BIOMETRICS 56, 909–914
September 2000

Interpreting Parameters in the Logistic Regression Model with Random Effects

Klaus Larsen,[1,*] Jørgen Holm Petersen,[1] Esben Budtz-Jørgensen,[1] and Lars Endahl[2]

[1]Department of Biostatistics, University of Copenhagen,
Blegdamsvej 3, DK-2200 Copenhagen N, Denmark
[2]National Institute of Occupational Health, LersøParkallé 105, DK-2100 Copenhagen Ø, Denmark
*email: k.larsen@biostat.ku.dk

SUMMARY. Logistic regression with random effects is used to study the relationship between explanatory variables and a binary outcome in cases with nonindependent outcomes. In this paper, we examine in detail the interpretation of both fixed effects and random effects parameters. As heterogeneity measures, the random effects parameters included in the model are not easily interpreted. We discuss different alternative measures of heterogeneity and suggest using a median odds ratio measure that is a function of the original random effects parameters. The measure allows a simple interpretation, in terms of well-known odds ratios, that greatly facilitates communication between the data analyst and the subject-matter researcher. Three examples from different subject areas, mainly taken from our own experience, serve to motivate and illustrate different aspects of parameter interpretation in these models.

KEY WORDS: Interpretation; Interval odds ratio; Logistic regression; Median odds ratio; Normally distributed random effects.

1. Introduction

Correlated binary data frequently occur in several fields of research, e.g., medical, social, psychological, economics, and agricultural research. The studies may have different designs, such as longitudinal studies, where responses are measured on the same subject repeatedly over time, genetic studies, where family members are closely related, or agricultural studies, where naturally defined clusters (e.g., pigs in pigsties) give rise to correlated responses within clusters. No matter how the study is designed, one must incorporate the correlation in the model to get the correct estimates and draw the proper conclusions.

The analysis of correlated binary data is not as straightforward as multivariate normal analysis. This is mainly due to the lack of a meaningful joint distribution (Bishop, Fienberg, and Holland, 1975; Liang, Zeger, and Qaqish, 1992). Therefore, at least three different alternatives to modeling the joint distribution have been proposed. These are threshold models (Ashford and Sowden, 1970), GEE models (Zeger, Liang, and Albert, 1988; Lindsey and Lambert, 1998), and random effects models (Williams, 1975; Stiratelli, Laird, and Ware, 1984). We will concentrate on one of the latter, namely the logistic regression model with normal random effects (LRRE).

To investigate the occurrence of *Ascaris suum* (roundworm) in pigs, samples of feces were taken from 1016 slaughter hogs and gilts in Denmark (see Roepstorff et al., 1998). Two different types of pigsties were included in the study—conventional and the so-called SPF. The data from the 72 conventional and 36 SPF pigsties are given as fractions in Table 1. The goal was to investigate whether pigs in SPF pigsties have a lower risk of being infected than pigs in conventional pigsties, and if so, to quantify the difference. The naive approach neglecting the clustering of pigs in pigsties provides an odds ratio for infection between conventional and SPF pigsties of 2.76 ($\chi^2_1 = 17.0$). Using this approach, it is implicitly assumed that the odds ratio of infection is the same for any two pigs—comparing conventional to SPF—which is not true, as will be shown in Section 2. To model the heterogeneity among pigsties, we introduce the logistic regression with random effects. Each pigsty is assumed to have its own level of infection; therefore, the fixed effects odds ratio no longer is an odds ratio between any two pigsties, and the above interpretation does not apply. Further, modeling the heterogeneity between pigsties calls for a heterogeneity measure in terms of odds ratios.

