

Estimating evolutionary parameters when viability selection is operating

Jarrold D. Hadfield*

University of Edinburgh, Edinburgh EH8 9YL, UK

Some individuals die before a trait is measured or expressed (the invisible fraction), and some relevant traits are not measured in any individual (missing traits). This paper discusses how these concepts can be cast in terms of missing data problems from statistics. Using missing data theory, I show formally the conditions under which a valid evolutionary inference is possible when the invisible fraction and/or missing traits are ignored. These conditions are restrictive and unlikely to be met in even the most comprehensive long-term studies. When these conditions are not met, many selection and quantitative genetic parameters cannot be estimated accurately unless the missing data process is explicitly modelled. Surprisingly, this does not seem to have been attempted in evolutionary biology. In the case of the invisible fraction, viability selection and the missing data process are often intimately linked. In such cases, models used in survival analysis can be extended to provide a flexible and justified model of the missing data mechanism. Although missing traits pose a more difficult problem, important biological parameters can still be estimated without bias when appropriate techniques are used. This is in contrast to current methods which have large biases and poor precision. Generally, the quantitative genetic approach is shown to be superior to phenotypic studies of selection when invisible fractions or missing traits exist because part of the missing information can be recovered from relatives.

Keywords: quantitative genetics; natural selection; breeders equation; selection bias; missing data

1. INTRODUCTION

Evolutionary change can be partitioned into change caused by an association of a trait with fitness and change caused by the transmission of the trait between generations (Price 1970). A central concern of evolutionary biology is to build realistic models of these processes and estimate the associated parameters. Quantitative genetics is the most well-developed theoretical and statistical framework for analysing the transmission of phenotypes between generations (Lynch & Walsh 1998). However, quantitative genetics has its roots in animal and plant breeding where interest is not on the nature of selection *per se*, but on the response to a selection regime imposed by the breeder (Falconer 1983). Nevertheless, early work in the field indicated that the relevant quantitative genetic parameters could not always be estimated without bias when selection was operating (e.g. Lush & Shrode 1950; Henderson *et al.* 1959; Curnow 1961). Subsequent work characterized methods of inference and patterns of selection that resulted in the unbiased estimation of variance components (Henderson *et al.* 1959; Curnow 1961; Thompson 1973; Rothschild *et al.* 1979; Meyer & Thompson 1984; Schaeffer 1987; Ouweltjes *et al.* 1988; Gianola *et al.* 1989) and prediction of random effects (Henderson *et al.* 1959; Henderson 1975, 1990; Goffinet 1983; Gianola *et al.* 1988; Fernando & Gianola 1990). In part, these results anticipated results derived from Rubin's

(1976) theory of missing data, which was applied in a quantitative genetic context by Im *et al.* (1989).

The key result from these studies is that selection does not have to be explicitly considered if complete information regarding the selection criterion is contained in the data. Selection is said to be ignorable. For example, if a farmer selects cows to produce a second lactation based on their performance in their first lactation, then a valid inference is only possible when the data include the performances of all cows in their first lactation. In animal breeding, it is generally assumed that any selective process takes this form and that the relevant data have been collected (Schaeffer *et al.* 1998). Explicit models of the selective process are therefore redundant, and methods for measuring selection have largely been developed in evolutionary biology (Manly 1985) from both quantitative genetic (Lande & Arnold 1983; Kirkpatrick *et al.* 1990; Rausher 1992; Janzen & Stern 1998) and ecological (see Metcalf & Pavard (2007) for a review) perspectives. In particular, Lande & Arnold's (1983) use of multiple regression has stimulated a large amount of empirical work on the strength and form of natural selection in wild populations (Kingsolver *et al.* 2001). With few exceptions, the methods developed for estimating the parameters of a selection process implicitly assume that the process is ignorable.

Two types of non-ignorable selective process have been discussed in the evolutionary literature. The first involves what I will call missing traits, where the trait of interest is correlated with unmeasured traits that directly affect fitness (Lande & Arnold 1983; Mitchell-Olds & Shaw 1987; Price *et al.* 1988). Here, unmeasured environmental variables are included in the set of missing traits, as they are non-ignorable in the same sense as unmeasured traits

*j.hadfield@ed.ac.uk

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2007.1013> or via <http://journals.royalsociety.org>.

One contribution of 18 to a Special Issue 'Evolutionary dynamics of wild populations'.

(Mitchell-Olds & Shaw 1987; Rausher 1992). When missing traits exist, the causal effect of the focal trait on fitness cannot be estimated from observational data, and selection on the genetic component of the phenotype differs from selection on the environmental component (Rausher 1992; van Tienderen & de Jong 1994). When missing traits are thought to exist, two recommendations have been suggested. The first is to estimate patterns of selection on the genetic component, as this may represent causal patterns better than the estimates of selection on the phenotype (Rausher 1992; Stinchcombe *et al.* 2002). The second suggestion is to use path analytic techniques to impose a structure on the model that may capture patterns of causality better than multiple regression-based techniques (Scheiner *et al.* 2002). A related problem to that of missing traits is when some individuals die before a trait is measured or expressed. These individuals are referred to as the invisible fraction (Grafen 1988). The problem of the invisible fraction has received less attention, despite empirical work on plants suggesting that it can compromise both quantitative genetic (Wei & Borralho 1998) and selection studies (Bennington & McGraw 1995). Accounting for the fact that selection may have already modified the distribution of a trait prior to it being measured is difficult, although Lynch & Arnold (1988) proposed a correction based on the strength of observable selection.

This paper discusses the concept of ignorability in the context of quantitative genetic—selection studies. Section 2 gives a brief overview of selection models that have been developed from a quantitative genetic perspective, and §3 gives a brief introduction to missing data theory that follows Rubin (1976) closely. In §4 use is made of the missing data theory to give the conditions under which natural selection is ignorable, and this section shows that nearly all forms of viability selection are non-ignorable. Section 5 shows that the force of viability selection is generally underestimated and develops some simple models for non-ignorable selection by extending basic survival analyses. Section 6 discusses why selection on developing traits is always non-ignorable and reviews developments in the medical sciences for dealing with this type of problem. Section 7 discusses traits that are not expressed until a stage in life where some mortality has already occurred. This paper discusses the problem of ascertaining whether the phenotypes of those individuals that survive to express the trait are a random sample of the possible phenotypes that would have been expressed in the absence of mortality. Although there does not seem to be a general solution in non-experimental systems, using pedigreed data, it is possible to ask whether they are a random sample with respect to breeding value. Section 8 highlights the problem of using predicted breeding values (PBVs) to address these problems and suggests that their use is easily avoided.

