eq 22.3 is a linear mixed model for Z_i! Therefore, computation of variance components and *ICC*s for Z_i-variables follows the rules of Chapter 21:

$$\text{var}(Z_i) = \text{var}(u_{\text{herd}(i)}) + \text{var}(\varepsilon_i) = \sigma^2_h + \pi^2/3$$
$$\rho = \sigma^2_h / (\sigma^2_h + \pi^2/3) \qquad \textit{Eq 22.4}$$

We demonstrate the procedure in Example 22.2. To summarise, the latent variable approach allows interpretation in terms of variance components and *ICC*s by fixing the error variance at

$\pi^2/3$. We should keep in mind that the strict interpretation is for the latent variables, and the values are only approximate for the binary outcomes. In particular, as noted, the variances and correlations are not constant for the binary outcomes but depend on the predictors; this dependence has disappeared for the latent variables. Experience with different methods for computing *ICC*s indicates that the latent variable *ICC* tends to be somewhat larger than the true *ICC* for the binary outcome (see the above-cited papers). Intraherd correlations ranging from 0.04 to 0.42 (with most values <0.2) have been observed for a number of infectious diseases of animals (Otte & Gumm, 1997).

The variance σ^2_h can also be interpreted in terms of an odds-ratio between the risk in 2 randomly selected clusters, where the animals and herds compared should have the same fixed effects. The *OR* between the larger and smaller of the 2 risks is ≥1, and the median in its distribution (cluster-median odds-ratio, MOR_c) can be calculated as

$$MOR_c = \exp(0.954 * \sigma_h) \qquad \textit{Eq 22.5}$$

The interpretation of the *MOR* is that when comparing (identical) subjects from 2 randomly selected clusters (herds), the odds ratio will, with probabilities 0.5, take values above and below the *MOR*. One advantage of the *MOR* is that the heterogeneity between clusters is now comparable to the impact of the fixed effects (Larsen *et al*, 2000). Example 22.2 illustrates the procedure for the pig pneumonia data.

22.3 POISSON REGRESSION WITH RANDOM EFFECTS

A Poisson random intercept regression model with exposure n and herd random effect u can be written:

$$\begin{aligned} \ln(\lambda_i) &= \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_{\text{herd}(i)} \\ Y_i &\sim \text{Poisson}(n_i \lambda_i) \end{aligned} \qquad \textit{Eq 22.6}$$

with the assumption: $u_{\text{herd}(i)} \sim N(0, \sigma^2_h)$. Thus, the random effect is added to the fixed effects in a similar way as for logistic regression. The differences described in the previous section between models with and without random effects to a large extent carry over to Poisson regression. We briefly discuss the interpretation of fixed and random effects parameters from Eq 22.6, and illustrate by an example (see Example 22.3).

22.3.1 Interpretation of fixed effects parameters

The distinction between cluster-specific and population-averaged parameters largely vanishes because it is only the intercept among the fixed effect parameters that takes different values in its cluster-specific and population-averaged versions (Diggle *et al*, 2002, Section 7.4). This, perhaps somewhat surprising fact, is related to the log link function and holds true regardless of the form of the random effects, *ie* also if they include multiple hierarchical levels or random slopes.

22.3.2 Interpretation of variance parameters

In a linear mixed model (Chapter 21), the *ICC* (for observations within the same cluster) could be computed by decomposing the total unexplained variance into terms for each level of the

Example 22.2 Interpretation of fixed and random parameters for pig-pneumonia data
data = pig_adg

Based on the model presented in Example 22.1, we calculate an odds-ratio for -ar_g1- as exp(0.437)=1.55. It can be interpreted either as a cluster-specific value (valid when comparing pigs with and without atrophic rhinitis from the same farm) or as the median odds ratio (*MOR*) across the population of farms. For the population-averaged odds ratio, we first convert the parameter to marginal scale (using Eq 22.2):

$$\beta^{PA} = 0.437/\sqrt{1+0.346*0.877} = 0.383,$$

and then compute the odds ratio the usual way as exp(0.383)=1.47. This is the mean odds ratio across the population farms. The difference between the 2 *OR*s is modest here, due to the moderate (not very large) between-farm variation. An 80% range for the *OR* of pigs selected from randomly selected farms is computed as:

$$\text{80\% range:} \quad \exp(0.437 \pm 1.282\sqrt{2*\sigma_h^2}) = \exp(0.437 \pm 1.698) = (0.28, 8.46),$$

where 1.282 is the Z_α for α=0.2 from Chapter 2 (the 90% percentile of *Z*). The range is wide and spans well across 1, essentially stating that the impact of the between-farm variation is stronger than of the predictor (atrophic rhinitis).

Turning next to the variance parameter, we calculate by the latent variable approach a total variation of 0.877+3.290=4.167, and an *ICC* (and proportion of variance at the herd level) of

$$\rho = 0.877/4.167 = 0.21.$$

Finally, the cluster-median odds-ratio for the random effect is calculated as:

$$MOR_c = \exp(0.954\sqrt{0.877}) = 2.44.$$

The median odds ratio for 2 comparable pigs (same fixed effects) from 2 randomly chosen herd is 2.44, which is quite large compared to the odds ratio comparing pigs from the same farm with and without atrophic rhinitis (1.55). This shows that even a moderate between-farm variation has an appreciable impact on the risk of individual pigs.

hierarchy (see Examples 21.1 and 21.3). For a random effects logistic regression, a similar (approximate) calculation was enabled by the latent variable approach. For Poisson regression, exact formulae exist (Stryhn *et al*, 2006), but the resulting variance decomposition and *ICC* depend on the predictor values, so there is no longer a simple and unique *ICC* across the entire dataset (except for the 'null' model with no predictors). To simplify the notation, denote by βX a set of predictor values, including the logarithmic offset, of interest:

$$\beta X = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \ln(n).$$

Then the variances at level 1 (lowest) and level 2 (highest) as well as the *ICC* (and proportion of variance at the highest level) are given by:

$$\begin{aligned} &\text{level 1:} && \sigma(1) = \exp(\beta X + \sigma_h^2/2) \\ &\text{level 2:} && \sigma(2) = \exp(2\beta X + 2\sigma_h^2) - \exp(2\beta X + \sigma_h^2) \\ &ICC: && ICC = \sigma(2)/(\sigma(2)+\sigma(1)) \end{aligned} \qquad \textbf{\textit{Eq 22.7}}$$

It is recommended to calculate the *ICC* across a range of βX values of interest. Example 22.3 illustrates the procedure.

Example 22.3 Random effects Poisson model
data = tb_real

In Examples 18.1 and 18.3, Poisson and negative-binomial models were fit to data on the incidence of new TB cases in cattle and cervid herds in Canada. The simple Poisson model (Example 18.1) was clearly inappropriate due to overdispersion. Below are the results from a random effects Poisson model with random herd effects which were assumed to have a normal distribution.

log likelihood = -143.56

	Coef	SE	Z	P	95% CI	
type=beef	-0.394	0.333	-1.18	0.236	-1.046	0.258
type=cervid	-0.238	0.487	-0.49	0.625	-1.192	0.716
type=other	-0.104	0.800	-0.13	0.896	-1.673	1.464
sex=male	-0.339	0.208	-1.63	0.103	-0.747	0.069
age=12-24 mo	2.717	0.747	3.64	<0.001	1.252	4.181
age=24+ mo	2.467	0.726	3.40	0.001	1.044	3.889
constant	-11.056	0.830	-	-	-12.682	-9.428

In addition, the estimated variance of herd random effects was

$$\sigma_h^2 = 1.688\,(0.593).$$

Compared with the negative binomial model, the type of animal remains completely insignificant while the coefficients for sex and age groups have generally moved slightly away from the null and their P-values have gone down. The random effects Poisson model fits the data substantially better than the negative binomial model because the log-likelihood is -143.6 compared with -157.7; the models are not nested (so a likelihood ratio test does not apply) but they have the same number of parameters (so log-likelihood values can be compared directly). The Poisson model was also fit with gamma-distributed random effects, but showed a slightly poorer fit (log-likelihood=-146.53). The herd-level variance (on log-scale) was estimated at 1.613 (0.477), and also the fixed effects parameters were similar to those shown above.

