

Understanding Field Disease Data

S.C. Bishop^{*} and J.A. Woolliams^{*}

Introduction

Disease genetics studies typically require observations on many animals in order to quantify genetic variation or perform a genome scan of sufficient power. This usually necessitates utilizing field data because, with the exception of some aquacultural species (e.g. Moen et al., 2009), challenge experiments of a sufficient scale are not possible. For example, data may be captured from a population undergoing an epidemic such as bovine tuberculosis (Brotherstone et al., 2010) or from endemic diseases with predictable prevalence. However, such field data is very ‘noisy’: diagnosis of infection or disease may be imprecise; it can be difficult to determine when infection of an individual occurred; and it is often unclear whether or not apparently healthy individuals have been exposed to the infection. Exposure and diagnostic test sensitivity or specificity are important concepts when studying the spread of disease in a population, and their impact on the estimation of genetic parameters has only recently been recognised by geneticists (Bishop and Woolliams, 2010). These authors derived formulae, in the context of viral or bacterial infections, to illustrate the biases caused to estimated heritabilities by these factors. They demonstrated that although incomplete exposure or poor diagnostic properties bias estimated heritabilities downwards, they are not fatal flaws. Indeed, the presence of genetic variation for disease resistance in field data is indicative of possibly much stronger underlying genetic control.

This paper extends the concepts proposed by Bishop and Woolliams (2010), demonstrating the similarity of the impact of incomplete exposure, diagnostic test sensitivity and incomplete data recording. It then extends the concept of exposure, demonstrating how it relates to force of infection, with impacts on modes of expression of resistance genotypes. Lastly, it compares diseases where phenotypes are binary classifications and diseases where quantitative measures of relative resistance are possible.

Material and methods

Quantifying the impact of non-genetic factors. The mathematical properties of incomplete exposure to infection are derived by Bishop and Woolliams (2010). Consider that a proportion ε of the population has been exposed to the pathogen; defining the virtual prevalence (p) as the proportion of individuals that have been exposed to the pathogen that become infected, and assuming that exposure is random and independent of host genotype, then the observed prevalence is εp . Heritabilities on the observed and liability scales (rearranged from Bishop and Woolliams, 2010) are then $\varepsilon^2 \phi(x_p)^2 (\varepsilon p)^{-1} (1 - \varepsilon p)^{-1} h^2$ and $\varepsilon^2 \phi(x_p)^2 \phi(x_{\varepsilon p})^{-2} h^2$, respectively, where h^2 is the liability heritability with complete

^{*}The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, UK

exposure. Further, animals will be classified into *healthy* and *diseased* categories by means of a diagnostic test, which will have defined values for specificity and sensitivity. Specificity (S_p) is the probability that a truly *healthy* individual is classified by the diagnostic test as *healthy* and sensitivity (S_e) is the probability that a truly *diseased* individual is classified by the diagnostic test as *diseased*. Observed prevalence is $p' = (1 - S_p) + (S_p + S_e - 1)p$, and heritabilities on the observed and liability scales are then, respectively, $(S_p + S_e - 1)^2 \phi(x_p)^2 p^{-1} (1 - p')^{-1} h^2$ and $(S_p + S_e - 1)^2 \phi(x_p)^2 \phi(x_{p'})^{-2} h^2$.

The form of the biases caused by incomplete exposure, imperfect diagnostic test sensitivity, and incomplete data recording are identical, and relate to the ability to correctly classify potentially susceptible animals as susceptible. In these equations, diagnostic test properties and exposure are considered independently, however often both factors will jointly affect the observed data. It is readily shown that with incomplete exposure and imperfect diagnostic tests, the observed prevalence is $p' = (1 - S_p) + (S_p + S_e - 1)\epsilon p$, where p is the true prevalence when $\epsilon = 1$. It then follows directly that heritabilities on the observed and liability scales are $\epsilon^2 (S_p + S_e - 1)^2 \phi(x_p)^2 p^{-1} (1 - p')^{-1} h^2$ and $\epsilon^2 (S_p + S_e - 1)^2 \phi(x_p)^2 \phi(x_{p'})^{-2} h^2$, respectively.