2. SELECTION AND QUANTITATIVE GENETICS

The following three traits are considered: relative fitness ω ; the trait of primary interest y ; and a nuisance trait x . Any phenotype can be partitioned into a population mean (\bar{y}), a breeding value (g_y) and an environmental deviation (e_y), $y = \bar{y} + g_y + e_y$. (2.1)

The population mean is defined with respect to the reference (base) population and the environmental

deviations and breeding values are deviations from this mean. The breeding value is twice the expected deviation of an individual's offspring's phenotype under certain assumptions such as random mating. In the simplest case, environmental deviations of different individuals are independent, whereas breeding values of different individuals may show dependence. Any dependency is accounted for by the additive genetic relationship matrix (A). Environmental deviations for the three traits may be dependent if they are from the same individual. Breeding values for the three traits may be dependent if they are from the same individual or related individuals. All breeding values are independent of all environmental deviations. These assumptions are made for simplicity and may not hold when common environmental effects, non-additive genetic effects or major genes exist.

The expected variances and covariances of the three traits can be decomposed into the relative contribution of environmental deviations and breeding values, respectively,

$$\begin{bmatrix} \sigma_{G\omega}^2 & \mathbf{s}'_G \\ \mathbf{s}_G & \mathbf{G} \end{bmatrix} = \begin{bmatrix} \sigma_{G\omega,y}^2 & \sigma_{G\omega,y} & \sigma_{G\omega,x} \\ \sigma_{Gy,\omega} & \sigma_{Gy}^2 & \sigma_{Gy,x} \\ \sigma_{Gx,\omega} & \sigma_{Gx,y} & \sigma_{Gx}^2 \end{bmatrix}, \quad (2.2)$$

$$\begin{bmatrix} \sigma_{E\omega}^2 & \mathbf{s}'_E \\ \mathbf{s}_E & \mathbf{E} \end{bmatrix} = \begin{bmatrix} \sigma_{E\omega,y}^2 & \sigma_{E\omega,y} & \sigma_{E\omega,x} \\ \sigma_{Ey,\omega} & \sigma_{Ey}^2 & \sigma_{Ey,x} \\ \sigma_{Ex,\omega} & \sigma_{Ex,y} & \sigma_{Ex}^2 \end{bmatrix}.$$

(Co)variances involving fitness have been included in the matrices to emphasize the common form of the parameters that need to be estimated. However, the standard convention will be followed and the vector of these covariances (selection differentials) will be denoted by \mathbf{s} , or s for individual elements. Vectors and scalars will be subscripted with the trait(s) to which they correspond (x , y and ω) preceded by either G or E if they pertain to genetic or environmental (co)variances, respectively. Phenotypic (co)variances will simply be subscripted by the trait, and are simply the sum of the respective genetic and environmental (co)variances.

The matrix elements in equation (2.2) can be used to define some important quantities and relations from quantitative genetics (table 1). Original references, when they exist, are given in table 1, together with useful introductory texts and previously used notation. An important distinction is between selection differentials and selection gradients. Whereas selection differentials measure the association of the traits with fitness, irrespective of whether that association is causal or not, selection gradients (β), on the other hand, capture the causal effect of a trait on fitness. In this article, we are mainly concerned with selection on the focal trait y , although we need to consider additional traits that are also correlated with y but also directly affect fitness. These 'nuisance' traits compromise our ability to measure direct selection operating on y . In reality, there are likely to be several nuisance traits, although for our purposes we can work with a single hypothetical nuisance trait x . To do this, we need to define x in a precise way. It is defined such that $\beta_{Gy} = \beta_{Ey}$ in the regression (Rausher 1992)

$$\begin{aligned} \omega &= \bar{\omega} + \beta_{Gy}g_y + \beta_{Ey}e_y + x + \epsilon_\omega, \\ &= \bar{\omega} + \beta_y(g_y + e_y) + x + \epsilon_\omega, \\ &= \bar{\omega} + \beta_y y + x + \epsilon_\omega, \end{aligned} \quad (2.3)$$

Table 1. Some important quantities and relations, which hold when the standard assumptions of quantitative genetics are met. ($\Delta\bar{y}$ represents the change in the population mean of y over a generation.)

parameter	previously used notation	notation
(phenotypic) selection differential; Pearson (1903), Lande (1979), Lande & Arnold (1983), Falconer (1983) and Phillips & Arnold (1989)	s	$s_y = s_{Gy} + s_{Ey}$
genetic selection differential; Robertson (1966), Rausher (1992) and Rice (2004a)	\mathcal{G}	$s_{Gy} = \Delta\bar{y}$
heritability; Falconer (1983) and Lynch & Walsh (1998)	h^2	$h^2 = \sigma_{Gy}^2 / \sigma_y^2$
(phenotypic) selection gradient; Pearson (1903), Lande (1979), Lande & Arnold (1983) and Phillips & Arnold (1989)	β	$\begin{bmatrix} \beta_y \\ \beta_x \end{bmatrix} = \mathbf{P}^{-1} \mathbf{s}$
genetic selection gradient; Rausher (1992)	\mathbf{B}	$\begin{bmatrix} \beta_{Gy} \\ \beta_{Gx} \end{bmatrix} = \mathbf{G}^{-1} \mathbf{s}_G$
(multivariate) breeders equation; Pearson (1903), Lande (1979), Falconer (1983), Lande & Arnold (1983) and Phillips & Arnold (1989)	$\mathbf{GP}^{-1} \mathbf{s}$	$\mathbf{GP}^{-1} \mathbf{s} = \begin{bmatrix} \Delta\bar{y} \\ \Delta\bar{x} \end{bmatrix}$

where x is a weighted sum of traits that contains the minimum amount of information required to correctly estimate the causal relationship between the focal trait and fitness (Rausher 1992). If the causal relationship can be correctly estimated from y alone, then x does not exist due to the additional constraint that $s_y = 0$ (see appendix A in the electronic supplementary material). This is for mathematical convenience only. When x does exist, the traits that make up x are traits in the loosest sense; they can include not only the phenotypes of individuals, but also the environmental variables that those individuals experience. Biologically, the distinction is important (Mitchell-Olds & Shaw 1987), but statistically their effect is identical. Defining x in such a way allows us to condense all relevant information into a single vector. From a practical perspective, it is enough to measure the set of traits (or environments) that have non-zero weights. Appendix A in the electronic supplementary material derives some properties of x which may be useful.

