ESTIMATION OF CORRELATIONS BETWEEN DISEASE RESISTANCE AND PRODUCTION TRAITS – CONCEPTS AND PROBLEMS

H. Simianer
Department of Animal Husbandry and Animal Breeding,
University of Hohenheim, P.O. Box 700562, 7000 Stuttgart, FRG

SUMMARY

The problem of estimating genetic correlations between resistance to diseases and production traits is considered. Often, only binary observations whether or not an animal has suffered a certain disease are available. An estimation procedure accounting for this discrete observation structure was suggested by Simianer and Schaeffer (1989). Problems due to restrictions in the methodology as well as to disturbances in the observed data are adressed. The impact of pleiotropy or linkage of major loci causing correlations between resistance and production traits is discussed.

INTRODUCTION

A case of illness of an animal is the result of two interacting components: the individual's present status of resistance against (or, to say it negatively, its susceptibility to) the disease on one hand and the environmental conditions being more or less favourable for an outbreak of the disease on the other hand. Disease resistance surely is not measurable as milk yield or lifeweight is, it is rather a product of complex physiological processes which are only partly understood. It is possible though to measure components of these regulatory systems like the efficiency of certain immunological pathways.

As an alternative, disease resistance may be quantified by statistical means. Animal breeders are mainly interested in the additive genetic background of traits. In dairy cattle there is evidence for both variance of disease resistance and covariance between disease resistance and production traits due to additive gene effects, as recently reviewed by Emanuelson (1988) and Shook (1989). For resistance against mastitis, a heritability of around .05 and a genetic correlation to milk yield of about –.30 must be assumed (Emanuelson et al., 1988), i.e. high yielding cows tend to show low resistance against mastitis. As a consequence, breeding for yield only will increase disease problems under constant environmental conditions.

The example of mastitis and milk yield will be used to illustrate statistical concepts of modelling disease resistance and of estimating genetic correlations to production traits. A number of difficulties particular to this estimation problem will be adressed and solutions will be suggested.

MODEL AND OBSERVATION STRUCTURE

The problem will be restricted to the quantitative genetic aspect, i.e. small additive gene effects at an infinite number of loci on both disease resistance and productive performance are considered exclusively. If disease resistance were measurable, a multi–trait mixed linear model could be fitted and used to estimate the dispersion parameters for, among others, the random additive genetic effects. It is not measurable though, so the best information available about individual disease resistance is whether or not an animal shows the disease. This phenotypic discrete observation ('healthy' or 'ill') can be linked to the underlying continuous resistance variable by the threshold concept (Wright, 1934). Each animal under its specific environmental conditions has its individual score on the underlying resistance scale, determined by its additive genetic value but also by effects of herd environment, age, season etc. Only if an individual's score falls short a fixed threshold value on the underlying scale, a disease is observed. For this conceptual 'resistance'–variable, the same model can be used as if it were observable, it is only necessary to define the form of the distribution, which usually is chosen to be normal.

This concept is superior to the use of measures attempting to describe the severeness of the disease, by using diverse 'mastitis scores', number of treatments, or the logarithm of somatic cell count in milk (SCC). SCC is measured routinely for all cows in some populations (see e.g. Emanuelson et al., 1988). An inflated SCC is however a *consequence* of mastitis, while decreased resistance in turn is a *causal factor*, so that SCC and mastitis resistance are *correlated traits* (Emanuelson et al. (1988) and Shook (1989) give estimates for the respective genetic correlation of around 0.6).

ESTIMATION PROCEDURES

Using binary disease observations and continuous production traits, the problem is to estimate variance and covariance components under a bivariate mixed linear model with one variable dichotomized. The most common approach is to ignore the non–normality and discontinuousness of the disease recordings and to use conventional linear model techniques, as reviewed by Emanuelson (1988). In the univariate case, such a strategy is known to lead to severely underestimated heritabilities, for which Dempster and Lerner (1950) have given an approximate transformation rule. A similar rule to transform estimated correlations between binary and continuous variables on the observed scale to the underlying scale is given by Kendall and Stuart (1961), showing that these correlations are similarly underestimated. The effect on estimated *genetic* correlations, however, is not at all clear, so that it is impossible to judge the reliability of estimates given in the literature, ranging from –.10 to +.66 for the genetic correlation of milk yield and clinical mastitis (Emanuelson, 1988).

Based on work of Foulley et al. (1983), Simianer and Schaeffer (1989) have suggested a method to estimate location and dispersion parameters which accounts for the discrete observation structure of the disease trait. The algorithm is derived from Bayesian arguments and uses a number of unavoidable approximations. In a simulation study, the method is shown to yield unbiased estimates of the variance components under certain conditions. Although computationally demanding, it can be used to analyze large scale data sets with more than 200,000 pairs of observations.