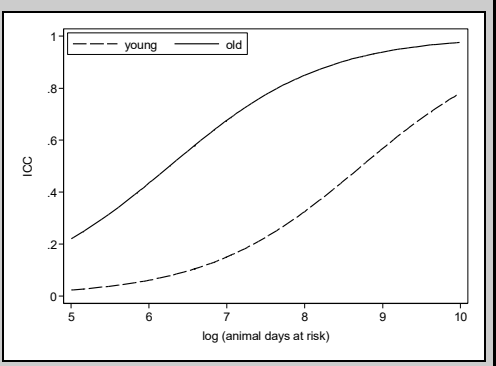

Fig. 22.1 Intra-class correlations for animal groups within herds, across a range of populations of risk and two age groups

Fig. 22.1 shows the estimated intra-class correlation (*ICC*) for 2 animal groups within the same herd, of the same age group (young: <12 mo, or old: >24 mo) across a range of sizes of the population at risk (on log scale) from approximately 50 to 22,000 animal-days at risk. The *ICC* depends very strongly on the age group and the population at risk because they both have a strong impact on the mean number of reactors, and the *ICC* is increasing as a function of the mean. In datasets with strong variations in the mean (due to predictors or the population at risk), the *ICC* therefore, does not seem particularly useful to illustrate the herd-level clustering.

22.4 GENERALISED LINEAR MIXED MODEL

The examples of mixed models in the first 2 sections extend to a larger class of models called generalised linear mixed models. These models are constructed by adding the desired random effects on the transformed scale specified by the link function in the same way as we did in the logistic and Poisson regressions. The random effects are, as a rule, assumed normally distributed with mean zero but possibly involving some non-zero correlations (between random effects at the same level). The model assumptions listed in Section 16.5 are still valid, although the distributional form and the equation for the linear predictor are now conditional on the values of the random effects (Section 22.2.1). Also, the general discussion of correlation structure and interpretation of fixed effects and variance parameters from Section 22.2 carries over to GLMMs, but some of the specific procedures for binary/binomial data do not, *eg* the latent variable approach for computing variance components and *ICC*s. To keep things simple, we will confine ourselves throughout to random intercept models (*ie* no random slopes are included), but adding random slopes and contextual effects to GLMMs is possible (and relevant) in the same way as in linear mixed models (Sections 21.3 and 21.4).

In this section we set out by discussing in general terms the population-averaged and cluster-specific interpretations of parameters in GLMMs (Section 22.4.1), and then we move on to specific models for binary, count or multinomial data (Sections 22.4.2-22.4.5).

22.4.1 Population-averaged versus cluster-specific parameters

The distinction between **population-averaged** (PA) and **cluster-specific** (subject-specific; SS) modelling for clustered data was introduced in Sections 22.2.1 and 22.3.1 where GLMMs were referred to as cluster-specific. Here we give more details and examples (largely following Diggle P *et al,* 2002). First a note on the term 'subject-specific'. It originates from repeated measures data consisting of several observations (*eg* over time) on different subjects (Chapter 23); in this case, measurements are 'clustered' within subjects. In the context of our usual hierarchical clustering, we might instead have our subjects clustered in groups (*eg* animals in herds). To avoid any confusion of this double use of 'subjects', we shall refer to the upper level of the structure as clusters or groups (instead of subjects). Our next observation is that the PA and SS interpretations of regression coefficients are equivalent for linear mixed models. This is not due to their usual normal distribution assumption but to the fact that the linear predictor is modelled on the original scale; in the terminology of GLMMs, the link function is the identity function and there is no shift of scale. Therefore, the proper reference for our discussion is a GLMM with non-identity link, and we also assume a 2-level structure.

The difference between the PA and SS approaches is in the way the clustering or grouping of the data is dealt with. As previously seen, subject-specific (or cluster-specific) models include a random effect for each cluster in the linear predictor of the model. The assumptions for the random effects (*ie* their distribution and correlation) imply a particular form of the distribution of the set of observations within a cluster, including their correlation structure. Population-averaged or marginal models involve only the **marginal** means, *ie* the expected values for a particular set of predictors **averaged** across the **population of clusters**, and do not include specific effects for each cluster. To show the difference between the parameters involved in the 2 types of model in simple formulae, denote by Y our observations and by u the random effects for the clusters (in an SS model). Then, using the vector-matrix notation introduced in Section 21.2.2, GLM(M)s of SS and PA types are based on the equations:

$$\begin{aligned} &\text{cluster-specific:} && \text{link}[E(Y|u)] = X\beta^{SS} + Zu \\ &\text{population-averaged:} && \text{link}[E(Y)] = X\beta^{PA} \end{aligned} \qquad \textit{Eq 22.8}$$

where, as before (Eq 21.8), X and Z are our shorthand for the fixed and random part predictors of the model, and E($Y|u$) is the mean of Y conditional on the value of u (as discussed in Section 22.2.1). As indicated by the notation in Eq 22.8, the SS and PA regression parameters are not identical (unless the link is the identity link or there is no clustering). Generally the PA parameters are closer to the null ('attenuated') than their SS counterparts; we already noted this attenuation to be absent for identity and log links (except the intercept). The difference depends on the amount of clustering and is often small relative to estimation error. Formulae for specific models are given in Sections 22.4.1 and 22.4.2.

The selection of the most appropriate model type (SS or PA) depends on the predictor(s) being examined. Consider, as an example, a clinical trial of the effect of a treatment (compared with a placebo) on the risk of a cow developing milk fever. The study is carried out in multiple herds and the breed of the cow (Holstein versus Jersey) is also recorded. The final model includes terms for the 2 dichotomous variables: treatment and breed. The β^{SS} for treatment in a cluster-specific model estimates the effect of the treatment in a **particular herd** on the risk of milk fever (compared with the risk in the same cow if she was not treated). This makes biological sense for cows staying in the same herd. On the other hand, the β^{PA} gives the effect (assumed decrease in risk of milk fever) of introducing a programme for treatment against milk fever **across all herds**. Thus, interest has been shifted from the individual herd to effects across herds. The parameters for breed are interpreted similarly, but the SS interpretation for breed would seem of less interest for herds with cows of a single breed (it refers to the altered risk of milk fever if all cows were replaced by cows of another breed), and the PA estimate seems more appropriate by comparing breeds across herds. Note also that an SS interpretation would become almost meaningless for herd-level predictors that are inherent in the herd (*eg* related to its location). This problem with an SS interpretation of a predictor that is unchangeable for the cluster is more common in repeated measures data where clusters are subjects (individuals), for example, with predictors such as sex or race/breed.

A final note on recommendations for the (not uncommon) situation that a dataset/model contains multiple predictors, some of which have a desired SS interpretation and others with a desired PA interpretation. If a conversion formula such as Eq 22.2 exists for the model used, one should convert (or "marginalise", in the terminology of Hedeker & Gibbons, 2006) the parameters with the desired PA interpretation. It is generally more difficult to convert from PA to SS estimates because most PA estimation procedures do not have information about the variances (see Section 23.5).

22.4.2 GLMMs for binary data

Random effects logistic regression was introduced in Section 22.2. Here we add some comments about alternative link functions and give additional formulae for the conversion between SS and PA estimates, and for the latent variable *ICC*. Example 22.4 introduces the data and model we use to illustrate statistical methods for binary GLMMs.

For a random intercept logistic regression model, the approximation was presented in Eq 22.2 and is repeated here for convenience:

$$\beta^{PA} \approx \beta^{SS} / \sqrt{1 + 0.346\,\sigma_h^2}$$

where σ^2_h is the (herd) cluster variance. By this formula, for PA parameters to be more than 10% lower than SS parameters, we need $\sigma^2_h \geq 0.68$. For a model with multiple random intercepts (*eg* 3+ levels in the data hierarchy), Eq 22.2 still applies after replacing σ^2_h by the sum of all variance components. When the model includes random slopes, the variance associated with the random effects is no longer constant across the data, so the conversion depends on the values of random-effects predictors (Z). The general formula for subject i can be written (see also Hedeker & Gibbons, 2006, Section 9.7):

$$\beta_i^{PA} \approx \beta^{SS} / \sqrt{1 + 0.346 * \mathrm{var}((Zu)_i)} \qquad \textit{Eq 22.9}$$

where the variance computation follows similar lines as Eq 21.11 (depending on the details of the model).

