Epidémiol. santé anim., 1997, 31-32

DESIGN AND EVALUATION OF FIELD STUDIES ON SUBCLINICAL MASTITIS DEALING WITH INTERDEPENDENCE AMONG QUARTERS OF THE BOVINE UDDER

Barkema H.W.[1], Schukken Y.H.[2], Lam T.J.G.M.[3]

Le problème fondamental lié à non indépendance des données est la sous estimation des variances et la possibilité de biaiser l'estimation des effets. La méthode la plus simple pour tenir compte de la non indépendance est celle qui utilise le facteur d'inflation de la variance (FIV). La variance estimée par la méthode classique est multipliée par le facteur FIV : variance robuste=variance x FIV. Le FIV est estimé par : 1 + (n-1) x Corrélation Intra Classe (CIC), où n est la taille moyenne des groupes. En supposant que toutes les vaches ont 4 quartiers en production : FIV = 1 + (3xCIC). Les équations d'estimation généralisées (GEE) peuvent être appliquées sur les données binaires. Elles ont l'avantage de préserver la distribution binomiale des données. Les inconvénients de GEE sont la difficulté d'ajuster les modèles complexes avec plusieurs niveaux de regroupement. D'autres méthodes comme l'algorithme de Schall ou les modèles linéaires généralisés mixtes peuvent être utilisées.
Le transfert des agents pathogènes entre les vaches et entre les quartiers d'une même vache est possible. Lorsqu'on cherche à modéliser ce type de transfert la corrélation entre les quartiers ne peut plus être considérée comme un facteur de nuisance. Lam et al. (1996) ont appliqué un modèle de contagion ajusté pour étudier la circulation de S. aureus entre les quartiers et les vaches. Le modèle permet également de tester les effets de covariables comme la parité et le jour de lactation.

INTRODUCTION

A considerable amount of research has been devoted to the understanding of subclinical mastitis, and, based on this understanding, guidelines to evaluate experimental data have been developed. Current guidelines to evaluate clinical and subclinical infection status have been based on the assumption that quarters within a cow are independent and that an equal probability exists that either the left or right quarters could be infected (Hogan et al., 1990). However, cross-infections of pathogenic bacteria within cows have been described.

Interdependence of quarters is either based on similar risk factors of the quarters within a cow or quarters within a herd, or on contagiousness of microorganisms. Transmission of intramammary infecition (**IMI**) does not only occur among cows but also among quarters within a cow. When several quarters are infected in the proximity of a non-infect quarter, risk of IMI in the non-infected quarter is high. Because known contagious pathogens such as *Strep. agalactiae* (Dodd, 1983) and *Staph. aureus* (Lam et al., 1996) are mainly transmitted during milking (including preparation before milking), interdependence of quarters based on transmission is more likely for these pathogens than for pathogens such as *Strep. dysgalactiae*, *Strep. uberis*, and *Escherichia coli*. Coagulase-negative staphylococci and *C. bovis* are contagious minor pathogens (Lam et al., 1997).

The difference between prevalence of subclinical mastitis at the cow and quarter levels can be corrected by using intraclass correlation (**ICC**), which describes the strength of clustering (Donald, 1993). In a Dutch study interdependency among udder quarters with subclinical mastitis was evaluated on 150 farms using a total of 35,828 udder quarters (Barkema et al., 1997). High somatic cell count (**SCC**) (>250,000 cells/ml) in zero, three, and four quarters occurred at a higher rate than would be expected based on independence of the quarters. For all bacterial species, intramammary infection in zero, two, three, or four quarters of the same cow occurred at a higher rate than would be expected based on independence of the quarters. Intramammary infection and high SCC were less often found in front quarters than in rear quarters. Interdependence between IMI and high SCC in quarters both within herd and within cow was strong (Table 1). The ICC within herd ranged from 0.01 to 0.11; ICC within cow ranged from 0.08 to 0.28. The difference in ICC among pathogens was considerable. Differences in ICC among herds were especially apparent for known contagious pathogens such as *Strep. agalactiae* (Dodd, 1983), *C. bovis*, coagulase-negative staphylococci, and *Staph. aureus* (Lam et al., 1996). The difference found between left and right quarters in infection status has implications for the design of experimental studies. If not corrected, this difference could lead to biased estimates of effect, especially in studies using a split udder design.

In the case of subclinical mastitis, the interdependence of quarters should be taken into consideration when experiments are designed and evaluated, which is particularly important for evaluating dry cow products, lactation treatments, and teat disinfection products. In such studies with measurements on the quarter level, ignoring interdependence among quarters could lead to a serious underestimation of variance and an associated increase of Type 1 error. In split udder trials, the assumption of independence of quarters could lead to an underestimation of treatment effects because of the contagiousness of microorganisms.