The research on LRRE has focused on estimation of the parameters since this has proven to be difficult due to the fact that the likelihood function involves multiple integrals, which cannot be derived explicitly (Zeger and Karim, 1991; Breslow and Clayton, 1993; Hedeker and Gibbons, 1996). It is well known that fixed effects parameters do not maintain their interpretation when random effects are introduced in the model. Zeger et al. (1988) derived an approximative relation between what they called the population-averaged parameters (from the GEE) and the subject-specific parameters (with random effects in the linear predictor). The interpretation of random effects parameters has, to our knowledge, not been discussed, and the scope of this paper is to discuss the possible

Table 1
The pigsty data

SPF								
0/16	0/15	0/15	0/15	0/15	0/15	0/15	1/15	2/15
0/14	0/14	0/12	0/10	0/10	0/10	0/10	0/10	0/10
0/10	0/10	1/10	1/10	2/10	2/10	0/9	1/7	1/7
0/5	0/5	0/5	0/5	1/5	1/5	4/5	1/3	0/1
Conventional								
0/15	0/15	0/15	0/15	0/15	1/15	1/15	1/15	1/15
2/15	3/15	3/15	3/15	4/15	4/15	6/15	2/11	0/10
0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/10
0/10	1/10	1/10	1/10	1/10	3/10	3/10	4/10	0/9
0/9	0/9	0/9	1/9	2/9	5/9	9/9	0/8	0/8
2/8	3/8	0/7	1/7	1/7	1/7	1/7	2/7	3/7
0/6	0/6	0/6	0/5	0/5	0/5	1/5	1/5	1/5
1/5	0/4	0/4	3/4	0/3	0/2	0/2	0/1	1/1

interpretations of both the fixed effects parameters and the random effects parameters.

The paper is organized as follows. In Section 2, we present the model and odds ratio interpretations are discussed using the pigsty example as motivation. We emphasize that the odds ratio is a random variable, and we propose median-based interpretations for both fixed effects and random effects. The fixed effects are supplied with intervals reflecting the heterogeneity in the model. To clarify the ideas and concepts, we return to the pigsty example. More subtle points of the median-based interpretation are illustrated in the following two examples (Sections 3 and 4). The salamander mating experiment has crossed random effects, and the morphine prescription study has a nested structure on the random effects.

2. Logistic Regression with Random Effects

Let $(Y_i)_{i=1,\ldots,N}$ be Bernoulli variables with probability distribution

$$P(Y_i = 1|\eta_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}, \qquad \eta_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}, \quad (1)$$

where $\boldsymbol{\beta}$ is the fixed effects parameters, $\mathbf{x}_i$ is the ith row of the $N \times p$ design matrix $\mathbf{X}$ for the fixed effects, $\mathbf{u}$ contains the random effects, which are normally distributed with mean $\mathbf{0}$ and variance matrix $\boldsymbol{\Sigma}$, and $\mathbf{z}_i$ is the ith row of the $N \times r$ design matrix $\mathbf{Z}$ for the random effects. One can think of $\mathbf{u}$ as unmeasured covariates, as a way to model heterogeneity, or as a way to model correlated data.

In the LRRE, conditioning on the random effects yields the same nice interpretation in terms of odds ratios as is the case for the ordinary logistic regression model. However, in practice, it is not possible to condition on the random effects because these are unobservable. The log odds ratio between the responses of two measurements, i_1 and i_2, is $\eta_{i_1} - \eta_{i_2} = (\mathbf{x}_{i_1} - \mathbf{x}_{i_2})\boldsymbol{\beta} + (\mathbf{z}_{i_1} - \mathbf{z}_{i_2})\mathbf{u}$, and the odds ratio takes the form

$$\text{OR} = \exp(\kappa + \omega \cdot v), \quad (2)$$

where $\kappa = (\mathbf{x}_{i_1} - \mathbf{x}_{i_2})\boldsymbol{\beta}$, $\omega^2 = (\mathbf{z}_{i_1} - \mathbf{z}_{i_2})\boldsymbol{\Sigma}(\mathbf{z}_{i_1} - \mathbf{z}_{i_2})^{\mathrm{T}}$, and v is normally distributed with mean zero and variance one. So the odds ratio is a random variable rather than a fixed parameter, and this should be reflected when interpreting the model.