For binary/binomial data, 2 occasionally encountered alternatives to the logit function are the so-called **probit** function (inverse cumulative probability for the standard normal) and the **complementary log-log** function. The choice of link function is largely dictated by the same considerations as for GLMs (Section 16.11); in practice, the difference in model fit and in the resulting statistical inference between the links is often minimal (see Example 22.4). For probit regression, the formulae (22.8) and (22.9) become exact when the constant 0.346 is replaced by 1. The latent variable calculation of *ICC*s is valid for probit regression as well, by replacing the constant $\pi^2/3$ by 1.

22.4.3 GLMMs for count data

Random-effects Poisson regression was introduced in Section 22.3. Compared with mixed models for binary data, the choice of mixed models for count data is considerably more diverse and confused (from an applied point of view). One reason for this is the larger selection of models for count data, including several versions of negative binomial models and zero-inflated models (Chapter 18), all of which could be extended with random effects. Another reason is that both the Poisson model and its extensions can incorporate random effects of different types and with different distributions. In this section, we briefly indicate some of the models and demonstrate their fit to the TB data of Example 22.3. A recent book on negative binomial models gives a full theoretical treatment (Hilbe, 2007).

An alternative version of the random effects Poisson regression model in Eq 22.6 assumes a log-gamma distribution instead of a normal distribution for the random effects $u_{\mathrm{herd}(i)}$. What this really means is that the herd random effects are $v_{\mathrm{herd}(i)} = \exp(u_{\mathrm{herd}(i)})$, and these are gamma-distributed variables which act as multiplicative random effects: $Y_i \sim \mathrm{Poisson}(n_i\, \lambda_i\, v_{\mathrm{herd}(i)})$. One technical advantage of this model is that its likelihood function is easier to compute which facilitates likelihood-based inference (*eg* maximum likelihood estimation).

The negative binomial distribution (in its standard form) is parametrised by the mean λ and an added dispersion parameter α (Chapter 18). Clustering of the data (*eg* in herds) may manifest itself as similarity of the means within herds (while the dispersion is constant) or conversely as similarity of the dispersion within herds (while the means are constant). These 2 scenarios would be modelled by incorporating random effects in the means or in the dispersion parameters, respectively. Perhaps the most intuitive extension of the Poisson regression model has normally distributed random effects on the log-scale for the means (in a similar fashion as

Example 22.4 Generalised linear mixed models (random effects logistic regression) for first service conception data

data = reu_cfs

In a study of reproductive measures, the success or failure of first-service conception (-fscr-) was one of the outcomes evaluated. The study comprised 3,027 lactations distributed on 1,575 cows in 50 herds. The data were analysed in a 3-level random effects logistic regression model (*ie* with random effects of cows and herds). (Lactations were not treated as repeated measures because there were few lactations in each cow). The model contained 2 dichotomous, lactation-level predictors: -heifer- (primiparous vs multiparous) and -ai- (artificial insemination vs natural breeding).

logL=-2010.85

	Coef	SE	Z	P	95% CI	
ai	-1.019	0.130	-7.81	<0.001	-1.274	-0.763
heifer	-0.064	0.097	-0.66	0.509	-0.254	0.126
constant	0.578	0.129	-	-	0.326	0.831

In addition, the estimated variances of the cow and herd random effects were, respectively:

$$\sigma_c^2 = 0.266\,(0.120) \quad \text{and} \quad \sigma_h^2 = 0.087\,(0.039)$$

Our first impression of these estimates is that there is no effect of parity and a clear, negative effect of artificial insemination on the conception rates. Both the random effects seem small but their significance is difficult to assess. The *ICC*s between 2 observations from the same cow (ρ_c) and between 2 observations on different cows in the same herd (ρ_h) can be estimated using the latent variable approach (Sections 21.2.1 and 22.2.2):

$$\rho_c(\text{lactations of same cow}) = \frac{0.266 + 0.087}{0.266 + 0.087 + \pi^2/3} = 0.097$$

and

$$\rho_h(\text{lactations of different cows in same herd}) = \frac{0.087}{0.266 + 0.087 + \pi^2/3} = 0.024$$

The next table shows the estimates for the corresponding probit random effects regression, which achieved virtually the same model fit (as measured by the log-likelihood).

logL=-2010.87

	Coef	SE	Z	P	95% CI	
ai	-0.626	0.079	-7.92	<0.001	-0.780	-0.471
heifer	-0.040	0.059	-0.68	0.499	-0.156	0.076
constant	0.355	0.078	-	-	0.201	0.509

In addition, the estimated variances of the cow and herd random effects were, respectively:

$$\sigma_c^2 = 0.100\,(0.044) \quad \text{and} \quad \sigma_h^2 = 0.033\,(0.014)$$

The probit regression estimates and SEs are scaled towards zero by a factor of roughly 1.6 (*eg* 1.019/0.626=1.63); in practice, the scaling is often in the range 1.6-1.8 and thus slightly less than the 'theoretical' scaling factor of $\pi/\sqrt{3}=1.81$ (Hedeker & Gibbons, 2006, Section 9.4). The variances are scaled by the square of this factor, and the latent variable *ICC*s are slightly larger (ρ_c=0.11).

in Eq 22.6), but it may be numerically difficult to estimate. Access to specific negative binomial random effects models varies between statistical software and may involve manual programming of the model. Example 22.5 illustrates 3 alternative Poisson and negative binomial models fit to the TB data.

An additional question arises when it comes to extension of various forms of models for zero-inflated counts (Chapter 16). As these models have different modelling equations for the zero and non-zero portions of the data, there is choice between inserting random effects in any one of these equations or in both of them. In the latter case, the model will contain 2 random effects per cluster and these should probably be correlated. Zero-inflation models with random effects are, at the current stand of statistical software, beginning to become available, and further research and applications are likely to be appear in this field.

Example 22.5 Random effects models for count data
data = tb_real

The table gives maximum likelihood estimates (SE) and the log-likelihood value for a Poisson regression model with log-gamma distributed random effects, a negative binomial regression model with normally distributed random effects for the linear predictor of the mean (log scale), and a negative binomial regression model with beta-distributed random effects for the dispersion parameter (see Cameron & Trivedie (1998) for details).

Outcome dist.	**Poisson**	**Neg. binomial**	**Neg. binomial**
Random effects dist.	**log-gamma**	**normal**	**beta**
type=beef	-0.349 (0.335)	-0.394 (0.333)	-0.363 (0.338)
type=cervid	-0.353 (0.469)	-0.238 (0.487)	-0.319 (0.479)
type=other	-0.241 (0.788)	-0.104 (0.800)	-0.167 (0.799)
sex=male	-0.352 (0.207)	-0.339 (0.208)	-0.333 (0.220)
age=12-24 mo	2.702 (0.746)	2.717 (0.747)	2.264 (0.758)
age=24+ mo	2.462 (0.726)	2.467 (0.726)	2.384 (0.742)
constant	-10.132 (0.833)	-11.056 (0.830)	-7.431 (2.234)
herd variance	1.613 (0.477)	1.688 (0.593)	n/a
log-likelihood	-146.53	-143.56	-146.39

The dispersion parameter of the negative binomial model with normally distributed random effects was estimated at zero, thereby effectively making this model identical to the Poisson model in Example 22.3 (so the original Poisson results are not repeated here). We can interpret this to say that when the overdispersion caused by the clustering is accounted for, the Poisson dispersion seems to be appropriate for the data. The negative binomial model with beta-distributed random effects is parametrised differently so that no direct comparison of the herd variance or the usual negative binomial dispersion parameter can be made. It is seen that the difference between the models in terms of their fit and also most of the parameter estimates is fairly small. When there is little difference in model fit, model choice is often guided by ease (or meaningfulness) of the interpretation of model parameters; in this case, the Poisson model with normally distributed random effects seems to be the obvious choice.

22.4.4 GLMMs for categorical data

The multinomial models of Chapter 17 can also be extended with random effects. Most attention in the literature has been paid to extensions of the proportional odds model (Section 17.5), and we illustrate the simplest of these in Example 22.6. The simple multinomial logistic regression model for nominal (Section 17.3) data can be extended with separate random effects in each model relative to the reference category (Hedeker & Gibbons, 2006, Chapter 11), and similar extensions can be proposed for other multinomial models, although such models are not generally available in statistical software.