Exploring the concept of exposure. Exposure has been defined as the proportion of the population exposed to the pathogen, and thus at an individual animal level it is a binary category (exposed or not). Genetic differences in animal resistance/susceptibility to infection are then manifested in the probability that an animal becomes infected following exposure.

Underlying exposure is the concept of force of infection which, during an epidemic, is proportional to βI , where β is the transmission coefficient and I is the number or density of infectious individuals. Exposure may be considered a binary categorisation of the force of infection, with animals that face a force of infection greater than a specified quantity defined as being exposed. Force of infection is analogous to dosage levels in deliberate challenge studies where dosage levels are calibrated to obtain, e.g., 50% mortality during the course of the trial. Thus, for low infectious doses expected population mean mortality is close to zero and it rises with increasing dosage. In cases where disease-related mortality tends to unity, this curve may be conveniently described by the cumulative Normal distribution function $\int_0^x \phi(t; \mu, \sigma^2) dt$, where x is the actual dosage, μ is the dosage level achieving a population mean mortality or infection level of 50%, and the σ determines the rate at which infection or mortality increases as dose level increases. Within a population, different genotypes may differ in μ and/or σ . Similarly, the force of infection varies both between epidemics and over time within an epidemic, leading to exposure probabilities varying similarly.

In addition to the observable (polygenic) genetic variation varying with exposure, the mode of inheritance of major genes effects (i.e. recessive, additive, dominant) may apparently vary with force of infection or exposure. This is essentially a scale effect which determines the observable genetic effect (see Results section), and this effect is likely to alter both during and between epidemics.

Results and discussion

The impacts of incomplete exposure on estimated heritabilities are shown in Figure 1, for both the observed and liability scale. In both cases the bias due to incomplete exposure is near linear. It is more severe on the observed scale, but less influenced by virtual prevalence.

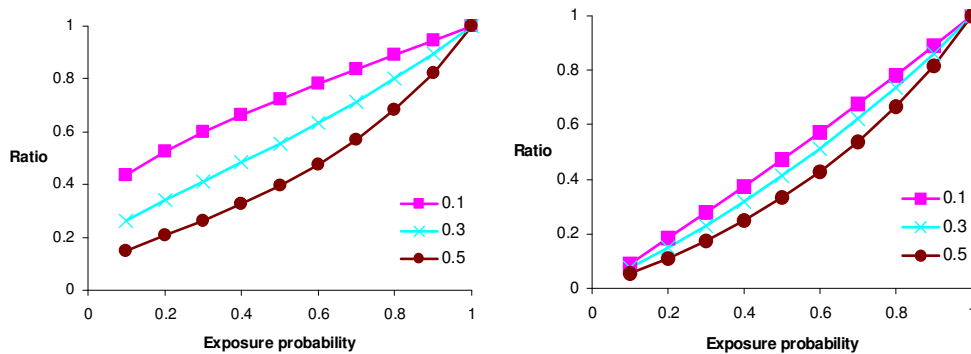


Figure 1: Ratio of observed to true heritability (liability scale left panel, observed scale right panel) for differing exposure probabilities and differing virtual prevalences.

The relationship between force of infection (or dosage level) and infection probability, assuming a cumulative Normal distribution function, is shown in Figure 2. Shown are three hypothetical major gene genotypes, differing in their resistance to infection. The vertical lines illustrate genotype comparisons at different dosage levels, and show apparently different modes of expression for resistance for the major gene. For the five cases shown, the genotype comparison would suggest (i) all animals resistant, (ii) resistance dominant, (iii) resistance additive, (iv) susceptibility dominant and (v) all animals susceptible. Noting that force of infection is proportional to βI , the apparent mode of inheritance is predicted to vary instantaneously during the course of an epidemic as well as between epidemics.

