MODELLING ANIMAL SYSTEMS PAPER

Effect of a pen in group-feeding trials: modelling data from suckler cows

L. JAUHIAINEN[1]*, M. MANNINEN[2] AND J. ÖFVERSTEN[1]

[1] *MTT Agrifood Research Finland, Services Unit, FI-31600 Jokioinen, Finland*
[2] *MTT Agrifood Research Finland, Animal Production Research, FI-31600 Jokioinen, Finland*

(*Revised MS received 7 March 2008*)

SUMMARY

This paper provides information on the extent and containment of intraclass correlation (ICC) concerning group-feeding trials on suckler cows. The research comprised the re-analysis of six previously reported experiments and a parallel simulation study. ICC coefficients were estimated for eight variables. They were seen to be higher in experiments on bulls compared with those on cows and more notable in the measurements of live weight gain compared with final weight. Moreover, the ICC coefficients were generally high for all variables measuring behavioural patterns. The simulation study showed that using a single animal as an experimental unit could be valid in certain situations, but it could not be extended to cover all the cases. The simulation study also showed that the common mixed model approach had significant problems when the ICC was slight. In general, the research strengthened the arguments that much more effort should be placed on the planning and statistical analysis of group-feeding experiments, especially in behavioural studies.

INTRODUCTION

An experimental unit is the portion of experimental material to which a treatment is applied (Gomez & Gomez 1984). The experimental unit lays the foundation for the definition of experimental error, which is defined as variation among observations that are treated alike. Correctly defined and estimable experimental error is the key element of the statistical analysis of experimental data collected during the study.

In an animal study, an experimental unit may consist of a single animal or a group of animals housed in the same pen, paddock or pasture. If the experimental unit comprises only one animal, the experimental design and the forthcoming statistical analysis are usually straightforward. Yet, in many cases, natural housing practices would necessitate experimental units with several animals. This usually leads to a low number of experimental units and a poor estimate of the underlying experimental error. Splitting the experimental units after the

experimentation is usually ruled out by a conceivable correlation of individual responses of animals within experimental units.

Lucas (1948) was among the first to discuss the problem of experimental units in experiments where animals were housed in pens. He defined group feeding as the practice of placing two or more animals in a single pen and feeding them from one trough. He concluded that the real problem was the size of the groups and the number of groups per treatment. Moreover, since correlated observations contain less information than uncorrelated ones and since the responses of individual animals within the same pen often correlate, he considered it wise to keep the size of a pen small and increase the number of pens. A review of papers published over the period 2004–07 showed that this recommendation has been well followed, especially in experiments on lambs, pigs and chickens (e.g. Swamy *et al.* 2004; Kiker *et al.* 2005).

During the last few decades, discussion on the practical problems associated with the correlation of individual responses of animals within pens and the low number of pens per treatment has persisted in the literature and several recommendations have been

* To whom all correspondence should be addressed. Email: lauri.jauhiainen@mtt.fi

given. For example Robinson *et al.* (2006) stated that, if animals were group fed, feed intake should be analysed using a pen as an experimental unit but, if the animals were weighed separately, weight gain could be analysed with a single animal as an experimental unit. A similar method for the analysis of blood parameters was also proposed.

From a statistical point of view it is clear that, if the individual responses of animals are correlated within pens, the pen should be seen as an experimental unit. On the other hand, if the individual responses are not, or only slightly, correlated within the pens, a single animal can be used as an experimental unit. In the statistical literature, the correlation of individual responses of subjects within experimental units is usually referred to as intraclass correlation (ICC). The key issue is to establish whether ICC is present or not, because this fact determines both the course of experimental design and statistical analysis. The key problem is that information on ICC is needed at a very early stage of designing a new experiment. To obtain such information on animal experiments, previously performed and reported experiments need to be examined.