The proportional odds model is simpler to extend by random effects than the multinomial models because the fixed effects are expressed in a single equation. Adding random effects to this equation corresponds to adding random effects to the latent (unobserved) variables S_i in Eq 17.7. The subject-specific interpretation of estimates and the latent variable method for computing *ICC*s are virtually unchanged from the logistic regression models because of the similarity of the modelling equation. Several extensions have been proposed to account for possible lack of fit, including the possibility of allowing the scale of the linear predictor to vary between suitably chosen predictor groups. For example, it could be of interest to allow the scale

Example 22.6 Random effects proportional odds model
data = beef_ultra

In Example 17.4, a proportional odds model was fit to data on ultrasound evaluation of beef cattle. The data were collected from 8 farms, which we here assume to represent a meaningful population of farms. The estimates from a model with farm random effects were as follows.

log likelihood=-374.80

	Coef	SE	Z	P	95% CI	
sex=steer	0.892	0.240	3.72	<0.001	0.422	1.363
backfat	-0.286	0.108	-2.66	0.008	-0.497	-0.075
ribeye	0.234	0.076	3.06	0.002	0.084	0.383
imfat	-0.568	0.113	5.01	<0.001	-0.791	-0.346
carc_wt	-0.019	0.003	-5.51	<0.001	-0.025	-0.012
cutpoint 1	-7.665	1.276				
cutpoint 2	-3.887	1.231				

In addition, the estimated variance of herd random effects was

$$\sigma^2_h = 0.420\,(0.264).$$

The random effects lead to some improvement in model fit (an increase in the log-likelihood of about 10 units). The estimated variance is fairly small, corresponding to an *ICC* computed by the latent variable method of 0.11. The regression coefficients are either unchanged or closer to zero, which is the opposite of what we would expect for parameters with a cluster-specific interpretation. However, a model with fixed effects for farms gave estimates even closer to zero (results not shown), so some confounding effect by farms seems to exist. The large changes in the estimated cutpoints are due to the change in the -carc_wt- coefficient and the fact that this predictor takes values in a range far beyond zero. Despite the low number of farms, the random-effects model seems to be an improvement of simple proportional odds model.

to vary between raters in datasets with items graded by multiple raters (Rabe-Hesketh & Skrondal, 2008, Chapter 7) or groups of subjects with particular characteristics related to smoking experience (Hedeker & Gibbons, 2006, Chapter 10). In Example 22.6, we illustrate the proportional odds model with random effects by the data on evaluation of beef cattle by ultrasound studied in Chapter 17.

22.4.5 Other random effects models

Mixed models with the random effects on an original scale (instead of the transformed scale as in a GLMM) do exist, and we briefly mention 2 of them here.

The **beta-binomial** model has been used extensively in veterinary epidemiology (*eg* Donald *et al,* 1994). As indicated by the name, it is a model for binomial data incorporating beta-distributed random effects for probabilities. If the 2 parameters (α_1, α_2) of the beta-distribution are expressed in terms of the mean (μ) and the *ICC* (ρ), the model can be used as a regression model by incorporating predictors into a linear predictor on logit (or probit) scale just as in a GLM. The expressions for this reparametrisation are:

$$\alpha_1 = \mu\,(1-\rho)/\rho \quad \text{and} \quad \alpha_2 = (1-\mu)(1-\rho)/\rho\,,$$

In this model, the regression parameters will have a PA interpretation. It has been recommended as one of the best models for estimating the *ICC* (Ridout *et al*, 1999). One major advantage of the beta-binomial model is that the likelihood function is given by a relatively simple and explicit formula (which is not the case for GLMMs), and therefore the model is numerically simpler to compute than GLMMs (Andreasen & Stryhn, 2008). As one of its drawbacks, it does not, in a natural way, allow for predictors at the lowest level, nor does it have any easy extension to several hierarchical levels; it is essentially a model for grouped or replicated binary data. Recall the assumed relation for the variance of the grouped (binomial) outcome (Eq 20.4); this assumption is different but not necessarily worse than other variance assumptions (*eg* the one implicit in a logistic random effects model); the fit of the beta-binomial model may be compared with that of other models by the log-likelihood or AIC statistics.

The **negative binomial** distribution was introduced in Chapter 18 as an extension of a **Poisson** distribution with overdispersion. Overdispersion could be understood as random variation in the mean (λ) of a Poisson-distributed variable (Y). Such variation may be attributed to 'inter-subject variability'—a heterogeneity between subjects not accounted for by the Poisson model. If λ has a gamma distribution with shape parameter $1/\alpha$ and scale parameter $\alpha\mu$ (equivalently: mean μ and variance $\alpha\mu^2$), then Y is a negative binomial distributed with mean μ and variance $\mu+\alpha\mu^2$, as shown in Eq 18.9. This distribution may also be called a compound or mixture Poisson model. Note that these random effects cannot be used for modelling of a hierarchical structure, because they are already incorporated into the negative binomial distribution and because they are at the lowest (subject) level. In Example 22.5, it was shown that once the farm effects of the TB data were accounted for, there was no longer any overdispersion left at the lowest (animal group) level.

22.5 STATISTICAL ANALYSIS OF GLMMS

Despite the apparent simplicity of models such as Eq 22.1 and Eq 22.6, analysis of GLMMs is not straightforward, even in the logistic and Poisson regression settings. In contrast to most

other models in the book, even the estimation of parameters is not clear-cut. Different methods exist, and they may give appreciably different results. No definitive answer exists at this point as to which method is generally preferable. The maximum likelihood method has been used throughout most parts of the book and is considered the standard choice here as well if it does not pose unmanageable computational challenges. Advances in computing power and software implementations over the last years has made maximum likelihood estimation feasible for moderate to large datasets and models. The implementation of GLMMs is still an active research area, and one is advised to investigate the options in different software before deciding on an approach (see also Section 22.6 for notes on current software). We outline briefly the methods available and indicate where they are discussed in this text.

1. Maximum likelihood estimation (Section 22.5.1): the likelihood function involves an integral over each random effect, which must be approximated by a summation and therefore makes ML estimation computationally demanding for large models.
2. Quasi-likelihood or iterative weighted least squares estimation (Section 22.5.2): algorithms for linear mixed models and GLMs are combined to produce multiple slightly different variants of an algorithm, which is fast and computationally simpler than ML estimation.
3. Bayesian MCMC (Markov chain Monte Carlo) estimation (Chapter 24): based on a different statistical approach (Bayesian statistics) and a simulation-based estimation that is computationally intensive.

All results shown so far in this chapter have been from ML estimation. But how does one determine which method is best, in general, for one's own data? One standard answer is to use simulation, *ie* generate artificial data from a model with known values of all parameters and then compare the results of different methods with those known values. Such simulation studies are regularly published in statistical journals (*eg* Browne & Draper (2006), or Masaoud & Stryhn (2009)), and you could also carry out your own simulation study for the data structure at hand (Stryhn *et al*, 2000).

Even if we as researchers are committed to always using the best possible method to analyse our data, it is useful to have a sense of when major differences between approaches might appear (see also the discussion of biases by quasi-likelihood procedures in Section 22.5.2). Estimation in GLMMs is most difficult if variances are large and/or the information contained in the data is limited, *eg* if replication is sparse. It is generally true that binary data are more difficult than count data, and that we should avoid fitting too ambitious models to even moderately sized binary datasets. Another possible cause of problems is if multiple clustering units should have 'extreme' predicted values, *eg* if in a logistic model all animals in a herd are negative (or all are positive). It almost goes without saying that whenever a dataset or model shows signs of being 'difficult' to estimate, one should be particularly careful with the analysis, and in such cases analysis by different procedures is often a fruitful approach.

22.5.1 Maximum likelihood estimation

Maximum likelihood (ML) estimation in GLMMs would, at first sight, seem to be our first choice, because of the overall strengths of the method (good statistical properties of the ML estimates) and the access to likelihood-based inference (*eg* likelihood ratio tests). However, ML estimation has, until recent years, had the reputation of being unfeasible for any GLMM beyond the simplest 2-level models, due to the massive and difficult computations required. Recent advances in computer power and software have changed this judgement, although the options

currently available vary considerably between statistical software. It is likely that, within a few years, ML estimation will become the standard estimation approach for all but huge GLMMs. Even if the method's numerical side now looks promising, we outline why computation of the likelihood function is so difficult and give some cautions (complex procedures always have pitfalls, even if the complexities are hidden in the software).