When interdependence is ignored and when quarters are treated as independent observations, statistical significance tests may be unreliable. Specifically, the tests may lead to an underestimation of Type 1 error

[1] Animal Health Service, PO Box 361, 9200 AJ Drachten, The Netherlands
[2] Department of Herd Health and Reproduction, Utrecht University, PO Box 80151, 3508 TD Utrecht, The Netherlands
[3] Department of Infectious Diseases and Immunology, Utrecht University, PO Box 80165, 3508 TD Utrecht, The Netherlands

because variance is underestimated. Additionally, understanding clustering of disease (IMI), either within a cow or within a herd, may be of considerable interest and may lead to further understanding of the dynamics of disease. Interdependence based on contagiousness may also lead to an underestimation of treatment effects (Lam et al, 1996).

IMPLICATIONS OF INTERDEPENDENCE FOR MASTITIS TRIALS

The choice of methods to account for interdependence depends on two main considerations: the objectives of the study and the assumptions that can be made about the nature of the correlation structure (McDermott et al., 1994). If the study objectives include the investigation of factors on both the quarter and cow level, or even on the herd level, some type of statistical control is required. However, if the study objectives are to test treatment efficacy and no inferences are to be made on risk factors for IMI or clinical mastitis at the cow or herd level, a more effective control of cluster effects can be achieved by the design of the study. One method to avoid some of the problems of ICC within cow when the udder quarter is the unit of interest is matching case quarters with controls within the same cow. This procedure allows the investigator to focus on individual quarter risk factors and to eliminate cow effects (McDermott et al., 1994). In clinical trials, this method can be applied using a split udder design (Schukken et al., 1993). Interdependence because of contagiousness of pathogens is not eliminated in this design but can be corrected (Lam et al., 1996).

The two fundamental problems when interdependence is present are that variance is underestimated (Lam et al., 1997, McDermott et al., 1994) and that effect estimation may be biased (Lam et al., 1996). Some methods to correct for interdependence among observations in the analysis are described subsequently. In the first set of methods discussed, variance inflation factor (**VIF**) and general estimation equations (**GEE**), the major goal is to reduce underestimation of variance.

A straightforward method to correct the variance for interdependence in the analysis of the data is to use VIF (McDermott et al., 1994). The calculated variance must then be multiplied by this VIF: robust variance = VIF x simple variance. The VIF is estimated as VIF = 1 + [(n - 1) x ICC], where n = mean cluster size. Because of the size of our study population, stable estimates of pathogen-specific ICC were obtained from our data. These estimates may be used in other trials, assuming that the nature of clustering (e.g., through milking machine) is similar to our study. Assuming that all cows have four quarters in production, VIF = 1 + (3 x ICC). The ICC for each pathogen is presented in Table I.

Table I
Variance components and intraclass correlation (ICC) of quarters infected per cow (Barkema et al., 1997)

	Variance component				
	Herd	Cow within herd	Error	ICC Within herd	ICC Within cow
Natural logarithm of SCC	0.173973	1.339574	1.517270	0.06	0.47
Streptococcus dysgalactiae	0.000602	0.001561	0.015787	0.03	0.09
Streptococci Lancefield group D	0.001583	0.002896	0.024477	0.05	0.11
Streptococcus agalactiae	0.000329	0.000852	0.002215	0.10	0.28
Streptococcus uberis	0.000073	0.001016	0.005703	0.01	0.15
Other streptococci	0.003053	0.002718	0.031101	0.08	0.08
Staphylococcus aureus	0.002229	0.004763	0.033876	0.05	0.12
Coagulase-negative	0.008797	0.020141	0.091822	0.07	0.18
staphylococci	0.023866	0.046445	0.138183	0.11	0.25
Corynebacterium bovis	0.001489	0.001394	0.010290	0.11	0.12
Bacillus spp.					

As an example, the results of a study of Hogan et al. (1987) are discussed. The prevalence of IMI in 16 herds using four different teat disinfection practices was studied. The mean number of animals per farm (cluster size) was 57 cows. Those researchers found an 8.5% (SD = 3.4; variance ≅ 11.6) quarter prevalence of *Staph. aureus* IMI in four herds that did not practice teat disinfection compared with a prevalence of 1.6% (SD = 1.3; variance ≅ 1.7) in four herds that used an iodophor teat dip as a postmilking teat disinfectant. The VIF within herd for *Staph. aureus* for that study (9) would be VIF = 1 + (56 x 0.05) = 2.8. The VIF within cow for *Staph. aureus* for that study would be VIF = 1 + (3 x 0.12) = 1.36. The VIF, corrected for both within cow and within herd clustering, would be VIF = 2.8 x 1.36 = 3.81, which leads to a substantial increase in variance. Results of a significance test would be affected by this correction.

Statistical models that account for clustering and simultaneously estimate effects have recently become available. Simultaneous estimation is preferable, because covariates that are included in the analysis may actually explain part of the correlation that is present in the raw data. For example, part of the within cow clustering of IMI could be explained by age, stage of lactation, or other diseases. Hence, the actual ICC that is present in the data after correction for these covariates is substantially smaller than the ICC estimated from the

raw data.