Sometimes it is easy to observe the presence of ICC, e.g. measuring conflicts between animals in an experiment on two animals in the same pen (e.g. Korhonen *et al.* 2001). In other instances the presence of ICC can be blurred, e.g. if there are dozens of animals housed in the same pen and innate behaviour, type of feeding or nature of the pen are not the main focus. Researchers have met the same kind of challenge in other fields of research too. Many school-based smoking prevention studies employ designs in which schools or classrooms are assigned to different treatments while observations are made on individual students (Siddiqui *et al.* 1996). In such studies there may be several reasons why individual responses correlate within cluster-levels. In such studies careful heed of ICC has proved to be useful. ICC has also proved to be of use in sample-size calculations (Donner 1992).

In the current study, previously collected data were employed to obtain information on the presence and extent of ICC in experiments with suckler cows. Methods for estimating the size of ICC and its impact on the proper analysis of experimental data were considered. Methods were based on mixed models, techniques and simulation. The results should be of use in the design of new experiments on suckler cows. The methods used are easily extendable to other sectors of animal experimentation.

MATERIAL AND METHODS

Group feeding trials

Six experiments were re-analysed to examine the extent of ICC in studies with suckler cows. These experiments were originally carried out for comparing different housing conditions, feeding strategies or feeding frequencies. They were conducted at the Tohmajärvi Research Station in Eastern Finland (62°20′N, 30°13′E). Three experiments were carried out on cows and their calves, while the rest were conducted on bulls. The principal practice used in the experiments was to place eight animals in a pen of 74 m² made up of 53 m² of bedding area and 21 m² of passage. Detailed information on the animals, experimental designs, treatments and measurement techniques used is available in Manninen *et al.* (2004), Manninen *et al.* (2005) and Manninen *et al.* (2006).

The re-analysis included the following six categories of variables:

(1) weight gain and final weight of the cows,
(2) body condition score (cows only),
(3) blood samples: urea, Hb, long-chain fatty acid, total protein, albumin (one experiment with cows),
(4) quality of meat (pH, tenderness, taste, consistency; one experiment with bulls),
(5) behavioural data (different behavioural categories, one trial), and
(6) initial and final weight of the calves.

Descriptive statistics of the response variables are presented in Table 1.

The statistical model used in the re-analysis was:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \tag{1}$$

where y_{ijk} is the observed value of the response variable for the kth animal in the ith treatment in the jth pen; μ the overall mean; α_i the fixed effect of the ith treatment; β_{ij} and ε_{ijk} are uncorrelated random pen and animal effects with zero means and variances σ^2_{pen} and σ^2, respectively. All the analyses and the estimation of the variance components were carried out using the SAS/MIXED-procedure and REML estimation method (SAS 1999). The estimated variance components were further used for estimating the ICC coefficient ρ as:

$$\rho = \sigma^2_{\text{pen}} (\sigma^2_{\text{pen}} + \sigma^2) \tag{2}$$

Design effect is a function of the average pen size and the ICC coefficient (Kish 1965):

$$\text{design effect} = 1 + (n-1)\rho \tag{3}$$

where n is the average number of animals housed in the same pen and ρ is the ICC. It is of use in the calculation of the proper number of experimental units when planning new experiments (Donner 1992).

In one of the re-analysed experiments, the animals were assigned to pens according to their initial weight, i.e. the pens were blocked. As suggested by Morris (1998), an extra variance component for blocking was then added to the above statistical model and

Table 1. *Descriptive statistics of response variables in six experiments (mean ± s.d.). Experiments 1–3 were conducted on cows and their calves, experiments 4–6 were carried out on bulls*