For simplicity, consider the 2-level logistic regression model (Eq 22.1) and let us begin by focusing on a single herd — herd 1. Given the value of u_1 (the random effect of herd 1), the conditional likelihood of the observations from that herd is binomial,

$$L_1(\beta|u_1) = \prod_{i:\,\text{herd}(i)=1} p_i^{Y_i}(1-p_i)^{1-Y_i}$$

and the full (sometimes denoted marginal) likelihood for those animals is obtained by integration over the distribution of the random effect u_1:

$$L_1(\beta) = \int L_1(\beta|u_1)(2\pi\sigma_h)^{-1/2}\exp(-\frac{1}{2}u_1^2/\sigma_h^2)\,du_1 \qquad \textit{Eq 22.10}$$

The integration weights the possible values of u_1 according to their likelihood in a normal distribution with mean zero and standard deviation σ_h. Integrals such as Eq 22.10 cannot be solved analytically, and therefore a numerical integration or **quadrature** becomes necessary. By this procedure, the integral is approximated by a weighted sum of values of the integrand (*ie* the function being integrated) at a number of selected quadrature points. Specific weighting schemes for integrals that involve exponential terms of a squared argument, as in Eq 22.10, are called Gauss-Hermite quadrature. In such schemes, you need to decide on the number of quadrature points and the way they are selected. Generally, increasing their number improves both accuracy and calculation time. Also, it is generally recommended that an **adaptive** approximation method be used, where the quadrature points (and their weights) are successively adapted to the integrand.

So far, we have dealt only with observations from one herd. Observations from different herds are independent, so the full likelihood function for the entire dataset is obtained as a product of terms such as Eq 22.10 over the total set of herds. We trust it is not necessary to write out the equation to make the point that, not only computing, but also maximising, a quadrature approximation to such a multiple integral with respect to the fixed and random effects parameters of the model can be a formidable task. Extension to multiple levels and/or multiple random effects at the same level rapidly increases the complexity of the problem.

To summarise, a few recommendations and cautions for the use of ML estimation for GLMMs:

- ML estimation might be computationally unstable or the approximation of the likelihood function may be insufficient; it is highly recommended, therefore, that the stability of results be checked by trying different starting values of the algorithm and/or different variants of the numerical integration procedure, such as a different number of quadrature points as well as adaptive procedures,
- ML estimation could be compared with other approaches (either quasi-likelihood estimation or other approaches for clustered data), and caution should be exercised if major differences appear; this is in particular recommended if the estimation problem is 'difficult' (as discussed above),
- ML estimation in GLMMs may be impractical for model selection (because of computational demands); it is then considered legitimate to use computationally simpler methods for (part of) the model selection and confirm the results by running

selected models also by ML estimation.

In Example 22.7, we examine the stability of the quadrature behind the ML estimates.

22.5.2 Quasi-likelihood estimation

A quasi-likelihood function could be thought of as a substitute for a (real) likelihood function whenever the latter does not exist or is too difficult to compute. In the early 1990s, when computers were much less powerful, several algorithms employing an iterative weighted least squares scheme were developed to maximise quasi-likelihood functions for GLMMs. These algorithms are referred to by many different acronyms, typically containing the letters QL (for quasi-likelihood), PL (for pseudo-likelihood) or ILS (for iterative and least squares), and often in combination with a G for generalised or a W for weighted or an R for reweighted or restricted. The main idea of the iterative weighted least squares methods is to compute an 'adjusted' variate on the scale given by the link function (*eg* logistic scale) in each step of the iteration. Technically, the adjusted variate is obtained by a Taylor expansion of Y around the current estimated mean, but one may think of it as a continuous version of the discrete outcome. Estimation for this adjusted variate is carried out using estimation procedures for linear mixed models (weighted REML or ML estimation). The procedure continues until convergence of the parameter estimates. Again, for the technically interested reader, some common options in the procedure are mentioned below:

- first- or second-order Taylor expansion, the latter being considered more accurate whenever the procedure converges,
- ML or REML estimation for the adjusted variate, the latter being the more commonly used,

Example 22.7 Checking maximum likelihood estimation of a GLMM
data = reu_cfs

In Example 22.4, ML estimation was used to fit a random effects logistic regression model to the Reunion Island first-service conception risk data. This model was refit using a range of number of quadrature points (at both the herd and cow level) in the estimation procedure. The fixed and random effects estimates from each estimation were:

	Number of quadrature points at (herd,cow) level			
	(1,1)	**(3,3)**	**(7,7)**	**(12,12)**
ai	-0.993	-1.019	-1.019	-1.019
heifer	-0.065	-0.064	-0.064	-0.064
constant	0.568	0.578	0.578	0.578
herd variance	0.083	0.087	0.087	0.087
cow variance	0.145	0.267	0.266	0.266
log likelihood	-2012.446	-2010.839	-2010.847	-2010.847

The default in the software used for the estimation is 7 quadrature points. Using a single (1) quadrature point is sometimes referred to as a (low order) Laplace approximation (Section 22.5.2). With 20 quadrature points, exactly the same estimates were obtained as with 12 points. This model seems very stable because the estimates changed very little as the number of quadrature points was increased, and the default number of quadrature points seems adequate.

- MQL or PQL form of the adjusted variate (M=marginal, P=predictive or penalised), the former being computationally more robust by omitting estimates of random effects in the linear predictor, and yields estimates with a PA interpretation (Breslow & Clayton, 1993), contrary to the other procedures for estimation in a GLMM.

These 3 options can be combined arbitrarily (depending on the facilities of the software package used). Example 22.8 shows results from some of these algorithms.

A long list of (statistical) papers from the 1990s discussed the different versions of algorithms and their implementation in software packages (*eg* Browne & Draper (2006); Zhou *et al* (1999)). For well-behaved data, the different variants of the algorithms give very similar results (taking into account the standard errors of the estimates). One should whenever possible use the 'best' possible of the above options (second order, REML, PQL). More importantly, any 'strange-looking' estimates or standard errors should cause the model to be examined carefully and the results to be confirmed with other models or estimation methods.

Early simulation studies showed that estimates from some of the iterative least squares algorithms for GLMMs could be markedly biased towards the null. The bias might affect both fixed and random-effect parameters, but the latter are particularly sensitive. The general consensus seems to be that particular caution should be exercised if:

- the number of replications at a hierarchical level is 'small' (*eg* less than 5),
- the corresponding random effect is 'large' (*eg* the variance exceeds 0.5).

Example 22.8 Quasi-likelihood estimation of a GLMM
data = reu_cfs

Three quasi-likelihood estimation procedures were applied to the first-service conception data of Example 22.4. The estimates were obtained using MLwiN software; for the 1st order procedures, the same estimates were obtained in SAS (Proc Glimmix) with slightly larger SE for the variance parameters.

	ML (Ex 22.4)	1st order MQL	1st order PQL	2nd order PQL
	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
ai	-1.019 (0.130)	-0.941 (.120)	-0.953 (.121)	-0.995 (.123)
heifer	-0.064 (0.097)	-0.062 (.092)	-0.062 (.093)	-0.064 (.093)
constant	0.578 (0.129)	0.540 (.120)	0.545 (.122)	0.567 (.123)
herd var.	0.087 (0.039)	0.079 (.032)	0.080 (.032)	0.088 (.034)
cow var.	0.266 (0.120)	0.100 (.076)	0.130 (.078)	0.153 (.080)

The estimates of all quasi-likelihood procedures are generally closer to zero than the ML estimates in Example 22.4. The largest disagreement is seen for the cow-level variance which is about 40-60% of the ML estimate. We interpreted the disagreement about this value as a bias of the quasi-likelihood estimation procedure; a simulation study confirmed that with these data the quasi-likelihood procedure would consistently give too low cow-level variance estimates (Stryhn *et al*, 2000). The bias is less pronounced with the 2nd order than the 1st order PQL procedure. The MQL estimates for the regression coefficients have a PA interpretation (and should therefore be closer to zero than the SS counterparts), but the variances are the most severely affected by the bias towards zero. Note that the quasi-likelihood procedures do not allow easy computation of test statistics or confidence intervals for variance parameters.