Pigsty example, continued. To take the heterogeneity of pigsties into account, we use the LRRE. The model is given by (1), where $\mathbf{X}$ is a 1016×2 matrix with elements $x_{ij} = 1_{(i\text{th pig is in a pigsty of type } j)}$, $\mathbf{Z}$ is a 1016×108 matrix with elements $z_{ik} = 1_{(i\text{th pig is in the } k\text{th pigsty})}$, and $\boldsymbol{\Sigma}$ is a 108×108 diagonal matrix with elements σ^2. The parameters are estimated using the SAS macro GLIMMIX (SAS Institute, 1996), yielding $\hat{\beta}_{\text{SPF}} = -3.03$, $\hat{\beta}_{\text{conv}} = -2.14$, and $\hat{\sigma}^2 = 1.38$.

The estimated β-contrast has a pigsty-specific interpretation: The odds for having *Ascaris suum* for a pig in a particular SPF pigsty would have been $\exp(\hat{\beta}_{\text{conv}} - \hat{\beta}_{\text{SPF}}) = 2.44$ times higher had the pigsty been of the conventional type. Or, in a conventional pigsty, the odds for having *Ascaris suum* would drop $\{1 - \exp(\hat{\beta}_{\text{SPF}} - \hat{\beta}_{\text{conv}})\} \times 100\% = 59\%$ for any pig in the pigsty if changing to the SPF type. To obtain population-averaged estimates, an approximate formula has been suggested by Zeger et al. (1988), yielding an estimated odds ratio of 2.08. This is the estimated odds ratio between a pig from a randomly chosen conventional and a pig from a randomly chosen SPF pigsty. Both approaches give relevant interpretations. However, neither of them use the heterogeneity of pigsties explicitly when interpreting the effects. In this example, the size of the variance component is relevant, partly as a measure of the heterogeneity of pigsties within type and partly because it determines the scale on which the fixed effects should be judged.

2.1 *Reporting Random Effects*

Usually the random effects parameters are reported "as is," i.e., on the linear scale. Since one of the appealing properties of the logistic regression model is the odds ratio interpretation of the fixed effects, the question arises whether one may report heterogeneity in terms of odds.

As seen, in LRRE, the odds ratios take the form of (2). When quantifying the random effects, we disregard the differences due to fixed effects by considering subjects with the same covariate values for the fixed effects. This corresponds to $\kappa = 0$ in (2). The odds ratio is an unobserved random variable for which we can report distributional characteristics. Here, we discuss mean- and percentile-based measures.

The expected odds ratio, $\mathrm{E}[\exp(\omega \cdot v)]$, does not provide a good quantification of the heterogeneity since it does not measure the variation of the odds ratio. The variance, $V\{\exp(\omega \cdot v)\}$, is also unsatisfactory since the distribution of the odds ratio is skewed. Instead, we consider $\exp(|\omega \cdot v|)$, which can be interpreted as the odds ratio between the unit at highest risk and the unit at lowest risk when randomly picking out two units. However, the expected odds ratio suffers from a drawback, namely that $\mathrm{E}[\exp(-|\omega \cdot v|)] \neq \{\mathrm{E}[\exp(|\omega \cdot v|)]\}^{-1}$, and the alternative $\exp(\mathrm{E}|\omega \cdot v|)$ has an awkward interpretation.

Let $\mathrm{med}(V)$ denote the median of a random variable V. We propose using the median odds ratio between the unit at higher risk and the unit at lower risk, i.e.,

$$\text{MOR} = \mathrm{med}\{\exp(|\omega \cdot v|)\}.$$

We have that $\mathrm{med}\{\exp(|\omega \cdot v|)\} = \exp\{\omega \cdot \Phi^{-1}(3/4)\}$, and it holds that $\mathrm{med}\{\exp(-|\omega \cdot v|)\} = [\mathrm{med}\{\exp(|\omega \cdot v|)\}]^{-1}$ and that $\mathrm{med}\{\exp(|\omega \cdot v|)\} = \exp\{\mathrm{med}(|\omega \cdot v|)\}$. The interpretations can easily be extended to percentiles.

