Contextual multilevel models: Effects and correlations at multiple levels

Stryhn, H.[1], De Vliegher, S.[2] and Barkema, H.W.[1]
[1] Atlantic Veterinary College, University of Prince Edward Island, Canada
[2] Faculty of Veterinary Medicine, Ghent University, Belgium

Introduction

Multilevel models are routinely used in veterinary epidemiology, but most commonly in their simplest form with random effects corresponding to each hierarchical level (above the lowest), often termed random intercepts models. Such models adjust a regression analysis for predictors (independent variables) of interest for any clustering present in the outcome (dependent variable). However, they may provide simplistic, and possibly misleading, answers to questions concerning the effect of predictors that themselves show variation at several hierarchical levels. The notion of a ”contextual” effect of a predictor originates from social sciences and captures the idea that although the predictor is recorded at an individual level, its effect mostly (or entirely) relates to the group, or context, to which the individual belongs (Snijders and Bosker, 1999). The objective of this paper is to discuss and illustrate the importance of contextual effects in veterinary epidemiology by way of a worked example involving milk samples from heifers in dairy herds. Our discussion outlines the procedures to detect and account for contextual effects in the statistical analysis.

Materials and Methods

Contextual modelling

We consider the two-level hierarchical structure of animals clustered in herds, and continuous outcomes Y_{ij} recorded for the *i*'th animal in the *j*'th herd together with values x_{ij} of a (continuous or dichotomous) predictor *x*. Our discussion will be in terms of the simplest 2-level mixed model,

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + \varepsilon_{ij} \qquad (1)$$

with herd random effects $u_j \sim N(0,\sigma_h^2)$ and animal errors terms $\varepsilon_{ij} \sim N(0,\sigma^2)$. However, the concepts and procedures generalise to structures with more hierarchical levels, and to non-normal and multivariable models. In this model, the predictor *x* is said to have a contextual effect if the following two conditions are both satisfied:

i. *x* varies both between and within herds,
ii. the between-herd and within-herd regressions of *Y* on *x* have different slopes.

Two situations where condition i. is *not* satisfied are: when *x* is a herd-level predictor (i.e., *x* is constant within each herd), and when the herd averages $\bar{x}_{\cdot j}$ are constant between herds (e.g., in a clinical field trial with treatment groups equally represented within each herd). For condition ii., the within-herd regression of *Y* on *x* refers to a regression equation corresponding to different animals within a single herd. Furthermore, the between-herd regression refers to a regression of herd mean outcomes $\bar{Y}_{\cdot j}$ on herd predictor means $\bar{x}_{\cdot j}$. Figure 1 (below) illustrates situations where the between-herd and within-herd regressions of *Y* on a continuous predictor *x* coincide (left-hand panel) and are completely different (right-hand panel). The within-herd regressions are indicated by straight lines (without showing individual data points), and the between herd regression is obtained by fitting a straight line to the dotted points (herd means of *Y* and *x*). In the right-hand panel, the within-herd slope is positive whereas the between-herd regression would have a negative slope.

It is important to realise the presence of contextual effects for a problem, because the two different regressions represent different effects, and therefore often have different interpretations. The single slope (β_1) in model (1) is a complex function of the two slopes and difficult to interpret (see Snijders and Bosker, 1999, Section 3.6, for a detailed discussion). Failure to account for contextual effects may lead to conclusions based on either ecological or atomistic fallacies (see e.g., Dohoo et

Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006
Available at www.sciquest.org.nz

al. (2003), Chapter 25). Snijders and Bosker (1999, p.56) note that ”for theoretically important variables in multilevel studies [the presence of contextual effects] is the rule rather the exception”.

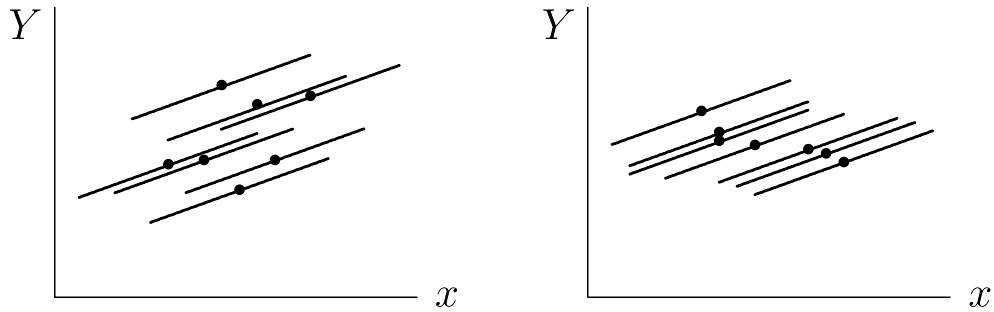

Figure 1 Schematic graphs showing no contextual effect (left) and a strong contextual effect (right).