In our Example 22.8, the number of cow-level replications was indeed small, with only an average of 1.9 observations per cow. The fact that the biases in the regression coefficients were still fairly small is due to the small variances at the 2 hierarchical levels, and the biases can be more substantial in datasets with larger variance components.

We finally also mention a related type of approximation method that goes under the name Laplace approximation. These methods have been continually developed in the last decade, involving high order (*ie* more accurate) approximations and now also providing useful approximations to the log-likelihood function (so as to enable likelihood-based inference). A recent paper concluded that Laplace approximation methods were fully acceptable for Poisson regression (Pinheiro & Chao, 2006), and a further simulation study explored the accuracy of the method for both count and binary data (Joe, 2008).

22.5.3 Confidence intervals and tests

Statistical inference in GLMMs is generally only approximative (asymptotically correct when the number of observations at all hierarchical levels is large). Fixed effects parameters are usually assessed by Wald-type confidence intervals and tests, however likelihood-based inference (profile likelihood CIs and likelihood ratio tests, see Section 21.5) may be preferable, in particular when the parameters are highly correlated or not well determined. However, likelihood-based inference is only feasible when ML estimation is used. As with GLMs, Wald-type statistics are useless for parameters that are 'out of bounds', *eg* in logistic regression when one category of a predictor has no cases. Such situations often signal separation issues (Heinze & Schemper, 2002), which would typically also affect estimation of the random effects (therefore, also ML estimates may be affected).

Reference distributions are most commonly 'asymptotic', *ie* the standard normal or χ^2-distributions. The resulting inference may be too liberal if replication is sparse at the level of the parameter of interest (Stryhn *et al*, 2000), and some software packages give the option of using similar approximations as for linear mixed models using *t*- and *F*-distributions. In general, no clear guidelines can be given about the accuracy of approximative inference in GLMMs. As in linear mixed models, Wald-type statistics are inappropriate for variance parameters, which should therefore be assessed by likelihood-based inference (Example 22.9) or alternative procedures such as bootstrapping.

Example 22.9 Statistical inference in a GLMM
data = reu_cfs

The tests and confidence intervals given for the fixed effects in Example 22.4 are 'asymptotic'; for example, the 95% CIs are computed as $\beta \pm 1.96 * SE(\beta)$. Both predictors vary (potentially) at the lowest level and even if some variation may reside at higher levels (in particular, the cow level) there would seem ample replication to justify inference based on the standard normal distribution.

To compute tests for the random effects of the model, we note the log-likelihood value of the fitted model (-2010.85) and refit the model without the random effect of interest. The models without cow-random effects and herd-random effects had log-likelihood values of (-2014.11) and (-2017.93), respectively, so that the corresponding χ^2-statistics with 1 df were 6.52 and 14.2, and thus both significant. Recall from Section 21.5.3 that P-values should be computed as half the tail probability from the $\chi^2(1)$-distribution to account for the one-sided alternative hypothesis. It is perhaps interesting to note that the herd-random effect was clearly the most significant of the 2; this is not at all obvious from the estimates and standard errors.

22.5.4 Prediction

The random effects in GLMMs can be predicted along similar lines as in linear mixed models (Section 21.5.4) and also exhibit a shrinkage towards the mean. However, some new issues arise for predictions of observations or their means, because the fixed and random effects reside on a different scale (*eg* logit scale). This is related to the distinction between SS and PA parameters. Following Skrondal & Rabe-Hesketh (2009), we describe in Example 22.10 3 different ways of computing predicted probabilities in a random effects logistic regression

Example 22.10 Prediction in a random-effects logistic regression
data = pig_adg

We consider again the logistic regression model for pneumonia in 341 pigs from 15 farms with the presence of atrophic rhinitis (-ar_g1-) as the sole predictor and farm random effects. The ML estimates for the intercept, coefficient for -ar_g1- and the between-farm variance were, respectively:

$$\beta_0 = 0.020, \quad \beta_1 = 0.437, \quad \sigma^2 = 0.877$$

We wish to predict the probability of pneumonia for pigs with and without atrophic rhinitis. Three possible interpretations exist for such probabilities:

1. **Probability for pigs in a hypothetical farm** For any given (hypothetical) farm random effect u, we can compute a farm-specific (conditional) probability as: $p(1)=\text{logit}^{-1}(\beta_0+\beta_1 \text{ ar\_g1}+u)$. Using $u=0$ gives the median probability across the population of farms. We can also insert $u=\pm 1.96\sigma$ to get a 95% range across the population of farms.

2. **Mean probability for pigs from any farm** The approximation formula Eq 22.2 gives this population-averaged probability as: $p(2)\approx\text{logit}^{-1}((\beta_0+\beta_1 \text{ ar\_g1})/\sqrt{(1+0.346\ \sigma^2)})$. In the table below, we used a more exact approximation for $p(2)$ based on quadrature (see Skrondal & Rabe-Hesketh (2009) for details).

3. **Probability for pigs in a study farm** When predicting for a study farm (say farm 1), we need to incorporate the information we have about its random effect. Due to the non-linearity of the logit function, simply inserting the predicted farm random effect does not work exactly; instead, we need to compute the mean probability averaged across the posterior distribution (in Bayesian terminology, see Chapter 24) of the random effect. Some statistical software will provide this calculation ($p(3)$).

The table below gives the 3 probabilities for the 2 categories of the predictor and some farms.

ar_g1	farm #	u(farm)	p(1) with u=0	p(2)	p(1) with u(farm)	p(3)
0	1	1.115	0.505	0.504	0.757	0.748
	3	-1.457			0.192	0.200
1	1	1.115	0.612	0.595	0.828	0.819
	3	-1.457			0.269	0.276

As the 2 fixed effects parameters are of moderate size, the first 2 probabilities are fairly close both to each other and to 0.5. The predicted probabilities in study farms with moderately sized random effects are very different, making again the point (from Example 22.2) that the random effects are stronger than the fixed effect. The calculation with the estimated random effect inserted gives only slightly different values than the correctly calculated probability $p(3)$.

model. The first 2 are commonly used; see *eg* McClure *et al* (2005). In order to demonstrate at least moderate differences between the approaches, we use the 2-level pig dataset from Example 22.1 where the between-herd variance is larger than in the Reunion Island dataset.

22.5.5 Residuals and diagnostics

The standard tools for model-checking—residuals and diagnostics—are even less developed and accessible for GLMMs than for linear mixed models (Section 21.5.5). The distinction between different types of standard errors still holds, but calculations are more difficult, and in practice one may need to accept whatever is offered by the statistical software (Skrondal & Rabe-Hesketh, 2009). The main new point for GLMMs (compared with linear mixed models) is that, because the model has no normally distributed error terms at the lowest level, the corresponding residuals and diagnostics at that level are difficult to assess. As an extreme example, in a binary model all the lowest-level residuals are dichotomous and cannot be expected to conform to a normal distribution. In this case, the residuals at the lowest level are not very informative. Unfortunately, the problems with the lowest-level residuals could penetrate to the higher levels if there is little replication. Reference distributions and points for residuals and diagnostics are therefore difficult to use rigorously, and one is advised instead to look for data points that are extreme in some way relative to the rest of the data. Example 22.11 presents residuals from our 3-level Reunion Island analysis.

GLMM analogues of some of the special statistics for discrete data, such as the Hosmer-Lemeshow test for goodness of fit in a logistic-regression model, are not available. A recent paper described a simulation-based goodness-of-fit test for GLMMs, but this procedure does not seem to be available in standard software (Waagepetersen, 2006).

22.5.6 Robustness against model misspecification

Much of the discussion of robustness of linear mixed model analysis to model specification in Section 21.5.7 carries over to GLMMs. One notable difference is that the use of robust standard errors is less obvious with non-normal data. A substantial body of research has been undertaken in the last decade on misspecification of GLMMs, in particular by McCulloch and Neuhaus whose work is summarised in McCulloch *et al* (2008), Chapter 12. Additional work includes Heagerty & Kurland (2001), and Litière *et al* (2008). One conclusion that seems to be common for this work is that misspecification of the random effects distribution may not be terribly serious (McCulloch & Neuhaus, 2009).