When x is defined in this way, then the selection gradients can be estimated correctly and the multivariate breeders equation can give accurate predictions of evolutionary change when the standard assumptions of quantitative genetics are met (table 1). However, those terms involving x also have univariate analogues which can be derived by treating x as a missing trait and setting (co)variances involving x to zero. These analogues which will be denoted by a tilde are the univariate breeders equation ($h^2 s_y = \tilde{\Delta}\bar{y}$), and the selection gradients estimated ignoring x (e.g. $s_y / \sigma_y^2 = \tilde{\beta}_y$). As a side note, it has been suggested that the univariate breeders equation can only predict evolutionary change ($\tilde{\Delta}\bar{y} = \Delta\bar{y}$) when the causal effects of selection have been measured correctly ($\beta_{Gy} = \beta_y$) (Rausher 1992). In appendix A in the electronic supplementary material, it is proved that this is not strictly true, and that the breeders equation can accurately predict evolutionary change if \mathbf{E} and \mathbf{G} matrices for all relevant traits are proportional (i.e. $\mathbf{E} = a\mathbf{G}$, where a is any constant). However, this possibility seems biologically unlikely (Willis *et al.* 1991; Hadfield *et al.* 2007). For a full discussion on the properties of the breeders equation, especially when the assumptions of quantitative genetics are not met, see Heywood (2005).

In statistical quantitative genetics, it is often assumed that the breeding values and environmental deviations follow a multivariate normal distribution, and so the matrices in equation (2.2) and the associated vector of trait means characterize their distribution completely. It should be stressed that for *some* parameters in some theoretical models, it is irrelevant whether this assumption is met or not. In such cases, it is the actual value of the (co)variances that matter, not the ability of those (co)variances to summarize a distribution. Nevertheless, this assumption is usually important when making statistical inferences about these (co)variances estimated from data (Mitchell-Olds & Shaw 1987). In certain sections of this paper, it will be necessary to relax these distributional assumptions in order to fit a more realistic statistical model. For those parameters where it is the actual value of the (co)variance that matters, it will generally be possible to derive estimates of these (co)variances as some function of the model parameters (Janzen & Stern 1998). For those parameters where it is important that multivariate normality holds, it may not be possible to directly apply the standard equations of quantitative genetics (Gimelfarb & Willis 1994). This should no longer be seen as a difficulty; in these instances, Rice (2000, 2002, 2004a,b) provides an alternative framework which can be applied to real data using Markov chain Monte Carlo (MCMC) techniques.

3. MISSING DATA THEORY

To evaluate the evolutionary parameters in table 1, the (co)variances in equation (2.2) must be estimated from the collected data. In the most general case, the data consist of pedigree information (which we will assume can be completely summarized by the additive genetic relatedness matrix \mathbf{A}) and the potential measurements that could be taken for the three traits: ω ; \mathbf{y} ; and \mathbf{x} , which will be jointly referred to as \mathbf{D} . These vectors are potential in the sense that some measurements may be absent. Here, those components that are observed are indexed with 'o' and those that are missing with 'm' (e.g. $\mathbf{y}' = [\mathbf{y}'_o \ \mathbf{y}'_m]$). It is assumed that \mathbf{A} is always observed. An important idea when dealing with missing data is that the pattern of missingness should be treated

as part of the data. These data will be represented by $\mathbf{M} = [\mathbf{m}_w, \mathbf{m}_y, \mathbf{m}_x]$ which has elements equal to one when the data are observed and zero when the data are missing. When dealing with missing data we also need to consider a new set of parameters (ϕ) which describe the distribution of missingness. However, these parameters are not of primary interest. The main aim of the analysis is to extract the information contained in the observed data regarding the evolutionary parameters (θ).

The likelihood of the observed data given the parameters can be expressed as

$$\Pr(\mathbf{D}_o, \mathbf{A}, \mathbf{M} | \theta), \quad (3.1)$$

although the probability model in this form is often hard to work with and it is easier to consider it in the form

$$\Pr(\mathbf{D}_o, \mathbf{A}, \mathbf{M} | \theta) = \int \int \Pr(\mathbf{D}, \mathbf{A} | \theta) \Pr(\mathbf{M} | \mathbf{D}, \mathbf{A}, \phi) d\mathbf{D}_m d\phi. \quad (3.2)$$

This splits the problem into two parts: first, we consider the likelihood of the full potential data given the evolutionary parameters and, second, we consider the probability of missingness given the full potential data and the parameters of the missing data model. However, we are not interested in the likelihood of hypothetical data \mathbf{D} , but in the likelihood of the data that we observe \mathbf{D}_o . The difficulty then becomes specifying and integrating over the distribution of the missing data \mathbf{D}_m .

However, this integration can be greatly simplified when certain assumptions for the model of the missing data mechanism are met. The most stringent of these assumptions is that there is no association between the data (observed or missing) and missingness; the data are ‘missing completely at random’ (MCAR),

$$\Pr(\mathbf{M} | \mathbf{D}, \mathbf{A}, \phi) = \Pr(\mathbf{M} | \phi). \quad (3.3)$$

Clearly, this is a strong assumption. A weaker assumption is that any association between missingness and the traits is mediated by the observed data and does not depend on the values of the missing data. In this scenario, any systematic pattern of missingness can be predicted from the observed data, and the data are said to be ‘missing at random’ (MAR),

$$\Pr(\mathbf{M} | \mathbf{D}, \mathbf{A}, \phi) = \Pr(\mathbf{M} | \mathbf{D}_o, \mathbf{A}, \phi). \quad (3.4)$$

A biological example will make these concepts more concrete, as the use of the term random is rather non-intuitive. Imagine that hatching order (x) is measured for a set of nestling birds, and that these data are completely observed. Some days later, the weight (y) of all surviving nestlings is measured, with the weight of nestlings that died in the intervening period being classed as missing data: the observed data, $\mathbf{D}_o = \{x, y_o\}$ and the missing data, $\mathbf{D}_m = y_m$. We will assume that nestlings which hatch early tend to weigh more. If we imagine the sole cause of mortality is hatching order, then the missing weights are said to be MAR, as missingness only depends on observed data. This does not imply that the observed weights are a random sample of all potential weights. They will not be; if chicks born early survive better, then the surviving nestlings would have been heavier than those that died, and the missing weights could not be classed as MCAR. An alternative scenario would be that weight does affect survival even after accounting for the effects of hatching

order. In this case, it is not possible to simplify the missing data model and the missing weights are said to be ‘not missing at random’ (NMAR).