For binomially distributed data, the regression models that allow for simultaneous estimation are either based on transforming the outcome variable and subsequently applying techniques assuming a normal distribution (Schall, 1991) or on models that allow for ICC and preserve the original distribution of the data. Zeger et al. (1988) introduced GEE that are an example of the latter strategy. A number of error distributions, including the binomial, and a number of link functions, including the logistic function, are allowed. These GEE methods are then comparable with simple logistic regression but allow simultaneous inclusion of correlation structures among observations. Several correlation structures have been proposed, including autoregressive, m-dependent, compound symmetry, or an unspecified structure. For the intracow correlation of quarters, likely a compound symmetry structure, assuming equal correlation among all quarters is applicable. This latter method was used by Leslie et al. (1994) in the analysis of dry cow treatment data. The ICC estimated from our data may be used as an initial estimate of correlation structure in GEE.

The limitation of GEE is that multilevel models cannot easily be fitted, which may occur when ICC is not only present at the cow level but also at the herd level. In the situation of multiple levels of clustering, a linearized approach, such as Schall's algorithm (1991) or a generalized linear mixed models procedure, such as applied in the macro GLIMMIX in SAS, may be utilized. These models allow the inclusion of multiple levels of correlation among observations. No applications using these latter methods in udder health data have been published to date.

In summary, VIF methods can be used to correct variance after estimation of effects, and GEE or similar methods can be applied simultaneously with effect estimation. The previously mentioned methods mostly eliminate underestimation of variance. Underestimation of efficacy because of contagiousness of microorganisms may still be present. This contagiousness may be modeled, and efficacy may be corrected.

Transfer of pathogens among cows and from quarter to quarter within cows may occur. When the pattern of infection among animals can be modeled, the correlation of quarters within a cow is not a nuisance factor, but the main interest of the modeling procedure. A model to describe the effect of the number of existing IMI on the number of new IMI has been described by Lam et al. (1996). An adjusted contagious disease model was applied to *Staph. aureus* IMI dynamics within and among cows. The described modeling procedure allowed estimation of effects of covariates such as parity and DIM. Possible biases in the effect estimation were then corrected in the analysis.

CONCLUSIONS

When studying IMI or subclinical mastitis prevalence on an animal or herd level and when measurement is done on the quarter level, correction for interdependence has to be considered. Statistical methods to correct for the underestimation of the variance are available.

REFERENCES

Barkema, H.W., Y.H. Schukken, T.J.G.M. Lam, D.T. Galligan, M.L. Beiboer, and A. Brand. 1997. Estimation of interdependence among quarters of the bovine udder wth subclinical mastitis and implications for analysis. J. Dairy Sci., accepted.

Dodd, F.H. 1983. Mastitis--progress on control. J. Dairy Sci. 66:1773.

Donald, A.W. 1993. Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. Prev. Vet. Med. 15:125.

Hogan, J.S., D.M. Galton, R.J. Harmon, S.C. Nickerson, S.P. Oliver, and J.W. Pankey. 1990. Protocols for evaluating efficacy of postmilking teat dips. J. Dairy Sci. 73:2580.

Hogan, J.S., D.G. White, and J.W. Pankey. 1987. Effects of teat dipping on intramammary infections by Staphylococci other than *Staphylococcus aureus*. J. Dairy Sci. 70:873.

Lam, T.J.G.M., M.C.M. De Jong, Y.H. Schukken, and A. Brand. 1996. Mathematical modeling to estimate efficacy of postmilking teat disinfection in split-udder trials of dairy cows. J. Dairy Sci. 79:62.

Lam, T.J.G.M., J.H. van Vliet, Y.H. Schukken, F.J. Grommers, A. van Velden-Russcher, H.W. Barkema, and A. Brand. 1997. The effect of discontinuation of postmilking teat disinfection. II. Dynamics of intramammary infections. Vet. Q. 19:accepted.

Leslie, K.E., K. Bateman, D. Barnum, and Y.H. Schukken. 1994. Effect estimation of repeated versus single dry cow treatment, correcting for intracow and intraherd correlation using generalised estimating equations. Kenya Vet. 18:149.

McDermott, J.J., Y.H. Schukken, and M.M. Shoukri. 1994. Study design and analytic methods for data collected from clusters of animals. Prev. Vet. Med. 18:175.

Schall, R. 1991. Estimation in generalized linear models with random effects. Biometrika 78:719.

Schukken, Y.H., J. van Vliet, D. van de Geer, and F.J. Grommers. 1993. A randomized blind trial on dry cow antibiotic infusion in a low somatic cell count herd. J. Dairy Sci. 76:2925.

Zeger, S.L., K.Y. Liang, and P.S. Albert. 1988. Models for longitudinal data: a generalized estimation equation approach. Biometrics 44:1049.