22.5.7 Over- and underdispersion in GLMMs

Non-distributional dispersion in GLMs was discussed in Section 20.5.3, and it was explained how an extra-binomial dispersion parameter could be added to a binomial model within the GLM framework. A similar multiplicative dispersion parameter φ can be added to a Poisson model by the specification that $\text{var}(Y_i)=\varphi\lambda_i$, where $\varphi=1$ corresponds to the usual Poisson distribution, and also to other models for count and categorical data. We discussed in Section 20.5.3 how the extra-binomial parameter could account for clustering, although this method was not as attractive as other modelling approaches such as mixed models. In Chapter 16, we also discussed other ways that an apparent over- or underdispersion could arise. The question we address here is the utility of allowing for extra-distributional dispersion in mixed models, where the random effects should account for the hierarchical structure.

Example 22.11 Residuals from a 3-level GLMM
data = reu_cfs

The 3-level logistic regression of Example 22.4 (Reunion Island first-service conception data) has residuals at all 3 hierarchical levels but the lowest-level residuals are of little use in this case so we disregard them completely. A normal (Q-Q) plot for the 1,575 cow-level standardised residuals is given in Fig. 22.2. The plot shows a curious pattern, far from a straight line but instead with 3 separate, almost straight, lines. One must realise that with typically only 1-3 observations per cow and only 4 different sets of predictor values, the cow-level residuals cannot realistically be expected to look like a normal distribution sample. With closer scrutiny, each approximately linear part of the graph correspond to cows with the same response pattern. For example, the lower part of the plot corresponds to cows without any first-service conceptions in the dataset and the upper part of the plot to cows that conceived at first service in all

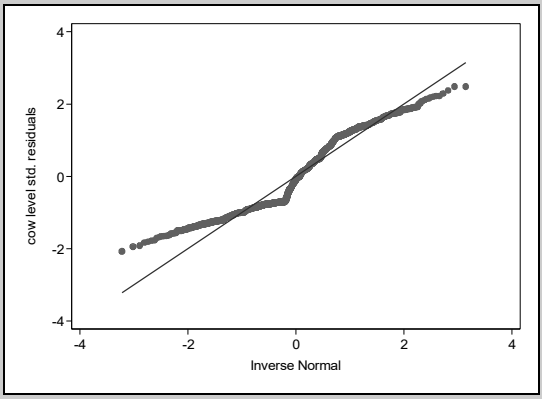

Fig. 22.2 Normal plot for cow-level residuals of 3-level model for Reunion Island data

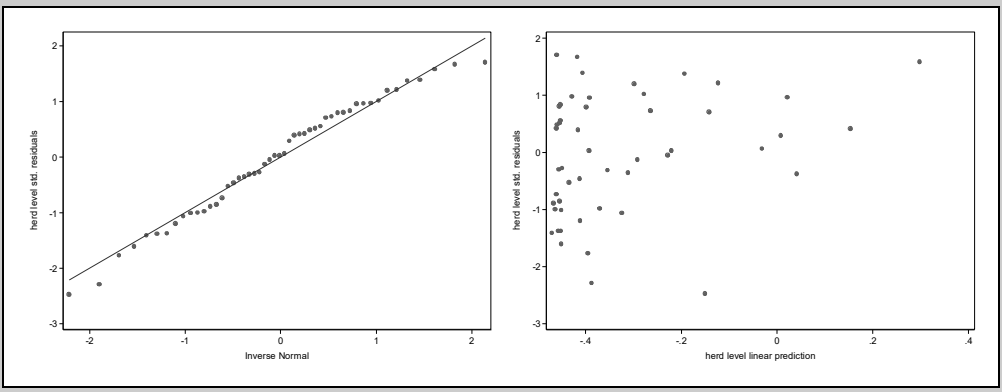

Fig. 22.3 Normal plot (left) and plot against predicted values (right) for herd level residuals of 3-level model for Reunion Island data

lactations. It seems almost impossible to assess from the plot whether there are problems with the model assumptions at the cow level.

Fig. 22.3 shows the herd-level residuals depicted in a normal plot and plotted against the herd-level predicted values (including cow-level predictors). The normal plot is somewhat skewed due to lack of herds with strongly positive residuals; however, when comparing with the lower tail of the distribution, you can see that only 2 negative residuals are more extreme than in the upper tail. The plot against the fitted values shows a grouping of predicted values at the lower end of the scale but no particular patterns in the residuals.

Our first observation is that extra-distributional parameters only exist within the GLM framework where models are incompletely specified and estimation is based on quasi-likelihood-type functions (Section 22.5.2). The fact that no data-generating mechanism exists for these models has been put forward as a major disadvantage of the approach (Skrondal & Rabe-Hesketh, 2007). It does indeed seem awkward to recommend the use of an approach that is only available in a subset of less attractive estimation procedures for GLMMs. Nevertheless, it is conceivable that a dataset could contain a dispersion that does not match the 'natural' distribution, even after the fixed and random effects have been incorporated into the model. In this sense, inclusion of an extra-distributional parameter may serve as a diagnostic tool. Values of φ substantially different from 1 would then lead us to either explore different distributions (where feasible), adopt the scaling of standard errors implicit in the quasi-likelihood estimation procedures, or perhaps ignore the finding.

If underdispersion is indicated, one should look for any reasons for negative correlations between observations, the standard example being competition in a group of animals for a limited resource (*eg* feed). If no such explanation can be found, as underdispersion means a better fit than expected to the data of our model, we often tend not to worry much about it (maybe it was just 'good luck'). By ignoring an appreciable underdispersion and pretending the dispersion to be as predicted by our model (when it is in reality smaller), our statistical inference becomes conservative—which may be considered the appropriate approach for 'a case of good luck'. Underdispersion (as well as very small values of one-sided test statistics) may, however, also indicate something strange to be going on in the data, so one should inspect the data critically (once more). To scale down the standard errors by an underdispersion factor is a serious decision because it may lead to spurious significance, and should probably only be done when there is a biological explanation of the phenomenon. It might be useful to also try robust standard errors (Section 20.5.4) to see if they point in the same direction. Example 22.12 presents a dataset where an extra-binomial parameter substantially less than 1 was encountered.

Overdispersion may be easier to understand intuitively and it may be considered less serious to inflate the standard errors; again, a comparison with robust standard errors might be useful. In some special cases, specific advice can be given on the modelling. First, if overdispersion is encountered in a Poisson model, it seems natural to try instead a negative binomial distribution

Example 22.12 Aggregation of the lowest level for pig-seroconversion data
data = ap2

Vigre *et al* (2004) observed that seroconversion to Actinobacillus pleuropneumoniae was strongly clustered in batches of pigs in multisite production systems. On average, each of the 36 batches of pigs consisted of about 30 pigs, and in 17 batches more than 90% of the pigs seroconverted, in 4 batches between 50% and 85% of the pigs seroconverted, and, in the remaining 15 batches, no pigs seroconverted. A 3-level logistic regression model (with predictors at all levels) showed a 'dispersion parameter' of φ=0.2, despite the fact that underdispersion cannot exist with binary data. It was unclear what model deviation (if any) the 'dispersion parameter' picked up in the data. However, the 3-level analysis also exhibited some numerical instability when fitted using quasi-likelihood estimation, and had a large variance component at the batch level. It was therefore decided to aggregate the data to the batch level by defining a batch as positive if at least one pig seroconverted, and as negative if otherwise. The pig's mean age at slaughter was computed for each batch as well, and the data were analysed by a 2-level model using this batch-level predictor. From a biological perspective, it was considered perfectly acceptable to designate batches as seroconverted or not, given the strong clustering in batches, so the 2-level model was preferred to the 3-level model.

(Chapter 18). Second, if overdispersion is encountered in a (mixed) model for grouped binary data (*ie* a binomial model with denominator >1), one may introduce a random effect at the group level which will then effectively remove the overdispersion (Browne *et al*, 2005; Skrondal & Rabe-Hesketh, 2007). Third, if the outcome is binary, extra-binomial dispersion cannot exist (Skrondal & Rabe-Hesketh, 2007); this is the same situation as in an ordinary GLM (Section 16.12). The same consideration exists for single categorical observations in a multinomial model. Notwithstanding this fact, many quasi-likelihood estimation procedures allow estimation of an 'extra-binomial' parameter for binary data, and many examples exist of such models fitted and published (Skrondal & Rabe-Hesketh, 2007). It is not clear what these estimation procedure actually estimate in the data, and interpretation of the estimated value of φ beyond a nondescript 'diagnostic' is hard to give. Skrondal and Rabe-Hesketh argue that the extra-dispersion parameter should be avoided in these instances.