When MAR or MCAR holds then integration over the missing data is simplified since

$$\begin{aligned} \Pr(\mathbf{D}_o, \mathbf{A}, \mathbf{M} | \theta) &= \int \int \Pr(\mathbf{D}, \mathbf{A} | \theta) \Pr(\mathbf{M} | \mathbf{D}, \mathbf{A}, \phi) d\mathbf{D}_m d\phi, \\ &= \int \Pr(\mathbf{D}, \mathbf{A} | \theta) d\mathbf{D}_m \int \Pr(\mathbf{M} | \mathbf{D}_o, \mathbf{A}, \phi) d\phi, \\ &= \Pr(\mathbf{D}_o, \mathbf{A} | \theta) \int \Pr(\mathbf{M} | \mathbf{D}_o, \mathbf{A}, \phi) d\phi. \end{aligned} \quad (3.5)$$

When the parameters ϕ and θ are ‘distinct’, the two models become independent and the model of the missing data process is said to be ‘ignorable’ with respect to θ ,

$$\Pr(\mathbf{D}_o, \mathbf{A}, \mathbf{M} | \theta) \propto \Pr(\mathbf{D}_o, \mathbf{A} | \theta). \quad (3.6)$$

In a Bayesian framework, the parameters ϕ and θ also have to be independent in their prior distribution ($\Pr(\phi, \theta) = \Pr(\phi)\Pr(\theta)$). These results are derived using probability theory, and therefore do not extend to methods of inference that do not explicitly use likelihood. Ignorability cannot be ensured for moment-based methods of inference such as ANOVA even when the data are MAR and the parameters are distinct. For these types of estimators, the weakest condition for ignorability is generally MCAR.

It is important to realize that ignorability of the missing data mechanism does not imply that the analysis can be restricted to complete cases (those individuals that have been measured for all the traits). It is paramount that the likelihood is based on all observed data and that the correct likelihood is used (Gianola *et al.* 1989). When recourse is made to the results of missing data theory, it is necessary to ensure that the software retains records of individuals that have some missing data, and that these missing values are dealt with correctly.

4. MISSING DATA THEORY AND VIABILITY SELECTION

This section considers a trait y that is expressed at birth, and once expressed, does not continue to develop. Conditions for MCAR and MAR are generally less strict for these types of trait than for traits expressed later in the life cycle, or traits that continue to develop during ontogeny. Sections 6 and 7 consider the problems associated with these types of traits in more detail.

For the purposes of this paper, I will assume that any covariance that exists between the trait and fitness exists owing to viability selection. This is not because fecundity and sexual selection are less important (Hoekstra *et al.* 2001), but because these forms of selection are less likely to result in missing data problems. The association between the trait and viability may be causal, incidental or both. However, any association between the nuisance trait x and viability is causal by our definition of x . For simplicity, I assume that x is either completely observed or completely missing (a missing trait). On the other hand, y may be partly observed, but I assume that all individuals alive at the time of measurement have equal probability of being measured. Mark-recapture methods can be used to relax this assumption (Lebreton *et al.* 1992).

Table 2. For different types of data, the patterns of selection that result in different patterns of missingness (MCAR, MAR and NMAR) are summarized. (Missing data theory is non-parametric in the sense that it is derived using basic probability theory. Given this, these results are general to any form of viability selection regardless of whether it is linear or nonlinear. x being observed is equivalent to x not existing.)

population	x	MCAR	MAR	NMAR
unrelated	observed	$s_y = 0$	$\beta_y = 0$	$\beta_y \neq 0$
unrelated	missing	$s_y = 0$	$s_y = 0$	$s_y \neq 0$
related	observed	$s_y = 0$ and $s_{Gy} = 0$	$\beta_y = 0$	$\beta_y \neq 0$
related	missing	$s_y = 0$ and $s_{Gy} = 0 = 0$	$s_y = 0$ and $s_{Gy} = 0$	$s_y \neq 0$ or $s_{Gy} \neq 0$

The sole cause of missingness in this case is lifespan, suggesting that if the data included lifespan then any missing data would be MAR and we could proceed by ignoring the missing data mechanism. However, the time of death is usually not known exactly but is only known to have occurred between two time points. A useful concept discussed by Little & Rubin (2002; §6) is that censored data such as these should be considered as missing data, and that this principle also extends to random effects. I will proceed as if any lifespan data is censored.

When y is measured at birth, it is either completely observed or missing values are MCAR. However, if time elapses before measurements can be made, then any pattern of missing data can arise depending on the nature of selection and the type of data that are collected (table 1). Datasets that include the fully observed nuisance trait x (if it exists) are distinguished from those that do not. In addition, data which are collected on individuals that are believed to be related are distinguished from those where individuals are believed to be unrelated. This distinction is necessary because the probability of a focal individual dying prior to measurement is not associated with the phenotype (or breeding value) of any other individual when individuals are unrelated. When individuals are related, however, the phenotype of a relative may be associated with the probability of missingness in the focal individual if there is genetic covariation between the trait and survival. The missing data mechanism is then no longer purely phenotypic. For different types of dataset, table 2 summarizes the conditions under which the three missing data mechanisms apply.

The conditions given in table 2 can be derived using four basic rules.

- (i) Missing values of y are MCAR when $s_y = 0$ and the data consist of y only. In this case, only phenotypic parameters can be estimated and it is assumed that the sampled individuals are unrelated $A = I$.
- (ii) When the sample of individuals is related, missing values of y are only MCAR when $s_y = 0$ and $s_{Gy} = 0$. When this condition is met, then x by our definition does not exist.
- (iii) When x is measured, missing values of y are MAR if $\beta_y = 0$. In this case, missing values of y are MAR even when the individuals are related.
- (iv) If x exists and is not measured, then missing values of y are NMAR.

These basic rules can be used to determine when valid inferences can be drawn without explicitly modelling the missing data mechanism. If y cannot be measured at birth and x is unobserved, then the selection gradient is

unestimable and the estimates of the selection differential are unbiased only when the selection differential is zero. If the sampled individuals are related, then the genetic selection differential must also be zero. If y cannot be measured at birth but x is observed or does not exist, then estimates of the selection gradient and selection differential are unbiased only when the selection gradient is zero. This holds for samples of related individuals.

These results can be placed in the context of the earlier example, where the focus was on viability selection operating on the weight of nestling birds (y) and the nuisance trait was hatching order (x). We will assume that hatching order has no heritable basis but that weight does. If we were unable to measure hatching order, then any estimate of the direct selection acting on weight (the univariate selection gradient $\tilde{\beta}_y$) would be biased by the shared effects that hatching order has on both weight and survival. However, this does not necessarily preclude us from measuring the selection differentials accurately. In the previous example, hatching early resulted in an increase in body mass and survival probability. However, if, after conditioning on hatching order, large chicks actually had a greater risk of dying, then it would be possible for these two processes to cancel, resulting in no observable relationship between weight and survival (the selection differential would be zero). When the individuals are unrelated, then any weights would be MCAR, but we can only measure phenotypic patterns. However, if the chicks were related, then weights would not be MCAR because individuals that belong to families subjected to high mortality will tend to be larger than expected from the observed data. Even if hatching order had been measured, then this pattern would still hold unless there was no direct relationship between weight and fitness ($\beta_y = 0$). If there was no direct relationship, then the missing weights would be MAR. From §§6 and 7, it will be apparent that for traits not expressed at birth these conditions will be even stricter. For developing traits, in particular, plausible models of missingness are always NMAR.

