Paper GM2.18 349 - 55th Annual Meeting of the EAAP, 5-8 September, 2004. Bled, Slovenia.
Corresponding author's e-mail: vitezica@germinal.toulouse.inra.fr

ESTIMATION OF RISK FACTORS USING GENOTYPE PROBABILITIES A SCRAPIE CASE

Z.G. Vitezica[1], C. Díaz[2], R. Rupp[1], J.M. Elsen[1]

[1] INRA, Station d'Amélioration Génétique des Animaux, 31326 Castanet Tolosan, France
[2] INIA, Depto. Mejora Genética Animal. Ctra. de la Coruña km 7,5, 28040 Madrid, Spain.

ABSTRACT

Susceptibility to scrapie is mainly controlled by polymorphisms at PrP gene. In some sheep breeds genotyping for the PrP gene is not extensively used therefore the information on PrP genotypes is incomplete. The objective of this work was to evaluate the potential use of genotype probabilities to handle records of non-genotyped animals in the estimation of risk associated to PrP genotypes. Original data consisted of 4049 Romanov sheep. Out of those, 447 animals died of scrapie. Data were analyzed using survival analysis techniques. Three models differing in the way that PrP genotype information was handled were tested. Firstly, records of untyped individuals were discarded; secondly those animals were assigned to an unknown group; thirdly probabilities of genotypes were assigned. In addition to the original situation, a second scenario where the non-genotyped individuals were not a negligible part of uncensored records was simulated. Probabilities of genotypes allowed to rank genotypes as expected. The use of genotype probabilities was a useful technique to deal with records of individuals with unknown genotype to estimate risk associated to PrP genotype and other transmission factors.

INTRODUCTION

Scrapie is one of several diseases known as Transmissible Spongiforme Encephalopathies that affect animals and humans. Resistance/susceptibility of sheep to scrapie is largely under the control of the PrP gene (Hunter *et al*., 1996). In sheep, several polymorphisms at codons 136 (A or V), 154 (R or H) and 171 (R or Q or H) have been associated to resistance-susceptibility to scrapie. In general, VRQ allele is associated with scrapie susceptibility and ARR is related to resistance.

Large-scale genotyping began a few years ago in order to breed for genetic resistance. Although in some population genotyping is extensively used, the information on genotypes is often incomplete due to the presence of non-genotyped animals. In these situations, the exclusion of non-genotyped animals from the data analysis may affect the estimates of other factors included in the model (Searle, 1971). Therefore, any tool to handle records of non-genotyped individuals is desired. The usefulness of genotype probabilities was pointed out by Kinghorn *et al.* (1993) in the context of major genes and by Meuwissen and Goddard (1997) in the context of QTL detection. However, genotype probabilities have not been explored in the

Corresponding author's e-mail: vitezica@germinal.toulouse.inra.fr

context of survival analysis. The objective of this work was to evaluate the potential use of genotype probabilities to handle records of non-genotyped individuals in the estimation of risk associated to PrP genotypes.

MATERIALS AND METHODS

Data. Data consisted of 4049 Romanov animals alive in an INRA experimental flock. ("Langlade") between the 1st of April 1993 and the 4th of March 2002. Among those, 447 died of scrapie (uncensored data). For the survival analysis, a record was considered as censored when the animal either died of other causes or did not die before March 2002. There were 1310 non-genotyped animals. Among them, 5 died of scrapie.

Compared strategies.
Three approaches relative to the treatment of missing genotype information might be used. Firstly the non-genotyped animals are excluded of the data analysis (P1) as in Elsen *et al.,* (1999). Secondly those non-genotyped animals are grouped in an unknown class (P2). Finally probabilities of genotypes are assigned (P3). P2 and P3 were compared by Díaz *et al.* (2003). The authors argued that results were similar because most uncensored records had genotype information. For this reason, P1, P2 and P3 were compared for two situations. The first situation corresponded to the original scenario (O) with only five individuals in the uncensored data. The second situation was simulated such that 50% of scrapie animals (uncensored data) were assumed to have unknown genotype (S).