	Experiment					
Variable	1	2	3	4	5	6
Body weight (kg)	789 ± 45·6	811 ± 80·8	760 ± 88·3	605 ± 9·5	618 ± 30·9	757 ± 89·1
Body weight gain (g/day)	0 ± 86	240 ± 120	240 ± 108	970 ± 96	1300 ± 204	460 ± 59
Body condition score	3·6 ± 0·48	3·8 ± 0·46	3·2 ± 0·22	NA	NA	NA
Calf's birth weight (kg)	40 ± 5·0	44 ± 5·1	44 ± 5·4	NA	NA	NA
Calf's target weight (kg)	240 ± 24·3	271 ± 31·1	256 ± 37·0	NA	NA	NA
Blood parameters						
- Urea	NA	NA	NA	2·9 ± 0·50	NA	NA
- Hb				0·30 ± 0·091		
- Long-chain fatty acid				0·69 ± 0·437		
- Total protein				76 ± 5·5		
- Albumin				40 ± 1·8		
Quality of meat						
- pH	NA	NA	NA	5·6 ± 0·08	NA	NA
- Tenderness				5·5 ± 0·62		
- Taste				4·7 ± 0·54		
- Consistency				8·5 ± 1·81		
Behaviour						
- Standing	41 ± 9·2	NA	NA	NA	NA	NA
- Outdoors	23 ± 11·6					
- Eating	19 ± 7·2					
- Resting	9 ± 5·0					

NA = not available.

estimation of the ICC coefficient and the design effect was otherwise performed as above.

Simulation study

The above analysis of six real experiments gave an insight into the range of possible ICCs appearing in real studies. As expected, the ICC seemed to be closely connected with the variable under consideration. It is obvious that ignoring a high ICC in experimental design has a notably deteriorating effect on the conclusions of any experimental study. A consequent simulation study therefore was focused on the effect of unheeded small ICCs. To screen out possible detrimental effects various model specifications and analysis techniques commonly used in the statistical analysis of data from animal experiments were used. Moreover, to obtain relevant practical results the analysis concentrated on the observed low end of the ICC coefficients encountered in the analysis of real data.

In the first simulation study, data for 10 000 simulated experiments were generated. Each experiment comprised a fictitious variable with an ICC coefficient of 0·004, which was the lowest value encountered with real data in the above studies, three different treatments with no real difference between them, two pens per treatment and eight animals per pen. The effect of each pen was randomly drawn from the normal distribution with a mean of 0 and variance 10^2. The effect of each animal was randomly drawn from the normal distribution with a mean of 0 and variance 500^2.

Each generated experiment was analysed separately using the following statistical model including the terms implied by the simulation process:

$$y_{ijk} = \mu + a_i + \beta_{ij} + \varepsilon_{ijk} \qquad (4)$$

where y_{ijk} is the generated observation, μ the overall mean, α_i the effect of the ith treatment (no difference), β_{ij} the effect of the jth pen within the ith treatment and ε_{ijk} the normally distributed residual error (= within-pen variation).

Depending on the distributional assumptions assigned to the terms of the above model specification, it was possible to carry out four different types of analysis. In the first analysis α_i and β_{ij} were assumed to be fixed effects. This analysis was performed using the SAS/GLM-procedure (SAS 1999). When treatments were compared using the F-test, β_{ij} was used as an error term. The rest of the analysis was performed using SAS/MIXED-procedure with the REML

Table 2. *Estimated ICC coefficients and design effects for eight different variables*

	Average ICC			Range of ICC		Design effect		
Variable	All animals	Bulls	Cows	Minimum	Maximum	All animals	Bulls	Cows
Body weight	0·0148	0·0296	0·0000	0·0000	0·0662	1·1034	1·2069	1·0000
Body weight gain	0·1042	0·2055	0·0029	0·0000	0·3682	1·7294	2·4386	1·0203
Body condition score	NA	NA	0·0012	0·0000	0·0031	NA	NA	1·0084
Calf's birth weight	NA	NA	0·0862	0·0231	0·2070	NA	NA	1·6033
Calf's final weight	NA	NA	0·0104	0·0000	0·0313	NA	NA	1·0730
Blood parameters	NA	0·0498	NA	0·0000	0·2276	NA	1·3488	NA
Quality of meat	NA	0·0005	NA	0·0000	0·0026	NA	1·0037	NA
Behaviour	NA	NA	0·1005	0·0642	0·1476	NA	NA	1·7035

NA = not available.

estimation method (SAS 1999) in which β_{ij} was assumed to be a random effect. The three additional analyses differed only in the calculation of the denominator degrees of freedom. The methods used were: (i) 'between/within' denominator degrees of freedom (i.e. no adjustments), (ii) Satterthwaite's method (Satterthwaite 1946) and (iii) Kenward–Roger method (Kenward & Roger 1997). The basic idea of the two latter adjustment methods is to correct the denominator degrees of freedom if the variance component of a random effect has been estimated to be non-positive. In the current simulation study, the ICC was chosen to be small and therefore negative or zero estimates for variance component of pens occurred frequently. Adjustment methods are described in more detail by Satterthwaite (1946) and Kenward & Roger (1997).