5. MAR AND NMAR MODELS BASED ON SURVIVAL ANALYSIS

As was noted in the preceding sections, when the probability of being measured depends on being alive at the time of measurement, a simple solution would be to include lifespan in the analysis. With the inclusion of lifespan, the missing data mechanism would then be ignorable. However, lifespan is rarely measured without error, as it is usually censored in the sense that we can only say that an individual died after a certain censoring point

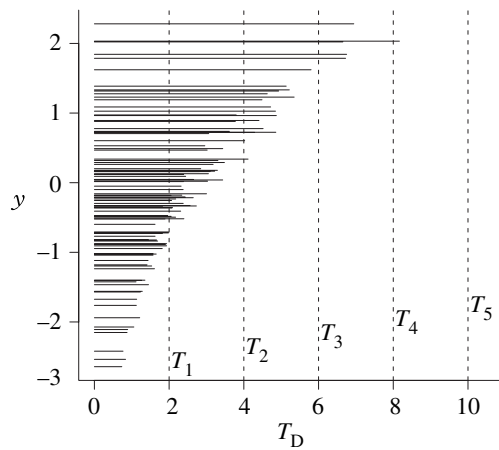


Figure 1. Hypothetical sampling scheme. The horizontal solid lines represent the period of time in which each individual was alive. This is related to the trait y , which is expressed from birth but only measured on day 2 (censoring point T_1). In the ideal case, the length of these lines is known exactly (lifespan data are uncensored), but a more reasonable scenario is that the length of the line is only known to fall between censoring points (vertical dashed lines).

or between censoring points. This section shows that a model including lifespan is MAR and gives correct inference when the correct likelihood is used. Furthermore, it also shows that a model which treats censored lifespans as if they were uncensored is NMAR and results in bias. Finally, a model is derived based on survival analysis that explicitly accounts for the missing data process and gives unbiased estimates. To illustrate the different approaches, use is made of simulated data generated according to the following process.

One hundred (n) unrelated individuals are followed from birth until they die, the timing of which is recorded. The time of death is either recorded exactly, or individuals are simply recorded as being dead or alive 2 days after birth, 4 days after birth and so on in 2-day intervals. Individuals express a trait y from birth, but it is only measured on the second day. Individuals that die before day 2 therefore have missing values for y , which is indicated as 1 in the vector that indicates missingness \mathbf{m} . The sampling schemes are represented in figure 1.

Lifespan is lognormally distributed and depends log-linearly on y which is normally distributed. Although biologically unrealistic, it is assumed that generations are discrete and that offspring production occurs at a constant rate per unit time across all individuals. The model is not intended to be realistic, it is intended to highlight the logic of the problem and a possible solution. It will be easier to work with log lifespan, which is denoted by l . The aim is to estimate the selection differential s_y , which is a function of the (co)variances for l and y (\mathbf{P}) and the trait means \bar{l} and \bar{y} . Here, we will concentrate on the regression coefficient of y on l , which is denoted by β_y .

When lifespan is known exactly, the likelihood is given by

$$\begin{aligned} \Pr(\mathbf{y}_o, \mathbf{l}, \mathbf{m} | \mathbf{P}) &\propto \Pr(\mathbf{y}_o, \mathbf{l} | \mathbf{P}), \\ &\propto \prod_{i=1}^n (1 - m_i) N(l_i | \bar{l}, \sigma_l^2) \\ &\quad + m_i N\left(\begin{bmatrix} \mathbf{y}_i \\ l_i \end{bmatrix} \middle| \begin{bmatrix} \bar{y} \\ \bar{l} \end{bmatrix}, \mathbf{P}\right), \end{aligned} \tag{5.1}$$

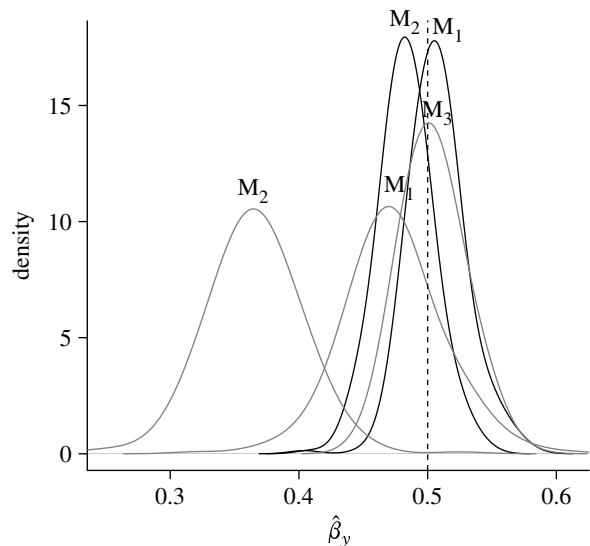


Figure 2. Distribution of ML estimators and posterior medians of β_y for five types of model. M refers to the type of model: 1, MAR; 2, complete case; 3, NMAR. The models were fitted to lifespan data that were known without error (black lines) or lifespan data that were censored (grey lines). The dashed line is the true underlying value.

where $N(y|\mu, \sigma^2)$ is the normal density function with specified mean μ and variance σ^2 . Importantly, this likelihood includes the probability of lifespans for those individuals that died prior to being measured (the first term on the right hand side). This is referred to as the MAR model. An alternative formulation would be to completely ignore those individuals that die prior to measurement, which is called the complete case model,

$$\prod_{i=1}^n m_i N\left(\begin{bmatrix} \mathbf{y}_i \\ l_i \end{bmatrix} \middle| \begin{bmatrix} \bar{y} \\ \bar{l} \end{bmatrix}, \mathbf{P}\right), \tag{5.2}$$

β_y was estimated using both models for 250 simulated datasets, with β_y having an expectation of 0.5. In addition, the likelihoods were maximized using the mid-point between the censoring points prior to death l^c and after death l^d as a replacement for actual lifespan. Maximum-likelihood (ML) estimates are known to give downwardly biased estimates of variances (Patterson & Thompson 1971), so parameter estimates are also given using the equivalent REML analysis. As was anticipated using missing data theory, the MAR model gives valid inference when lifespan is known (ML: $\hat{\beta}_y = 0.504 \pm 0.001$, $p = 0.006^{**}$; REML: $\hat{\beta}_y = 0.500 \pm 0.001$, $p = 0.31$), where the complete case model results in a downward bias (ML: $\hat{\beta}_y = 0.483 \pm 0.001$, $p < 0.001^{**}$; REML: $\hat{\beta}_y = 0.481 \pm 0.001$, $p < 0.001^{**}$). Applying these likelihoods to censored data also results in a downward bias (ML: $\hat{\beta}_y = 0.473 \pm 0.002$, $p < 0.001^{**}$; REML: $\hat{\beta}_y = 0.476 \pm 0.002$, $p < 0.001^{**}$), particularly in the complete case model (ML: $\hat{\beta}_y = 0.366 \pm 0.002$, $p < 0.001^{**}$; REML: $\hat{\beta}_y = 0.364 \pm 0.002$, $p < 0.001^{**}$). The p -values refer to the probability that the mean of the distribution of (RE)ML estimators differs from the true value (figure 2).