Pigsty example, continued. Consider two pigs from two randomly chosen different pigsties of the same type. The odds ratio between the pig at higher risk and the pig at lower risk has the form $\exp(|(2\sigma^2)^{1/2} \cdot v|)$, where v is normally distributed with mean zero and variance one. Note that it is not necessary to know which of the pigs has the higher risk. The estimated median odds ratio between the pig at higher and the pig at lower risk is $\widehat{\text{MOR}} = \text{med}\{\exp[|(2 \cdot 1.38)^{1/2} \cdot v|]\} = 3.06$, which indicates substantial heterogeneity between pigsties of identical type.

This quantification enjoys the property of being directly comparable to the fixed effect (interpreted as a subject-specific effect, say), and also in that perspective, the heterogeneity is of a substantial size.

2.2 *Reporting Fixed Effects*

When introducing random effects in the linear predictor in a logistic regression model, the interpretation of the fixed effects usually changes for one or more of the fixed effects. The odds ratio between two levels (or values) for a given variable takes the form of (2).

Mean-based measures must be disregarded for similar reasons as in Section (2.1), and we suggest the median odds ratio (MOR) between two randomly chosen units (each with a given pattern on the explanatory variables) as a measure of a fixed effect,

$$\text{MOR} = \text{med}\{\exp(\kappa + \omega \cdot v)\}.$$

Note that $\text{med}\{\exp(\kappa + \omega \cdot v)\} = \exp(\kappa)$ and that $\exp(\kappa)$ is the subject-specific effect, i.e., the effect of changing fixed effects covariates for a specific measurement. This interpretation does not always make sense, e.g., "What is the effect of changing sex?" as opposed to "What is the median odds ratio between women and men?" However, the median-based interpretation is always valid.

Let $\text{perc}_a(V)$ denote the $100 \times a$-percentile of a random variable V. The percentiles of the distribution of $\exp(\kappa + \omega \cdot v)$ characterize both the size and the variation of the subject-specific parameters. Therefore, we suggest reporting an interval around $\text{med}\{\exp(\kappa + \omega \cdot v)\} = \exp(\kappa)$, referred to as the interval odds ratio (IOR), reflecting the variation of the odds ratio due to the random effects in the linear predictor,

$$\text{IOR} = \Big[\text{perc}_{(1-a)/2}\{\exp(\kappa + \omega \cdot v)\};\ \text{perc}_{(1+a)/2}\{\exp(\kappa + \omega \cdot v)\}\Big].$$

The interpretation is that the odds ratio between two randomly picked measurements is contained in the interval with probability a. So the heterogeneity yields an interval as a measure of fixed effects rather than just an odds ratio. In a given study, one could choose to report, say, an 80% interval and the median. Note that the IOR has nothing to do with a confidence interval; it is merely an illustration of the joint point estimates of fixed and random parameters.

Pigsty example, continued. Consider two pigs from two randomly chosen pigsties of different type. As already noted, the estimated median odds ratio between the pig from the conventional and the pig from the SPF pigsties is 2.44 and the estimated 80% interval odds ratio is [0.291; 20.5]. The interpretation is that, when comparing two pigs from two randomly picked pigsties, the odds ratio will, with 80% probability, lie within the range of the interval and, with 50% probability, it will be larger than 2.44.

Apparently, even though the median odds ratio is 2.44, there still is a chance of 10% for the odds ratio being larger than 20.5 and a chance of 10% for the odds ratio being less than 0.291. An interval this broad suggests that, if one was to improve hygiene in the pigsties as a whole, the search for other risk factors could turn out to be more fruitful.

Because there are several pigs within pigsties, it is possible to check the model by comparing the empirical distribution of the odds ratios between pigsties with the estimated theoretical distribution. This check confirms the model.