Statistical Method.
Survival analysis techniques were used. This analysis models the hazard of an animal to be affected by scrapie at time t provided that it has not been showed till that moment. A Cox proportional hazards model was considered in the analyses. Under this model, the hazard of an animal to die of scrapie at time t is written as the product of a specified baseline λ_0 and a set of explanatory variables or stress factors $e^{x'\beta}$ modifying the baseline. Two models were run. Both models had a common part, including significant transmission effects and differed in the way that PrP genotype information was handled. Firstly, survival analysis was performed as

$$\lambda_i(t) = \lambda_0(t)\exp\left\{F_j + I_k(\pi) + Sx_l + PrP_m + R_o\right\},$$

where F_j is the flock effect representing the effect of the experimental group (j=1 to 5), I_k is a time dependent effect which is the interaction between individual's age and the level of infection assuming that changes occur at the beginning of each lambing season (π). Sheep were classified into three groups of age, animals up to 2yrs, 3 yrs and older than 3 yrs; Sx_l is the effect of sex; and R_o is the interaction between rearing type (maternal rearing or artificial rearing) and dam's disease status (scrapie or non-scrapie dams). PrP_m is the effect of the animal's genotype (known or unknown). Under P1, this model included ten classes for the effect of the PrP genotype of the animals. An additional class was included to account for non-genotyped animals under P2. In the second model, the hazard was modeled similarly except for the genotype effect. To take into account genotype probabilities for non-genotyped animals, each individual was assigned a corresponding vector of probabilities:

Corresponding author's e-mail: vitezica@germinal.toulouse.inra.fr

$$\lambda_i(t) = \lambda_0(t)\exp\left\{F_j + I_k(\pi) + Sx_l + \sum_m^{10} b_m x_{im} + R_o\right\}.$$

Probabilities x_{im} of the ten m possible genotypes were estimated for each animal i, with an effect b_m which is the regression coefficients that represents the effect of the m genotypes on the hazard. For a genotyped animal, the genotype probability was equal to 1.0 at the position corresponding to its genotype and 0.0 otherwise. The non-genotyped individuals but whose parents were homozygous were considered as genotyped animals. For non-genotyped animals, genotype probabilities were obtained using an iterative peeling approach. Details are provided in Vitezica (2003). This model was used in P3. The analyses were performed using the software package Survival Kit V3.12 (Ducrocq and Sölkner, 1998).

RESULTS AND DISCUSSION

The effect of the genotype on the risks of individuals was expressed as risks of carrying a given genotype relative to the risk of carrying ARQ-VRQ genotype. To compare results among the three strategies, relative risk of individuals associated to each genotype had to be calculated. Under P3 the relative risk of the genotype m was calculated from the estimate of b_m, as $\exp(b_m x_m)$ for x_m equal to one. Relative risks associated to the different PrP genotypes for P1, P2 and P3 under O and S are shown in Figure 1. Similar patterns were observed for the three strategies. The most susceptible genotypes were VRQ-VRQ, ARQ-VRQ and ARQ-ARQ. Identical ranking for the most susceptible genotypes was described in Elsen *et al.* (1999).