Fitting the correct NMAR model to censored data is a bit more involved, and the mechanics are probably best understood by breaking the problem into several conditional distributions, one of which is the conditional distribution of the missing values. These conditional

distributions are sampled from standard MCMC techniques (Gelman *et al.* 2004). Of particular interest are the conditional distributions of the regression slope,

$$\Pr(\beta_y | \mathbf{y}, \mathbf{I}^a, \mathbf{I}^d, \sigma_y^2) = \prod_{i=1}^n [\Phi(\mathbf{I}_i^d | \bar{l} + \mathbf{y}_i \beta_y, \sigma_y^2) - \Phi(\mathbf{I}_i^a | \bar{l} + \mathbf{y}_i \beta_y, \sigma_y^2)] \Pr(\beta_y), \quad (5.3)$$

and the missing values,

$$\Pr(\mathbf{y}_m | \mathbf{I}^a, \mathbf{I}^d, \bar{l}, \bar{y}, \mathbf{P}) = [\Phi(\mathbf{I}_i^d | \bar{l} + \mathbf{y}_m \beta_y, \sigma_y^2) - \Phi(\mathbf{I}_i^a | \bar{l} + \mathbf{y}_m \beta_y, \sigma_y^2)] N(\mathbf{y}_m | \bar{y}, \sigma_y^2), \quad (5.4)$$

where $\Phi(y | \mu, \sigma^2)$ is the normal cumulative distribution function with specified mean μ and variance σ^2 .

The conditional likelihood for the regression coefficient is the likelihood of the censored lifetime data conditional on the complete data \mathbf{y} . It is essentially the likelihood that individuals die between two censoring points given their phenotype and the lifespan model. Lawless (2003) gives a clear overview of likelihoods involving censored data. When the missing values of y are MCAR, they have conditional distribution $N(\mathbf{y}_m | \bar{y}, \sigma_y^2)$, which is the last term on the right hand side of equation (5.4). Since the data are not MCAR, the normal density function needs to be 'corrected' by the information contained in the (censored) lifetime data. Fitting this model to the 250 datasets gives unbiased estimates ($\hat{\beta}_y = 0.503 \pm 0.002$, $p = 0.206$).

In conclusion, when the probability of missingness depends on lifespan, the missing data mechanism can be made ignorable by including lifespan in the analysis. When lifespan data are censored, however, the missing 'information' is still NMAR and it is necessary to model the missingness explicitly. When this is not taken into account, the force of viability selection operating on a trait is always underestimated. Extensions to standard survival analyses are a natural choice for modelling the missing data mechanism, as they provide a (semi)parametric model that can be extrapolated over the unsampled time points. The Bayesian model presented in this paper is only applicable to data collected on unrelated individuals. When data are available on related individuals, the analysis is complicated by the fact that the association between lifespan and the trait may be different at the genetic and environmental level. For general pedigrees, this would require fitting a bivariate animal model for censored traits, which has recently been developed in an MCMC framework (Damgaard & Korsgaard 2006).

6. NMAR MODELS FOR DEVELOPING TRAITS

Generally, in quantitative genetic studies of developing traits, measurements are taken repeatedly on a set of related individuals. There are two main approaches to modelling the data. Each age-specific measurement can be treated as separate trait in a multivariate analysis (Lynch & Arnold 1988) or the phenotype can be modelled as some function of time, with the 'phenotypic' (co)variance of the function's parameters being decomposed into \mathbf{G} and \mathbf{E} . Commonly used functions are polynomials (Kirkpatrick *et al.* 1990) and nonlinear functions in the

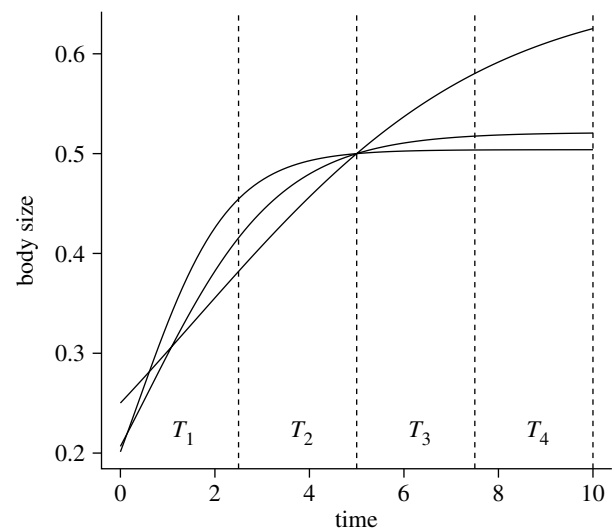


Figure 3. Logistic growth curves for three individuals, with four measurement times indicated by vertical dashed lines. When selection is acting consistently and continuously on body size, the instantaneous rate of survival is proportional to the logistic growth curves. Under a conditional MAR process, the probability that the three individuals pass through time period 3 should be constant, as they all have the same body size when entering that time period. Clearly, the individuals that would be larger at the fourth measurement, and this information cannot be obtained from the data observed at the third measurement. Missing information is NMAR.

family of Richards' (1959) growth curves, but alternative covariance functions exist (Pletcher & Geyer 1999). Section 7 applies to all these models but I will consider a model where body size is to be modelled as a logistic function of time. Several measurements are taken for each individual at set time points, but due to death later measurements are missing for some individuals (figure 3). Death is not random but is a consequence of viability selection directly either on the trait value as it is expressed or on the parameters of the logistic function (such as growth rate). Using ideas presented by Little (1995), it is shown why either process results in a non-ignorable missing data mechanism.