Data were generated so that all differences between the three treatments were caused by chance only. An *F*-test can be used to test the null hypothesis that differences are caused by chance. If the null hypothesis is true, *P*-value for the *F*-test is uniformly distributed between 0 and 1, i.e. all values between 0 and 1 are equally probable. Accordingly, the four different analyses mentioned above were compared using the distributions of the *P*-values obtained and then compared with the standard uniform distribution.

Eight additional simulation studies were carried out by varying the number of pens per treatment (two, four or eight) and the between-pen variance (10^2, 40^2 or 100^2). Each of them comprised 10 000 simulated experiments. The ICC coefficients used varied between 0·0004 and 0·0400.

RESULTS

Group-feeding trials

Considering the weight of an animal at the end of an experiment, the average estimated ICC coefficient was 0·0148. When experiments on bulls and cows were

examined separately, the results showed that intra-correlation coefficient was higher in experiments on bulls than those on cows. On average, the ICC was more notable in the analyses of live weight gain during the experiment than in analyses of a given target weight. Again, the ICC coefficients were higher in experiments on bulls than those on cows (Table 2).

The body condition scores and weights of calves were evaluated in experiments on cows only. As to the birth weight of the calves, in one experiment the estimated ICC coefficient was as high as 0·2070 while in other experiments it was less than 0·0300. The number of pens per treatment was small in all experiments and therefore the estimate for the variance component of the pen is poor. The results obtained showed, however, that the variation in ICC coefficients between experiments can be large. Blood parameters were measured for one trial only. Variation between ICC coefficients among the different blood parameters was seen to be large. The highest ICC coefficient was found in the analysis of total protein (0·2276) and the lowest for long-chain fatty acids (≈ 0.0). ICC coefficient for urea, Hb and albumin was 0·0167, 0·0001 and 0·0048, respectively.

The ICC coefficient was low for all the variables measuring the quality of meat. On the other hand, the ICC coefficients were high for all the variables measuring behavioural patterns. The ICC coefficients and design effects were estimated according to formulae (2) and (3), respectively. Table 2 shows the estimated ICC coefficients and design effects for eight different variables.