3. Salamander Mating Experiment

To illustrate how the points raised in the previous section apply to a more complex setting, in which two crossed random effects are present, we study data from a salamander mating experiment.

The salamander mating experiment (McCullagh and Nelder, 1989) was set up to study whether two populations of the same species of salamanders separated in their natural habitats can interbreed. The experiment involved two salamander populations—rough butts (R) and whitesides (W). The study outcome was whether or not mating occurred for each of the pairs of salamanders that were brought together for mating. Each salamander took part in six matings—three with its own species and three with the other. The data, being unusual in that the random effects are crossed, have been analyzed extensively (see Schall, 1991; Karim and Zeger, 1992; Breslow and Clayton, 1993; Kuk, 1995).

Let Y_{ij} indicate the result of the mating between the ith female and the jth male (success coded as 1, failure as 0). The linear predictor is then $\eta_{ij} = \beta_{\text{species}(i,j)} + u_i^{\text{F}} + u_j^{\text{M}}$, where $\beta_{\text{RR}}, \beta_{\text{RW}}, \beta_{\text{WR}}$, and β_{WW} are the fixed effects parameters, i.e., β_{RW} describes the effect of matings between a female rough butt and a male whiteside. The random effects u_i^{F} and u_j^{M}, being independent normally distributed with mean zero and variances σ^2_{F} and σ^2_{M}, respectively, are specific to each female and male salamander, respectively. The parameters σ^2_{F} and σ^2_{M} are sex-specific parameters, allowing different heterogeneity among the two sexes. Schall (1991) provides the estimates $\hat{\beta}_{\text{RR}} = 1.16$, $\hat{\beta}_{\text{RW}} = 0.78$, $\hat{\beta}_{\text{WR}} = -1.41$, $\hat{\beta}_{\text{WW}} = 1.02$, $\hat{\sigma}^2_{\text{F}} = 1.41$, and $\hat{\sigma}^2_{\text{M}} = 0.09$.

Different interpretations of contrasts between regression parameters are available, i.e., as either a comparison of the effect on a pair of salamanders, when changing the species on one or both of the animals, or as the difference in effect between two types of salamanders with the same value of their random effects, e.g., two average (both with random effects zero) salamanders. The two different interpretations yield exactly the same numerical result, i.e., the species-specific effect of salamander type. However, a more satisfactory interpretation of the fixed effects parameters of the model is as medians in the distributions of the odds of successful mating between many pairs of salamanders. Consider the experiment in which two rough butts, say, were mated repeatedly. From

these repeated matings, the odds of success could be determined. Mating two different rough butt salamanders would result in different odds of successfully achieving mating. Matings between many pairs of salamanders would yield a distribution of odds, of which the median is $\exp(\beta_{RR})$.

Three types of heterogeneity are present—between males when mated with a specific female, between females when mated with a specific male, and between matings involving two different males and two different females. The distribution of the ratio of the odds of a mating involving a female (i) and a male (j) and the odds of a mating involving the same female and another male (j') is $\exp\{\beta_{\text{species}(i,j)} - \beta_{\text{species}(i,j')} + (u_j^M - u_{j'}^M)\}$. Choosing a whiteside female and mating her with first a whiteside male and then a rough butt male yields $\widehat{\text{MOR}}_{\text{WW-WR}} = 11.4$ and $\widehat{\text{IOR}}_{\text{WW-WR}} = [6.59; 19.6]$. So 80% of this type of mating is expected to yield an odds of successfully mating with a whiteside male a factor 6.59 to 19.6 higher compared with that of mating with a rough butt. Whiteside females definitely prefer whiteside males!

Such IOR intervals are model-based but, in principle (as mentioned above), the distribution of relevant odd ratios could be validated. This was possible in the pigsty example, whereas in the salamander experiment, such validating would require repeated matings between the same salamanders.