When selection is acting consistently and continuously on body size, we can imagine that the instantaneous rate of survival is proportional to body size at that time. Therefore, the three logistic curves in figure 3 represent not only changes in body size but also changes in the probability of survival. Under an MAR process, we need to anticipate these changes in survival probability from the observed data, which is generally not possible since there will always be some error in estimating the trajectory of body size between measurement times. Consider the common but often unstated approximation that the survival probability across a time period depends on body size as that period is entered (i.e. the data are MAR). Under this assumption, the probability of all three individuals passing through time period 3 should be constant, although this is clearly not the case if survival depends directly on body size as it is expressed. The same arguments can be made for a selective process that depends not directly on body size but on one of the underlying parameters of the logistic growth curve. Imagine that survival probability depends on potential

asymptotic size and that individuals grow logistically without any developmental noise (or measurement error). If an individual is measured thrice, then the asymptote is known with certainty even if that individual died before it reached an asymptote. There is essentially no missing data. However, imagine that an individual was only measured at the first censoring point, leaving uncertainty regarding its asymptote. Under an MAR process, knowing whether failure to be measured was due to death or some logistical hitch should provide no additional information regarding that individual's asymptote. When selection is operating on asymptotic size, then there is additional information, however, and missingness is again NMAR. If the individual had not been measured because it had died, then we should expect its asymptote to be lower when under positive directional selection than if it had not been measured for other reasons.

Together, these results imply that if selection operates on an ontogenetic process such as growth rate, or if selection is acting continuously on the outcome of an ontogenetic process such as body size, then the missing data mechanism is never ignorable (Little 1995). In one of the few attempts to deal with this problem, Lynch & Arnold (1988) suggested that age-specific phenotypic (co)variances in the absence of selection could be recovered by correcting for the pattern of selection observed in the data. However, the correction should be seen as a way of recovering the expected phenotypic covariance under an assumed MAR process, which is only necessary because the pattern of selection is estimated using a variance-type estimator that assumes MCAR. The correction will of course reduce bias in the estimates, but is largely redundant with the advent of multivariate REML.

In order to correctly estimate all relevant parameters, it is necessary to model the drop-out mechanism explicitly through joint analysis of the longitudinal data and lifespan (Vonesh *et al.* 2006). In the medical sciences, these types of model have been used extensively for some years and good overviews can be found in Little (1995) and more recently in Tsiatis & Davidian (2004). Usually, these methods specify two models: one for the trait (body size) conditional on the random effects (in this case the individual's growth parameters) and another for lifespan conditional on the random effects. In most cases, these random effects have been individual-specific intercepts and slopes, which act as covariates in a generalized linear regression for lifespan (Wu & Carroll 1988; Faucett & Thomas 1996; Schluchter *et al.* 2001). However, applying these ideas to nonlinear growth models, or models in which pedigree information is incorporated, should be a relatively straightforward extension of existing Bayesian models in quantitative genetics (Blasco & Piles 2003; Damgaard & Korsgaard 2006).

7. NMAR MODELS WHEN INDIVIDUALS DIE BEFORE TRAIT EXPRESSION

For traits expressed at birth, it was implicitly assumed that the association of y with survival remains constant with age. This is often unreasonable, particularly with traits that are not expressed from birth. For these traits, the causal effect of a trait prior to expression must be equal to zero. However, this does not imply that the phenotypes of individuals which express the trait are a random sample

of the possible phenotypes that would have been expressed in the absence of selection: the potential trait value may be genetically and/or environmentally correlated with traits that are expressed earlier in ontogeny. This is the invisible fraction at its most awkward (Grafen 1988), and it seems unlikely that we can understand the evolution of these traits without considering the possibility that individuals which survive until trait expression are a non-random sample. Empirically, this is a difficult problem, and currently the main way to address this problem is through artificial selection experiments. Indeed, several studies have shown that artificial selection on a trait has induced a correlated response in a trait expressed earlier in ontogeny (Zwinn 1999).

In unmanipulated systems, there seems to be no general solution to this problem. Imagine that we are interested in estimating the association between a reproductive trait and survival prior to reproduction. The problem is obvious; any variation in pre-reproductive survival is not manifest in those individuals that we can measure reproductive traits in because they all must have survived the pre-reproductive period. On the surface, it seems that there is no information in the data regarding this association, and that we are therefore forced to make some assumptions about its form. The two most common assumptions are that any association between survival and a trait is non-existent prior to trait expression, or that the association observed post-expression also holds pre-expression. Both these assumptions seem unreasonable. Artificial selection experiments suggest that the former is not always true, and given that a trait can only have a causal effect on fitness once it is expressed suggests that the latter is also unlikely to hold. Although not generally appreciated, these assumptions can be relaxed a little when data are collected on related individuals.

When individuals are related and genetic variance exists, then we have some information regarding the phenotypes of individuals that have not been measured. The genetic covariance between pre-reproductive survival and reproductive traits can therefore be estimated in the same way that genetic correlations can be estimated between traits that are expressed in different sexes. Any residual environmental covariance will of course be unestimable, meaning that the phenotypic association between a trait and pre-trait survival cannot be estimated without some assumptions regarding the environmental covariance. Earlier, we were forced into assuming that this covariance is either zero or equal to the covariance measured post-trait expression. Perhaps a more reasonable assumption would be that the environmental correlation is equal to the estimated genetic correlation. Either way, the genetic component of the invisible fraction is at least partially visible, and for those datasets that have sufficient power it would be very interesting to see how different visible and invisible fractions really are.

8. MISSING TRAITS AND PBVS

The nuisance trait x is a weighted sum of traits. The traits and their weights are unlikely to be known *a priori*, but the weighted sum can be estimated if all the traits with non-zero weights have been measured. This is exactly what the multiple regression model of Lande & Arnold (1983) aims to do. However, we need some way of checking whether x

can be estimated. Since we define x such that the phenotypic and genetic selection gradients are equal, a simple test is whether $\tilde{\beta}_y = \hat{\beta}_{Gy}$ (Rausher 1992). As family sizes become larger, Rausher (1992) showed that PBVs (in his case, family means) could be used in place of true breeding values to give valid inference. Postma (2006) showed that the bias remains for general pedigrees and suggested an alternative test based on the genetic covariance between y and fitness (s_{Gy}). Again, this test involved estimating the covariance from PBVs. The number of studies that estimate either β_{Gy} or s_{Gy} using PBVs is now large (Kruuk *et al.* 2001, 2002; Merila *et al.* 2001a,b; Stinchcombe *et al.* 2002; Coltman *et al.* 2003; Sheldon *et al.* 2003; Charmantier *et al.* 2004; Garant *et al.* 2004b,a; Winn 2004; Gienapp *et al.* 2006). Below, it is shown that estimates of these quantities using PBVs are extremely biased. This bias arises because the breeding values are predicted using a model in which fitness is treated as missing. Clearly, fitness is not missing in these studies and it is suggested that an unbiased estimator of β_{Gy} or s_{Gy} is easy to estimate using REML (see Kruuk *et al.* 2002), Bayesian and, in some cases, moment-based procedures.