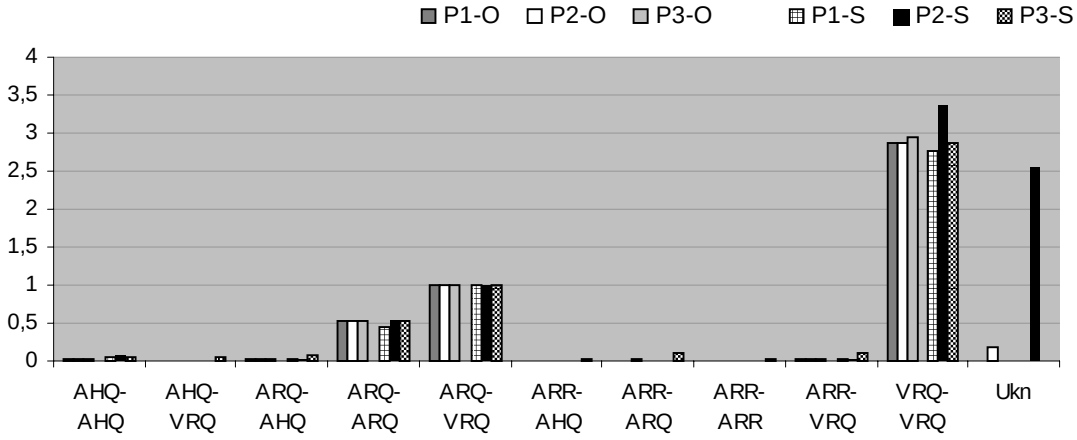

Figure 1. Relative risk of genotypes and the unknown class (Ukn) to ARQ-VRQ genotype with different approaches to treat missing genotype information: non-genotyped individuals were eliminated (P1), non-genotyped individuals formed the unknown class (P2), the probabilities of genotypes were included (P3), under the original (O) and the simulated situation (S).

Corresponding author's e-mail: vitezica@germinal.toulouse.inra.fr

Regardless the strategy (P1, P2, P3) and the situation (O, S), VRQ-VRQ genotype had a risk almost three times higher than the heterozygote ARQ-VRQ, and almost six times higher than the homozygotes ARQ-ARQ. Going from P2-O to P2-S resulted in an increase in the relative risk associated to the unknown class. The unknown class changed from the fourth position in P2-O to the second position in P2-S. These results showed an increase in the number of individual carrying susceptible genotypes assigned to the unknown group. In general, the risk associated to the unknown class will depend on the underlying distribution of susceptible genotypes. On the other hand, the strategy P3 allowed to include non-genotyped individuals in the analyses eliminating the uncertainty of the unknown class.

Relative risks among genotypes were studied for a range of values of probabilities. Relative risks associated to the genotypes computed as $\exp(b_m x_m)$ for different values of probability (x_m) are shown in Figure 2. Risks conferred to genotypes seemed to work in the expected direction when x_m increases. Thus, as it was expected, when the VRQ-VRQ probability increased the risk increased, and when the ARR-ARR probability increased the risk decreased.

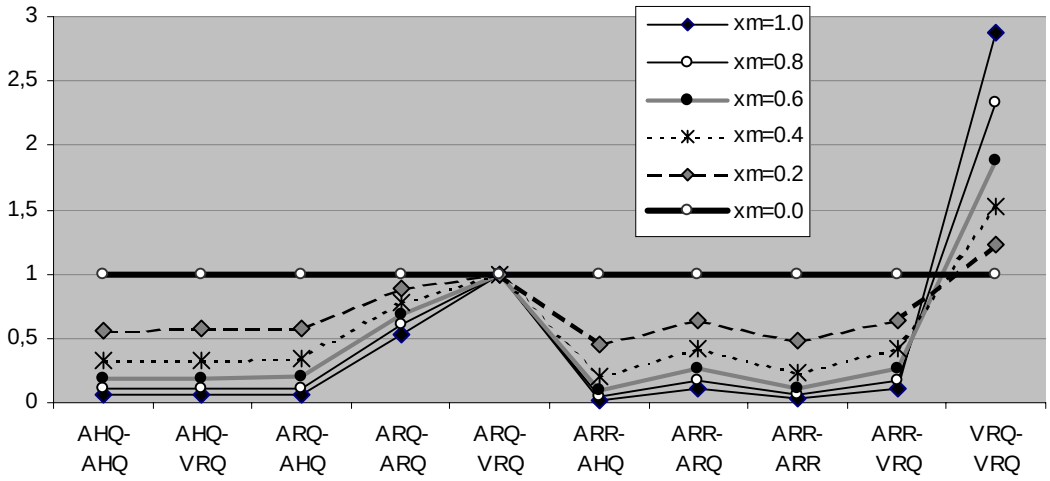

Figure 2. Relative risk of genotypes to ARQ-VRQ genotype for different values of probabilities (x_m) under the situation S.