The heterogeneity measure is interpreted as follows. From two randomly chosen males of the same species, the median ratio of the odds between the salamander with the highest propensity to yield a successful mating with some female salamander and the salamander with the lowest propensity to yield a successful mating with the same female is estimated to be $\widehat{\text{MOR}} = 1.33$. This is a low heterogeneity between males. Similarly, when mating the same male with two different females, one finds $\widehat{\text{MOR}} = 3.10$, and when mating two different pairs of salamanders, assuming that it is the same combination of salamander species in the two pairs, $\widehat{\text{MOR}} = 3.21$.

4. Morphine Prescription Study

A model with two sources of heterogeneity is considered but, contrary to the previous salamander experiment, the random effects are nested (a random main effect and a random interaction) and the structure of the fixed effects is more complex.

In 1993, Denmark had the highest consumption of analgesic drugs in the world. To avoid abuse, changes were made in the emergency service, and the Danish National Health Service sent out instructions on the prescribing of morphine. The purpose of this study is to investigate whether these actions had any effect on the tendency to prescribe morphine among the medical doctors in the emergency services of Copenhagen (see Andersen, 1998).

Data consist of records on 2153 visits made by 74 randomly chosen doctors from the emergency services of Copenhagen. Of these visits, 980 were made in 1993 and 1173 in 1996. Only a few of the doctors from the 1993 investigation reappeared in the 1996 investigation.

The response Y_i of the ith visit is whether morphine was prescribed on the visit (coded as 1) or not (coded as 0). The statistical analysis lead to a model with linear predictor $\eta_i = \beta_{\text{year}(i)} + \beta_{\text{time}(i)} + \beta_{\text{age}(i)} + \beta_{\text{cause}(i)} + u_{\text{doc}(i)} + u_{\text{doc}(i),\text{cause}(i)}$, where $u_1, \ldots, u_{74}$ and $u_{1,1}, \ldots, u_{74,6}$ are mutually independent normally distributed with means zero and variances σ^2_{doc} and $\sigma^2_{\text{doc}\times\text{cause}}$, respectively. The fixed effects are year (1993 or 1996), time (day or night visit), age of the patient (five levels), and cause of action (six levels). The variation not explained by the fixed effects stems from two sources. First, we have variation due to the fact that some doctors have a higher (overall) tendency to prescribe morphine than others (σ^2_{doc}). Second, we have variation due to the fact that different doctors rank the causes differently with regard to the necessity of morphine prescription ($\sigma^2_{\text{doc}\times\text{cause}}$).

Estimation is carried out using GLIMMIX (SAS Institute, 1996). The estimated fixed effects parameters are shown in Table 2. The estimated random effects parameters are $\hat{\sigma}^2_{\text{doc}} = 0.11$ and $\hat{\sigma}^2_{\text{doc}\times\text{cause}} = 0.33$. Due to the random interaction between cause and doctor, only a limited validation of the random part of the model may be performed, yielding no conclusive results.

Table 2 shows that the tendency to prescribe morphine decreases significantly from 1993 to 1996. If two doctors are picked randomly, one in 1993 and one in 1996, and then sent on the same type of visit, then the estimated median odds ratio for prescription in 1993 against 1996 is $\widehat{\text{MOR}}_{93-96} = 1.63$. This is also the doctor-specific estimate under the assumption that all doctors undergo the same change in tendency to prescribe morphine from 1993 to 1996. However, the heterogeneity between doctors is substantial, which is reflected by $\widehat{\text{IOR}}_{93-96} = [0.49; 5.43]$.

The effects of the variables time and age have straightforward interpretations as ordinary odds ratios conditioning out the random terms in the linear predictor.