PROBLEMS AND DISCUSSION

Simianer and Schaeffer (1989) indicate a number of restrictions of their approach: the proposed method of estimating the residual correlation is largely heuristic (although leading to satisfactory results), there is a considerable overestimation of the heritability of the binary trait when the average smallest subclass size is less than 2 (parallelled by findings of Hõschele et al., 1987), but this does not affect the unbiasedness of the estimates of the genetic correlation. Also, the method does not allow for missing observations for the continuous trait. This, however, is of practical importance, as a certain percentage of diseased animals will be excluded from the production process, thus being recorded in the 'diseased' category but having no proper record for e.g. milk yield. Similarly disturbed observation structures are to be expected: some diseases may affect the productivity of the animals, the individual production level will have an impact on the decision whether a cow will be immediately culled or treated, production performance of culled cows may be wrongly extrapolated to a biased production level etc.. Unless suitable data are available and these implications are correctly modelled, estimates will be biased. Simianer (1990) has suggested an analytical approach to this sort of problem, considering the estimation of residual correlation in a fixed model. In this study it is shown, that increased mortality or reduced productivity due to a disease lead to an underestimation of the absolute value of residual correlations when not correctly accounted for, the downward bias increasing with a decreasing disease frequency. These results, however, can not be readily transferred to the estimation of genetic correlations. There the problem is much more complex and requires extensive simulation studies using the approach of Simianer and Schaeffer (1989).

Other problems arise when the disease frequency is low: if one per cent of the animals show the disease and the average number of daughters per sire is 20, the number of sires expected to have at least one diseased daughter is 18 per cent if resistance is not genetically determined and is even less if resistance is inherited. The sires having uniform non–diseased offspring do not show any variability *within* and *between* offspring groups, so that in this case at least 82 per cent of the sires do not contribute any information to the estimation problem.

Disease resistance is one classical area where hope is placed in the existance of single major loci having large effects, such as e.g. the major histocompatibility complex which has been shown to be effective in various species including cattle as reviewed by Lewin (1989). A single locus is assumed, whose effect accounts for a considerable share of the variability in disease resistance. This locus may be neutral or have a positive or negative pleiotrope effect on a production trait or may be linked to a major gene affecting a production trait. Both traits are also affected by quantitative genes at other loci. Recently, a close association between a recessive gene for *Weaver* and a major quantitative effect on milk and fat production in Brown Swiss dairy cattle was described (Hoeschele and Meinert, 1990). Ignoring the major gene effect, the additive genetic correlation technically can still be estimated, although the basic assumptions are violated. Selection based on these estimated parameters however would not lead to the predicted selection response. Molecular genetics will provide much better insight into the underlying genetic processes in the future. If an antagonism between disease resistance and production is due to *pleiotropy* in a major gene, the only prospect is to achieve an optimum gene frequency at this locus determined by economical considerations. If the antagonism is due to *linkage* between loci affecting resistance and production respectively, selection can aim at increasing the frequency of linkage between the favourable alleles. A stable association between the two traits can not be established unless both loci are fixed with respect to the favourable alleles. Smith and Simpson (1986) have demonstrated the limitations of marker assisted selection on major loci, so that the prospects have to be judged rather cautiously.

REFERENCES

DEMPSTER, E.R. and LERNER, I.M. 1950. Genetics **35**, 212–235.

EMANUELSON, U. 1988. Livest. Prod. Sci. **20**, 89–106.

EMANUELSON, U., DANELL, B. and PHILIPSSON, J. 1988. J. Dairy Sci. **71**, 467–476.

FOULLEY, J.L., GIANOLA, D. and THOMPSON, R. 1983. Genet. Sel. Evol. **15**, 401–424.

HOESCHELE, I., GIANOLA, D. and FOULLEY, J.L. 1987. J. Anim. Breed. Genet. **104**, 334–349.

HOESCHELE, I. and MEINERT, T.R. 1990. Subm. to J. Dairy Sci.

KENDALL, M. and STUART, A. 1961. The advanced theory of statistics. Vol. 2, Hafner, New York.

LEWIN, H.A. 1989. J. Dairy Sci. **72**, 1334–1348.

SHOOK, G.E. 1989. J. Dairy Sci. **72**, 1349–1362.

SIMIANER, H. 1990. Subm. to Biometrie und Informatik in Medizin und Biologie.

SIMIANER, H. and SCHAEFFER, L.R. 1989. Genet. Sel. Evol. **21**, 303–315.

SMITH, C. and SIMPSON, S.P. 1986. J. Anim. Breed. Genet. **103**, 203–217.

WRIGHT, S. 1934. Genetics **19**, 506–536.