However, this graph suggested that the probability of genotypes have to be assigned with a degree of certainty to be able to discriminate between genotypes. For an individual with known genotype, the probability of its genotype is equal to the unity. For a non-genotyped individual, this probability is theoretically divided among all possible PrP genotypes depending upon the information available for the estimation of probabilities. Thus, if animals received the same probability for each genotype (x_m equal to 0.1) then differences in risks associated to genotypes would almost be negligible. However, if the estimation of probabilities allows to

assign probabilities of at least 0.5, relative risks associated to each genotype may be discriminated from the others. In addition, P3 might tend to assign a certain relative risk to resistant genotypes because scrapie animals may have some probability of carrying ARR-ARR genotype provided the information available in this population. The probability assigned to animals with unknown genotypes will depend on the information (genotypes plus pedigree) available in each specific population.

An important issue of this type of analysis is the identification of other transmission factors included in the models, for instance, the flock or the sex. To estimate the risk associated to the transmission factors, P3 uses the same amount of information than P2 because the genotype probabilities of the animals included in the unknown class are estimated while for P1 they are discarded. In addition, to estimate the risk associated to genotypes, P3 strategy allows to eliminate the unknown class and the risk associated to it.

Therefore, probabilities of genotypes, P3, allow an effective use of data of non-genotyped animals with uncensored records to estimate risk factors associated to the different genotypes and other transmission factors.

ACKNOWLEDGEMENTS
Zulma Vitezica was funded by Ministerio de Educación, Ciencia y Técnica de la República Argentina and INRA. Clara Díaz was a Collaborator, via a fellowship under OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agriculture Systems and via INRA financial support. We are grateful to D. Lajous, and the staff of the experimental INRA farm of Langlade.

REFERENCES

- Díaz C., Vitezica Z.G., Rupp R., Elsen, J.M. 2003. El uso de la predicción de genotipos en el análisis de supervivencia: un caso de scrapie. *ITEA* 24 (II): 456-458.
- Ducrocq, V., Sölkner J. 1998. 'The Survival Kit'. A FORTRAN package for the analysis of survival data. *In: 6th WCGALP, Armidale, Australia.* 27: 447-448.
- Elsen, J.M., Amigues, Y., Schelcher, F., Ducrocq, V., Andreoletti, O., Eychenne, F., Tien Khang, J.V., Poivey, J.P., Lantier, F. and Laplanche, J.L. 1999. Genetic susceptibility and transmission factors in scrapie: detailes analysis of an epidemic in a closed flock of Romanov. *Arch.Virol.* 144: 431-445.
- Hunter, N., Foster, J.D., Goldman, W., Stear, M.J., Hope, J. and Bostock, C. 1996. Natural scrapie in a closed flock of Cheviot sheep occurs only in specific PrP genotypes. *Arch.Virol.* 141: 809-824.
- Kinghorn B. P., Kennedy B.W., Smith, C., 1993. A method of screening for genes of major effect. *Genetics*, 134: 351-360.
- Meuwissen, T.H.E., Goddard, M.E. 1997. Estimation of effects of Quantitative Trait loci in large complex pedigrees. *Genetics* 146: 409-416.
- Searle S.R., 1971. Linear Models, John Wiley & Sons, New York Inc.
- Vitezica, Z.G. 2003. Analyse of complex pedigrees for estimating genotypes. Application to the study of genetic resistance to scrapie. PhD Thesis, Institut National Agronomique Paris-Grignon, France.