When the fixed effects parameters of cause are interpreted, it is important to note that $\text{MOR}_{\text{cause1},\text{cause}_2}$ are not doctor-specific odds ratios. This is a consequence of the random interaction between doctor and cause. Looking at the contrast between back pain and other, the median odds ratio is 6.90. If the two causes are treated by the same randomly chosen

Table 2
Estimates from the morphine prescription study

Variable	p-Value	Estimated contrast
Year	0.017	
1993 vs. 1996		0.49
Age	<0.0001	
0–15 vs. ≥70		−3.03
16–29 vs. ≥70		−0.76
30–39 vs. ≥70		−0.42
40–69 vs. ≥70		0.36
Time	<0.0001	
Night vs. day		0.89
Cause	<0.0001	
Stomach pain vs. other		1.46
Back pain vs. other		1.93
Heart pain vs. other		2.58
General pain vs. other		2.94
Migraine vs. other		3.37

doctor, then $\widehat{\text{IOR}}_{\text{back pain–other}} = [2.45; 19.4]$. However, if the causes are treated by different doctors, then the 80% interval becomes wider, reflecting the difference in the overall tendency to prescribe morphine between the doctors, $\widehat{\text{IOR}}_{\text{back pain–other}} = [2.08; 22.9]$.

In this model, it is not straightforward to interpret $\hat{\sigma}^2_{\text{doc}\times\text{cause}}$ alone through MOR. If the same doctor is sent on two identical visits except that the causes are different, then the odds ratio for prescription is $\exp\{\beta_{\text{cause}_1} - \beta_{\text{cause}_2} + (u_{\text{doc}_1,\text{cause}_1} - u_{\text{doc}_1,\text{cause}_2})\}$. If we consider two causes that are (hypothetically) considered to be equally serious in the population of all doctors, i.e., $\beta_{\text{cause}_1} = \beta_{\text{cause}_2}$, then the odds ratio has the form of $\exp(u_{\text{doc}_1,\text{cause}_1} - u_{\text{doc}_1,\text{cause}_2})$. Under these circumstances, the estimated median odds ratio between the visit with higher risk and the visit with lower risk is $\widehat{\text{MOR}} = 1.72$. This means that, even though the causes are considered equally serious in the population of all doctors, there is a 50% chance that a randomly chosen doctor will consider one of the causes to be at least 1.72 times more serious than the other.

If two doctors are sent on identical visits, then the estimated median odds ratio between the visit with higher risk and the visit with lower risk is $\widehat{\text{MOR}} = 1.88$, a moderate increase from 1.72, reflecting that the heterogeneity of the doctors primarily is caused by different perceptions of which causes should be treated with morphine.

Above, a significant decrease from 1993 to 1996 in the tendency to prescribe morphine has been quantified. However, the heterogeneity between doctors is so large that, if a patient is visited by two doctors right after each other, the ratio of the odds exceeds the doctor-specific odds ratio between 1993 and 1996 with a probability of more than 50%.

5. Discussion

Although the main purpose of a study may be to estimate and draw inference about the effect of some treatment on a disease, it is usually equally important to understand and describe the ramifications of the different sources of heterogeneity and associations between outcomes. Our focus in this paper is a logistic regression model in which these sources of heterogeneity and association are described by introducing random variables. This has implications for the interpretation of possible fixed effects, but it also opens up a more detailed description of sources and magnitudes of variation.

The interpretation of fixed effects parameters in the presence of random effects is not always straightforward. As an alternative to the subject-specific interpretation of fixed effects parameters, we prefer interpreting these parameters as the median in the distribution of the odds ratio, an interpretation that is always valid, which is not the case for the subject-specific interpretation.

The distribution of the odds ratio adds to the understanding of the clinical relevance when evaluating the effect of some treatment because, in an unambiguous way, it spells out the implications of the heterogeneity between patients on the treatment effect. This basically amounts to what any statistically trained analyst does to get an impression of the clinical relevance of some effect when informally comparing the log-odds ratio estimate with the estimated random effects parameters. In the presence of more random effects parameters, a more formalized set-up, as presented in this paper, is needed for quantifying the variation and heterogeneity.

Different measures of heterogeneity have been considered, the aim being to find a function of the relevant random effects parameters that has a nice interpretation in terms of an odds ratio. Our recommendation is to use the median in the distribution of the odds ratio of the subject of the higher risk to that of the lower risk. This measure, which is based purely on the random effects parameters in the model, quantifies the variation between subjects and, when communicating the results, we can draw on the well-established understanding of ordinary odds ratios and relative risks.