An example of this can be seen from a meta-prevalence analysis of published results for resistance to the viral disease infectious pancreatic necrosis in Atlantic salmon, where a single QTL has been shown in repeated studies to account for nearly all the genetic variation in the resistance of salmon fry (Moen et al. 2009, Houston et al. 2010, Gheyas et al. 2010). In the studies of Moen et al. and Gheyas et al., where in the challenge tests the population average mortality was close to 60% and 70%, respectively, the QTL effect was additive with no evidence of dominance. However, in the study of Houston et al., population average mortality was 9%, and the resistance haplotype was almost completely dominant over the susceptibility haplotype. Under the assumption that the meta-prevalence differences arise from differences in dosage level (or force of infection) due to the experimental protocols, these results fit almost exactly the 2nd and 3rd vertical bars in Figure 2.

Two further inferences may be drawn from the concept of the force of infection. Firstly, for natural microparasitic epidemics, because the force of infection varies as the epidemic progresses the order in which animals become infected provides information on relative

resistance. Assuming random contacts, animals that are infected earlier will have succumbed at a lower force of infection and hence are likely to be more susceptible than animals infected later. Secondly, so far exposure has been defined for diseases with binary outcomes. However, consider macroparasitic diseases such as ruminant nematodiasis, where challenge levels are continuous and outcomes are measured in terms of disease severity which is assumed to be a linear function of liability. Here, exposure may be considered to be relative rather than absolute and heritabilities for disease severity should increase as relative exposure increases. Indeed, this is often reported.

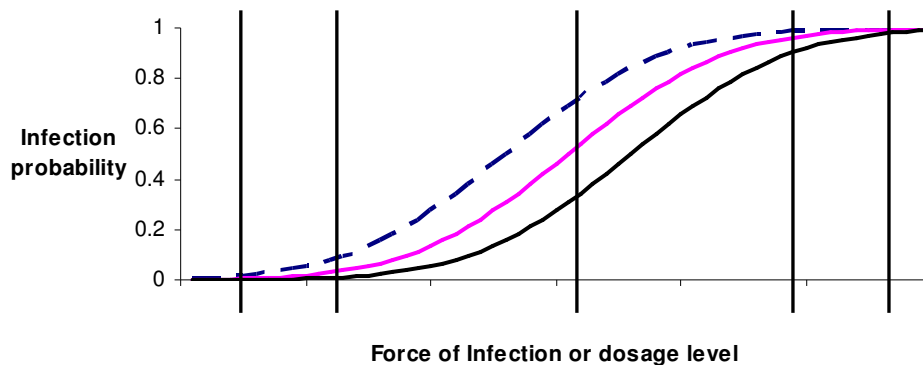


Figure 1: Infection probability as a function of force of infection (or dosage), for three hypothetical major-gene genotypes. The vertical lines represent different dosage levels, illustrating how the mode of expression of such genes may change with dosage level.

Conclusion

We have presented formulae which quantify the impact of incomplete exposure or imperfect diagnostic test sensitivity (or specificity) on estimated heritabilities. The form of the equations for these different scenarios are identical, and relate to the ability to correctly classify potentially susceptible (hence infected or diseased) animals, accounting the impact that this misclassification has on apparent disease prevalence. The concept of exposure is extended in relation to force of infection or dose-dependence effects, as may be seen during ongoing epidemics. These effects can result in resistance alleles manifesting as additive or dominant, as illustrated in the case of QTL for resistance to the disease IPN in salmon. Finally, during an epidemic, order of infection is also indicative of relative resistance.

References

- Bishop, S.C. and Woolliams, J.A. (2010). *PLoS ONE* 5(1): e8940.
- Brotherstone, S., White, I.M.S., Coffey, M. *et al.* (2010). *J. Dairy. Sci.* 93: 1234-1242.
- Gheyas, A., Houston, R.D., Mota-Velasco, J.C. *et al.* (2010). *Anim. Genet.*, In Press,
- Houston, R.D., Haley, C.S., Hamilton, A. *et al.* (2010). *Hered.*, doi:10.1038/hdy.2009.171
- Moen, T., Baranski, A., Sonesson, A.K. *et al.* (2009). *BMC Genomics* 10: 368.