Statistical procedures for contextual effects

In order to detect a contextual effect of the predictor *x* we add the herd means $\overline{x}_{.j}$ to model (1) as an additional predictor (while retaining the predictor *x*), that is,

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 \overline{x}_{.j} + u_j + \varepsilon_{ij} \tag{2}$$

A contextual effect is (significantly) present when the estimate of the regression coefficient β_2 is statistically significant. If a contextual effect is present, we recommend (in order to reduce collinearity and to obtain more easily interpretable estimates) to reformulate model (2) by replacing the original predictor *x* by its within-herd centered version, $z_{ij} = x_{ij} - \overline{x}_{.j}$, as follows:

$$Y_{ij} = \beta_0 + \beta_1 z_{ij} + \beta_2^* \overline{x}_{.j} + u_j + \varepsilon_{ij} \tag{2*}$$

The equations (2) and (2*) represent the *same* model, and the coefficients for *x* and *z* are identical (β_1), whereas $\beta_2^* = \beta_1 + \beta_2$. The parameter β_2^* is the slope of the between-herd regression of *Y* on *x*, and the parameter β_1 in models (2) or (2*) is the slope of the within-herd regression of *Y* on *x*. The models (2) and (2*) may be fitted using standard software for mixed models.

Somatic cell count data example

De Vliegher et al. (2004, 2005) presented data on milk yield and somatic cell counts (SCCs) of heifers in Belgian herds collected during the years 2000-2001. The heifers were followed by approximately monthly recordings throughout the first lactation (until dry-off), but for simplicity we consider only a single record for each heifer obtained during 76-105 days in milk; this subset of the data comprises 10,996 heifers in 3095 herds. The predictor of primary interest was the (natural) log SCC in early lactation (days in milk 5-14; *lnsccel*), and our focus here is on predicting log SCC later in lactation. Therefore, our outcome was the measured log SCC (*lnscc*) within the time frame 76-105 days. Somatic cell counts were measured in 1000 cells/mL milk. Due to the expected close relationship between SCC and the actual milk yield, *yield* (kg milk) was included as an additional predictor in the model. For simplicity, additional predictors of the analysis of De Vliegher et al. (2004) were omitted here: the season, the breed, and the days in milk at which the early lactation SCC was obtained, as well as an interaction between the latter and *lnsccel*. The conclusions and interpretations from this simplified model were qualitatively the same as from the full model. All analyses assumed a linear mixed model for *lnscc* with herd random effects, and were carried out using SAS software (proc mixed, version 9.1).

Results and Discussion

Table 1 displays parameter estimates for three models: with *lnsccel* as a sole predictor, with an additional effect of *yield*, and finally a contextual model with herd means of both predictors included as well. The herd-level regression coefficients (β_2^*) of the contextual model are added to the regression coefficients listed in the table (β_2).

Table 1 Parameter estimates with standard errors (SE) for three linear mixed models (with herd random effects) for log somatic cell counts in Belgian heifers recorded during 76-105 days in milk.

Parameter	Univariable model		Multivariable model		Contextual model			
	estimate	SE	estimate	SE	estimate		SE	
constant	4.095	.012	4.110	.012	4.097		.012	
β(lnsccel)	0.262	.008	0.259	.008	0.242	0.320	.009	.017
*β(*herd mean *lnsccel)*	--	--	--	--	0.078		.019	
β(yield)	--	--	- 0.014	.002	- 0.002	- 0.026	.003	.003
*β(*herd mean *yield)*	--	--	--	--	- 0.024		.004	
σ^2(herd)	0.121	.011	0.110	.011	0.102		.010	
σ^2(heifer)	1.037	.016	1.041	.016	1.042		.016	

The first two models show that both *lnsccel* and *yield* have strongly significant effects, but exhibit only little confounding. As both predictors are recorded at the heifer level, it is tempting to interpret their effects at the heifer level. For example, a one-unit difference (between heifers) in the early lactation log SCC seems to be associated with an average difference in actual log SCC of 0.259 units. Also, the negative regression coefficient for *yield* could be interpreted as a dilution effect.

There are contextual effects for *lnsccel* and *yield* because both herd mean predictors are clearly significant (P<0.001). For *lnsccel*, the results show significant regressions at the animal and herd levels. The animal-level slope is 0.242, and the herd-level slope is 0.32. The above interpretation of the implication of a one-unit difference in *lnsccel* between heifers in the same herd was therefore correct with a slight adjustment of the coefficient (from 0.259 to 0.242). In addition, herds with a high mean *lnsccel* also, and somewhat more strongly, tend to have a higher actual mean *lnscc*. This added effect may be interpreted as a reflection of management; herds with good management practices are likely to have low SCC levels both early and later in lactation. For the effect of *yield*, a different picture is seen because in the contextual model the animal-level predictor is non-significant. The proper interpretation of the effect of *yield* is therefore at the herd level and most likely again a reflection of management practices whereby herds with low SCCs also tend to have high milk yields, or vice versa. When the herd-level association has been accounted for there is no (or very little) correlation between log SCCs and milk yield of heifers. The previously indicated interpretation as a dilution effect at the animal (or sample) level was therefore false, in effect an ecological fallacy.

References

De Vliegher, S., Barkema, H.W., Stryhn, H., Opsomer, G. and de Kruif, A. (2004). Impact of early lactation somatic cell count in heifers on somatic cell counts over the first lactation. *Journal of Dairy Science*, 87, 3672-3682.

De Vliegher, S., Barkema, H.W., Stryhn, H., Opsomer, G. and de Kruif, A. (2005). Impact of early lactation somatic cell count in heifers on milk yield over the first lactation. *Journal of Dairy Science*, 88, 938-947.

Dohoo, I.R., Martin, S.W. and Stryhn, H. (2003). *Veterinary Epidemiologic Research*. AVC-Inc., Charlottetown, Canada.

Snijders, T.A.B. and Bosker, R.J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications, London.