Here we have only considered point estimates, in the sense that the distribution of the estimated parameters is not taken into account. To obtain confidence intervals for these estimates, we need the joint distribution of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$, which is usually not easy to obtain.

We have focused attention on the logistic regression model with normally distributed random effects entering additively in the linear predictor. Basically, the same arguments could be carried out for nonnormal generalized linear mixed models. Consider, e.g., log-linear models where the linear predictor, η_i, is the logarithm of the mean response, i.e., $\eta_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}$. Here one finds that the natural heterogeneity measure corresponding to MOR is interpreted as the median ratio of the higher expected response to the lower expected response. Similarly, fixed effects are given a median interpretation. The important point to note in connection with generalized linear models with possibly nonnormal random effects is that the effect of a covariate and the effect of heterogeneity are often not best understood on the linear predictor scale.

Acknowledgements

This work has been supported by the Danish Medical Research Council, the Danish Work Environment Fund, the Danish Research Academy, and Sygekassernes Helsefond. The authors wish to thank Allan Roepstorff and John Sahl Andersen for letting us use their data and for helpful discussions on the subject matter of their research. The authors are grateful to Niels Keiding and Lene Theil Skovgaard for helpful comments and suggestions.

Résumé

La méthode de régression logistique avec effets aléatoires permet d'étudier la relation entre une variable binaire et un ensemble de covariables lorsque les observations de la variable binaire ne sont pas indépendantes. Dans cet article, nous examinons en détail l'interprétation simultanée des effets fixes et des effets aléatoires. Comme mesures de l'hétérogénéité, les paramètres représentent les effets aléatoires dans les modèles ne sont pas faciles à interpréter. Nous envisageons différentes mesures de l'hétérogénéité et proposons celle de l'odds ratio médian, qui est une fonction des paramètres originaux des effets aléatoires. Cette mesure présente une interprétation simple—en termes d'odds ratio—ce qui facilite grandement le dialogue entre le biostatisticien et le chercheur confronté à de telles données. Trois exemples prélevés dans des domaines distincts et relevant de notre expérience personnelle permettent de mieux comprendre et d'illustrer les différents aspects de l'interprétation des paramètres dans ces modèles.

References

Andersen, J. (1998). Copenhagen out of hours service—Activity and quality. Ph.D. dissertation, University of Copenhagen, Copenhagen, Denmark.

Ashford, J. R. and Sowden, R. R. (1970). Multivariate probit analysis. *Biometrics* **26,** 535–546.

Bishop, Y., Fienberg, S., and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Massachusetts: MIT Press.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Hedeker, D. and Gibbons, R. (1996). Mixor: A computer program for mixed-effects ordinal analysis. *Computer Methods and Programs in Biomedicine* **49,** 157–174.

Karim, M. R. and Zeger, S. L. (1992). Generalized linear models with random effects: Salamander matings revisited. *Biometrics* **48,** 631–644.

Kuk, A. Y. C. (1995). Asymptotically unbiased estimation in generalized linear mixed models with random effects. *Journal of the Royal Statistical Society, Series B* **57,** 395–407.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society, Series B* **54,** 3–40.

Lindsey, J. K. and Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* **17,** 447–469.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman & Hall.

Roepstorff, A., et al. (1998). Intestinal parasites in swine in the Nordic countries: Prevalence and geographical distribution. *Veterinary Parasitology* **76,** 305–319.

SAS Institute. (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.11.* Cary, North Carolina: SAS Institute.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78,** 719–727.

Stiratelli, R., Laird, N., and Ware, J. (1984). Random effects statistical model for serial observations with binary response. *Biometrics* **40,** 961–971.

Williams, D. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31,** 949–952.

Zeger, S. L. and Karim, M. (1991). Generalized linear mixed models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86,** 79–86.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44,** 1049–1060.

Received December 1998. *Revised November* 1999. *Accepted December* 1999.