Simulation study

Statistical theory implies that whenever the null hypothesis is true, the observed *P*-value follows the standard uniform distribution. The simulation study confirmed that the fixed model analysis (SAS/GLM)
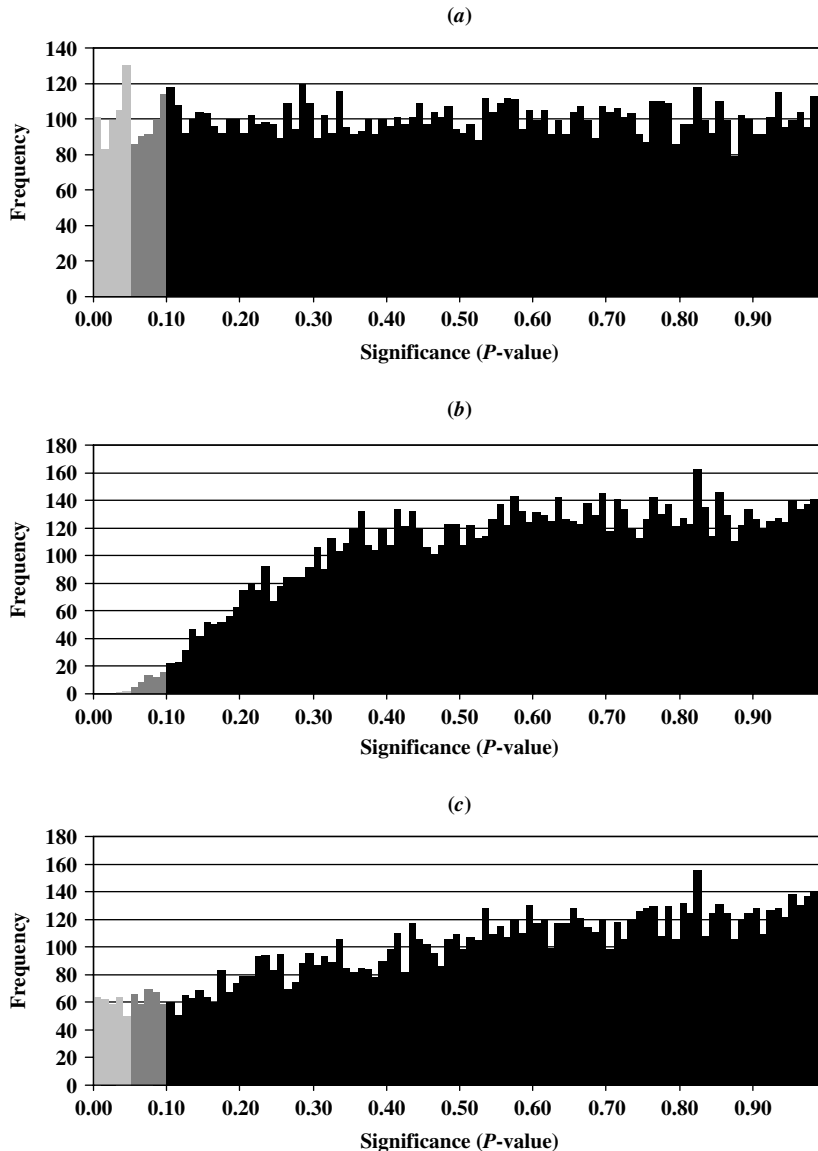
Fig. 1. Distribution of *P*-values in the simulation study ($N = 10\,000$) where the null hypothesis is true, the pen effect is slight and the number of pens/treatments is two. The analysis used three models: (*a*) standard fixed-effects model, (*b*) mixed-effects model, and (*c*) mixed-effects model with corrected denominator degrees of freedom. For each analysis, the domain of *P*-values was divided into 100 sections. If the analyses were valid, the computed *P*-values would be distributed evenly over these sections.

fulfilled this property when the ICC within pens was slight and the number of pens per treatment was two (Fig. 1*a*). The study further showed that the common mixed model approach did not work adequately with the presence of a slight ICC (Fig. 1*b*). In the simulation study, the probability of getting a statistically significant *P*-value (<0.05) by chance turned out to be only <0.001 instead of 0.05. Furthermore, the

probability for a *P*-value of <0.10 was only 0.006 instead of 0.10. Elevated *P*-values appeared when the estimated value of the variance component for the pen effect was non-positive. For positive estimates both models resulted in exactly the same *P*-value. The methods of Satterthwaite (1946) and Kenward–Roger (1997) were both used for correcting the degrees of freedoms when non-positive estimates of variance

Table 3. *Results of simulation: probability of obtaining a P-value less than 0·05 by chance when the number of pens per treatment was set at two, four or eight, the variance between pens was set at 10^2, 40^2 or 100^2 and the data were analysed using the fixed effects model (GLM), the mixed model (MIXED), and the mixed model with Kenward–Roger correction (KR MIXED). Each pen contained eight animals, the within-pen variance was set at 500^2 and 10 000 datasets were generated for each combination. The ICC coefficients represent the proportion of true variation that can be attributed to differences between the pens*