22.6 SUMMARY REMARKS ON ANALYSIS OF DISCRETE CLUSTERED DATA

Throughout this chapter we have emphasized the distinction between cluster-specific (SS) and marginal (PA) modelling and interpretation of effects. We have also noted that SS parameters reside on a different scale than PA parameters, and that the difference between the 2 scales depends on the magnitude of the variance components (*eg* Eq 22.2). This scaling of SS parameters (relative to PA parameters) by a factor depending on the variances has the, perhaps undesired, consequence that SS parameters become difficult to compare between different datasets and analyses. As the estimates of variance parameters are particularly sensitive to the choice of estimation procedure, the fixed effects will, whenever the variances are large, be equally sensitive. This was the main reason behind our recommendation in Section 22.5 to exercise particular caution with the analysis when variances are large. In situations where the interest is in PA parameters, it seems awkward to start the process by obtaining estimates on another scale that may be difficult to establish firmly (when variances are large) before converting back to the scale of interest. Such reasoning has spurred the development of **marginalised models**, in which fixed effects are modelled on PA scale while a random structure is retained on SS scale (Diggle *et al*, 2002, Chapter 11). Although first results with this new class of models were promising (Heagerty & Zeger, 2000), these models have not gained much popularity because they are not available in standard statistical software.

Several topics for mixed models covered in Chapter 21 (*eg* sample size) has not received a special treatment in this chapter because the coverage in Chapter 21 largely carries over or gives the relevant pointers also for GLMMs. The literature on GLMMs is huge and still rapidly expanding, including in recent years many excellent textbooks (often also covering linear mixed models, see the brief overview in Section 21.5, and/or repeated measures data, see Chapter 23). Let us at this point attempt a brief summary of the current stand of statistical software for GLMMs. The field is more diverse and confusing than for linear mixed models, due to the existence of different estimation procedures and the continuing emergence of algorithms improved in speed and flexibility. For **maximum likelihood estimation** by numerical integration, Stata is arguably the most versatile statistical software package, because it offers both standard multilevel routines for binomial and count data with no restrictions on the number of hierarchical levels and cross-classification, and the powerful Generalised Linear Latent And Mixed Models (-gllamm-) macro for multilevel modelling implemented in Stata (Rabe-Hesketh & Skrondal, 2008). The -gllamm- software also implements a wide range of models involving latent variables (Skrondal & Rabe-Hesketh, 2004). Implementations of ML

estimation in other packages is more limited, but updates are likely to occur rapidly. For **quasi-likelihood estimation,** many different implementations exist, both in general-purpose statistical software (*eg* SAS, R/S-Plus) and in specialised multilevel packages (MLwiN, HLM), with variable accuracy and flexibility of the algorithms. Another specialised package for mixed models (AD Model Builder) offers high order Laplace approximations.

A variety of approaches for dealing with clustered data has been presented in this and previous chapters, and 2 more are to come in Chapters 23 and 24. We conclude with a comparative table of estimates for the pig-pneumonia data (Example 22.13).

Example 22.13 Summary of analyses for pig-pneumonia data
data = pig_adg

Variable	Model	β	SE
ar_g1	logistic	0.647	0.220
	robust variance	0.647	0.276
	fixed effects	0.365	0.268
	stratification	0.349	0.261
	GLMM	0.437	0.258
	GEE	0.354	0.215
	Bayesian	0.438	0.260

The GEE estimation used an exchangeable working correlation structure (Section 23.4). The Bayesian analysis reported the posterior median and SD from a mixed model with standard flat priors (Chapter 24).

Accepting that the GLMM and GEE estimates are the best SS and PA estimates respectively, it is surprising to see the fixed effects and stratified estimates (also SS) closer to the GEE results. The logistic and robust variance estimates fail to account for the strong confounding effect of herd.

REFERENCES

Andreasen C, Stryhn H. Increasing weed flora in Danish arable fields and its importance for biodiversity Weed Research (Oxford). 2008; 48: 1-9.

Breslow N, Clayton D. Approximate inference in generalized linear models J Am Stat Assoc. 1993; 88: 9-25.

Browne W, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models Bayesian Analysis. 2006; 3: 473-514.

Browne W, Subramanian S, Jones K, Goldstein H. Variance partitioning in multilevel logistic models that exhibit over-dispersion J R Stat Soc A. 2005; 168: 599-613.

Cameron A, Trivedie P. Regression Analysis of Count Data. Cambridge University Press; Cambridge. 1998.

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

Donald AW, Gardner IA, Wiggins AD. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors Prev Vet Med. 1994; 19: 167-87.

Goldstein H, Browne W, Rasbash J. Partitioning variation in multilevel models Understanding Statistics. 2002; 1: 223-32.

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

Heagerty P, Zeger S. Marginalized multilevel models and likelihood inference Statistical Science. 2000; 15: 1-26.

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

Heinze G, Schemper M. A solution to the problem of separation in logistic regression Stat Med. 2002; 21: 2409-19.

Hilbe J. Negative Binomial Regression. Cambridge University Press; Cambridge. 2007.

Joe H. Accuracy of Laplace approximation for discrete response mixed models Computational Statistics & Data Analysis. 2008; 52: 5066-74.

Larsen K, Petersen JH, Budtz-Jørgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects Biometrics. 2000; 56: 909-14.

Litière S, Alonso A, Molenberghs G. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models Stat Med. 2008; 27: 3125-44.

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.

McClure CA, Hammell KL, Dohoo IR. Risk factors for outbreaks of infectious salmon anemia in farmed Atlantic salmon, Salmo salar Prev Vet Med. 2005; 72: 263-80.

McCulloch C, Neuhaus J. Misspecification of the Distribution and Prediction of Random Effects in Generalized Linear Mixed Models. University of California; San Francisco. 2009.

McCulloch C, Searle S, Neuhaus J. Generalized, Linear, and Mixed Models, 2nd Ed. Wiley; New York. 2008.

Otte MJ, Gumm ID. Intra-cluster correlation coefficients of 20 infections calculated from the results of cluster-sample surveys Prev Vet Med. 1997; 31: 147-50.

Pinheiro J, Chao E. Efficient laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models J Comput Graph Stat. 2006; 15: 58-81.

Rabe-Hesketh S, Skrondal A. Multilevel and Longitudinal Modeling using Stata, 2nd ED. Stata Press; College Stn. Tx. 2008.

Ridout MS, Demétrio CG, Firth D. Estimating intraclass correlation for binary data Biometrics. 1999; 55: 137-48.

Rodriguez G, Elo I. Intra-class correlation in random-effects models for binary data The Stata J. 2003; 3: 32-46.

Skrondal A, Rabe-Hesketh S. Generalized Latent Variable Modelling. Chapman & Hall/CRC; London. 2004.

Skrondal A, Rabe-Hesketh S. Redundant overdispersion parameters in multilevel models for categorical responses Journal of Educational and Behavioral Statistics. 2007; 32: 419-30.

Skrondal A, Rabe-Hesketh S. Prediction in multilevel generalized linear models J R Stat Soc A. 2009; 172: 659-87.

Snijders T, Bosker R. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling. Sage Publications; London. 1999.

Stryhn H, Dohoo I, Tillard E, Hagedorn-Olsen T. Simulation as a tool of validation in hierarchical generalized linear models. In: Proceedings of International Symposium for Veterinary Epidemiology and Economics, Breckenridge, Colorado. 2000.

Stryhn H, Sanchez J, Morley P, Booker C, Dohoo I. Interpretation of variance parameters in multilevel Poisson regressions models. In: Proceedings of International Symposium on Veterinary Epidemiology and Economics, Cairns, Australia. 2006.

Vigre H, Dohoo IR, Stryhn H, Busch ME. Intra-unit correlations in seroconversion to *Actinobacillus pleuropneumoniae* and *Mycoplasma hyopneumoniae* at different levels in Danish multi-site pig production facilities Prev Vet Med. 2004; 63: 9-28.

Waagepetersen R. A simulation-based goodness-of-fit test for random effects in generalized linear mixed models Scandinavian Journal of Statistics. 2006; 33: 721-31.

Zhou X, Perkins A, Hui S. Comparisons of software packages for generalized linear multilevel models American Stat. 1999; 53: 282-90.