Number of pens per treatment	Variance between the pens	ICC	GLM	MIXED	KR MIXED
2	10^2	0·0004	0·052	<0·001	0·030
2	40^2	0·0064	0·052	<0·001	0·039
2	100^2	0·0400	0·048	0·006	0·068
4	10^2	0·0004	0·050	<0·001	0·032
4	40^2	0·0064	0·048	0·018	0·039
4	100^2	0·0400	0·046	0·039	0·046
8	10^2	0·0004	0·050	0·029	0·038
8	40^2	0·0064	0·052	0·037	0·045
8	100^2	0·0400	0·049	0·048	0·049

components occurred. With the simulated data both methods resulted in exactly the same corrections. After correction, distribution for the *P*-value was closer to the standard uniform distribution (Fig. 1*c*). The lack of small *P*-values was still clear: the probability of obtaining a *P*-value of 0·05 and 0·10 by chance was only 0·030 and 0·062, respectively.

For some simulated datasets, the choice of statistical model made a big difference. In one of the most extreme cases differences between treatments were statistically significant ($P<0·001$) when the GLM model was used, but a *P*-value of 0·51 was obtained when a mixed model analysis without denominator degrees of freedom adjustment was applied. Both adjustment methods corrected the *P*-value to 0·44.

Distribution of the *P*-values of mixed models followed the uniform distribution more accurately when the number of pens per treatment was increased from two to eight, or when the between-pen variance was increased from 10^2 to 100^2 (Table 3). However, if a mixed model without denominator degrees of freedom adjustment was used, the resulting *F*-tests were still too conservative in all combinations.

DISCUSSION

The current study provides information on the presence and extent of ICC in six experiments on suckler cows. Similar appearance of ICC may be expected in future studies. The results obtained show that ICC varies with the response variable considered. Depending on the response variable, it may be significant or negligible.

It is not always possible to include more than one pen per treatment. In particular, when different housing or growing conditions are compared, the conventional analysis of variance technique necessitates more than one cow house per treatment or

at least two housing systems are needed. In such experiments the researcher is easily tempted to use the animal as an experimental unit. Careless trust in the validity of the procedure may, however, lead to unsound conclusions. The only trustworthy way to make sure is to repeat the experiment at least twice and analyse the data using experiment as the experimental unit.

As expected, one of the highest ICC coefficients was found in the analysis of behaviour data. It is well known that cows are highly social animals with synchronized feeding and resting behaviours (e.g. Bouissou *et al.* 2001). At pasture they have enough space for synchronizing their behaviour and offset each other, whereas housed in a pen they try to minimize agonistic interactions by avoiding each other's individual zones (Miller & Wood-Gush 1991). Thus, in pens, all cows might not perform the same behaviours at the same time. This means that the ICC coefficient is usually higher at pasture than in a small pen.

Negative ICC may exist in small pen sizes due to competition for food and space. Under the mixed model structure, negative correlation coefficients are inadmissible. The common correlation model does, however, allow for negative values for the ICC coefficients (Murray & Blitstein 2003), and may therefore provide feasible alternatives in studies with small pen sizes and antagonism between the animals.

The simulation study showed that the common mixed-model approach had striking problems when the ICC was small. The simulation study, however, indicated that the mixed model approach could be improved by approximating the error degrees of freedom by using the Kenward–Roger correction (Kenward & Roger 1997) or Satterthwaite's method (Satterthwaite 1946). Both correction methods are implemented in several commercial statistical packages

and are recommended if mixed-model analysis is applied. Current results, however, showed that the consequent *F*-tests could still be too conservative, especially if there were only two pens and the ICC was small (< 0.0400).

In the current study, ICC was used for measuring dependence of normally distributed variables only. The same techniques can also be used for any continuous and binomial variables (Gulliford *et al.* 1999). Animal behavioural studies are, however, more complicated than traditional feeding experiments. For instance, a fence can physically restrict animals, but it does not have a restrictive influence on vocalization. Therefore, the current study covered only a small aspect of the necessary investigations into the effect of a pen on the animals studied. In particular, much more effort should be placed on the planning process and statistical analysis of behavioural studies.

CONCLUSIONS

The current study gives some indication that, in certain situations, within-pen units may be valid experimental units. This does not, however, extend to a general result. Therefore, if pens are used in animal experiments, the only safe and sound technique to analyse the resulting data is also to use a pen as the experimental unit. If the ICC is small, then the fixed effects model is the only valid model. The mixed model approach is useful when the ICC is more notable. Degrees of freedom adjustment methods should be used when the mixed models are fitted to the data.

REFERENCES

Bouissou, M. F., Boissy, A., Le Neindre, P. & Veissier, I. (2001). The social behaviour of cattle. In *Social Behaviour in Farm Animals* (Eds L. J. Keeling & H. W. Gonyou), pp. 113–145. Wallingford, UK: CAB International.

Donner, A. (1992). Sample size requirements for stratified cluster randomization designs. *Statistical Medicine* **11**, 743–750.

Gomez, K. A. & Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research*, 2nd edn. New York, USA: John Wiley and Sons.

Gulliford, M. C., Ukoumunne, O. C. & Chinn, S. (1999). Components of variance and intraclass correlation for the design of community-based surveys and intervention studies: data from the health survey for England. *American Journal of Epidemiology* **149**, 876–883.

Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997.

Kiker, W., Salisbury, M., Green, B. & Engdahl, G. (2005). Effects of protein and energy feeding on ovine oocyte production and developmental capacity. *Proceedings, Western Section of the American Society of Animal Science* **5**, 312–316.

Kish, L. (1965). *Survey Sampling*. New York, USA: John Wiley and Sons.

Korhonen, H.T., Niemelä, P. & Jauhiainen, L. (2001). Effect of space and floor material on the behaviour of farmed blue foxes. *Canadian Journal of Animal Science* **81**, 189–197.

Lucas, H. L. (1948). Techniques in animal science research. In *Proceedings of the Auburn Conference on Statistics Applied to Research*, pp. 62–73. Auburn, Alabama, USA: Alabama Polytechnic Institute.

Manninen, M., Saarijärvi, K., Huhta, H., Jauhiainen, L. & Aspila, P. (2004). Effects of winter feeding strategies with alternative feeds on the performance of mature suckler cows and their progeny. *Agricultural and Food Science* **13**, 348–362.

Manninen, M., Virkajärvi, P. & Jauhiainen, L. (2005). Effect of whole-crop barley and oat silages on the performance of mature suckler cows and their progeny in outdoor winter feeding. *Animal Feed Science and Technology* **121**, 227–242.

Manninen, M., Sormunen-Cristian, R., Jauhiainen, L., Sankari, S. & Soveri, T. (2006). Effects of feeding frequency on the performance and welfare of mature Hereford cows and their progeny. *Livestock Science* **100**, 203–215.

Miller, K. & Wood-Gush, D. G. M. (1991). Some effects of housing on the social behaviour of dairy cows. *Animal Production* **53**, 271–278.

Morris, T. R. (1998). *Experimental Design and Analysis in Animal Sciences*. Wallingford, UK: CAB International.

Murray, D. M. & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomised trials. *Evaluation Review* **27**, 79–103.

Robinson, P. H., Wiseman, J., Uden, P. & Mateos, G. (2006). Some experimental design and statistical criteria for analysis of studies in manuscripts submitted for consideration for publication. *Animal Feed Science and Technology* **129**, 1–11.

SAS Institute Inc. (1999). *SAS/STAT User's Guide, Version 8*. Cary, NC, USA: SAS Institute.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110–114.

Siddiqui, O., Hedeker, D., Flay, B. R. & Hu, F. B. (1996). Intraclass correlation estimates in a school-based smoking prevention study: outcome and mediating variables, by sex and ethnicity. *American Journal of Epidemiology* **144**, 425–433.

Swamy, H., Smith, T. & MacDonald, E. (2004). Effects of feeding blends of grains naturally contaminated with *Fusarium* mycotoxins on brain regional neurochemistry of starter pigs and broiler chickens. *Journal of Animal Science* **82**, 2131–2139.