Here, we will concentrate on the difference between the actual genetic covariance s_{Gy} and the expected covariance between fitness and PBVs for y ($E[\text{cov}(\omega, \hat{\mathbf{g}}_y)]$). This genetic covariance is central to testing $\tilde{\beta}_y = \hat{\beta}_{Gy}$ when it is assumed that x does not exist (Postma 2006). The full derivation can be found in appendix B in the electronic supplementary material, but the key result is

$$E[\text{cov}(\omega, \hat{\mathbf{g}}_y)] = s_{Gy} E \left[\lambda \frac{\lambda - 1 + \frac{s_y}{s_{Gy}}}{\lambda - 1 + \frac{\sigma_y^2}{\sigma_{Gy}^2}} \right], \quad (8.1)$$

where the expectation is taken over λ , the eigenvalues of the additive genetic relationship matrix (\mathbf{A}). $E[\lambda]$ is equal to the mean inbreeding coefficient in the population, so in the absence of inbreeding, PBVs can yield unbiased estimates but only when $s_y/s_{Gy} = \sigma_y^2/\sigma_{Gy}^2$. These are the same conditions for which the breeders equation is predictive ($\tilde{\Delta y} = \Delta y$), which includes the condition to be tested ($\tilde{\beta}_y = \hat{\beta}_{Gy}$) (see appendix B in the electronic supplementary material). When this condition is not met, then any estimate is biased towards this equality. However, as the amount of information in the data regarding an individual's breeding value increases, the variance in eigenvalues increases and this bias drops. The distribution of eigenvalues depends critically on the distribution of relatives in the population. For example, large numbers of measured offspring may allow an individual's breeding value to be predicted accurately, whereas a large number of siblings would not, as the Mendelian sampling variation and residual variation are confounded. To gauge the magnitude of the problem, the expectation on the right hand side of equation (8.1) has been calculated using the eigenvalues obtained from the complete red deer (*Cervus elaphus*) pedigree of Rum (Nussey *et al.* 2008) and also a social pedigree from a population of blue tits (*Parus caeruleus*; Hadfield *et al.* 2006). It is assumed that all individuals are measured for both fitness and y , and also include the expected bias that would arise when no pedigree information is available

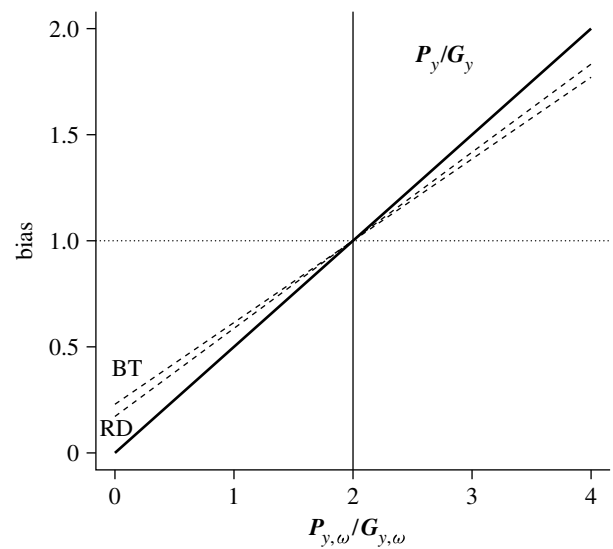


Figure 4. The expected bias (the expectation in equation (8.1)) as a function of the ratio between phenotypic and genetic selection differentials. The expected bias was calculated for a trait with a heritability of 0.5 (vertical solid line). The solid diagonal line is the degree of bias when there is no information in the data regarding s_{Gy} (all individuals are unrelated). The dashed diagonal lines represent the degree of bias for a blue tit (BT) and a red deer (RD) pedigree. The horizontal dashed line represents unbiasedness, which is close to the intersection of the two solid lines given the low levels of inbreeding.

($\mathbf{A} = \mathbf{I}$). The results are shown in figure 4 and demonstrate that the bias (the expectation on the right hand side of equation (8.1)) is severe. Measuring the genetic selection gradient ($\tilde{\beta}_{Gy} = s_{Gy}/\sigma_{Gy}^2$) using PBVs is even more troublesome (Postma 2006), as the variance is underestimated by the predictive error variance (Henderson 1975). Moreover, if the heritabilities of the trait and fitness are not equal, then information can pass between the traits (Mrode 1996), suggesting that estimates using PBVs lack efficiency even if $s_y/s_{Gy} = \sigma_y^2/\sigma_{Gy}^2$. This may not always be apparent because most studies do not account for error within, and dependency between, PBVs, which may give rise to anti-conservative measures of confidence.

In conclusion, estimating selection parameters on the genetic and environmental components of phenotype should be done explicitly in a single bivariate analysis of the trait and fitness, rather than trying to obtain them *post hoc* using PBVs. However, although a single bivariate analysis allows a more accurate and precise test for the presence of missing traits, it does highlight the power issues that are likely to be involved; we are essentially trying to estimate the genetic covariance between two traits, one of which (fitness) is likely to have a very low heritability (Gustafsson 1986; Kruuk *et al.* 2000). Nevertheless, these types of analysis can easily be done using REML (Gilmour *et al.* 2002) and standard errors for functions of these (co)variances such as selection gradients and breeders equation predictions could be obtained using the delta method. Alternatively, exact standard errors can be obtained simply using Gibbs samples (VanTassell & VanVleck 1996). In the rare instances when it is necessary to work with breeding values directly, such as assessing evolutionary change, Bayesian procedures must be seen as

more robust alternative to best linear unbiased prediction, particularly when datasets are small (Sorensen *et al.* 1994; Blasco 2001). However, care still needs to be taken when inferring evolutionary change in the presence of mortality selection (Henderson *et al.* 1959).

9. DISCUSSION

Measuring the strength and form of natural selection is a central part of evolutionary biology. With few exceptions, statistical methods used to estimate natural selection assume that any missing data are MAR. When viability selection acts on a trait, missing data are almost always NMAR. For traits that have a relationship with viability that is constant with age, the force of viability selection will be systematically underestimated. If the relationship between the trait and viability changes with age, then even the direction of viability selection may be estimated incorrectly. To measure the force of viability selection accurately, it is necessary to model the joint distribution of lifespan and the trait of interest, taking into account the phenotypes of those individuals that die prior to trait measurement or trait expression. Since the distribution of unmeasured phenotypes is never directly observable, we need to make some assumptions about the relationship between viability and unmeasured phenotypes. In general, it is assumed that no relationship exists, irrespective of whether a relationship is observed between viability and those phenotypes that are measured. I suggest a more parsimonious model where the relationship seen in the observed data holds for the missing data, and show how standard survival analyses can be extended to do this.

When interest is on the quantitative genetics of viability selection, then the problem becomes a little more complicated as the genetic and environmental associations with viability are not necessarily equivalent in the presence of missing traits (Rauscher 1992). In such cases, separately modelling the genetic and environmental components for the phenotypes of unmeasured individuals needs to be considered. Although this adds additional complexity to model fitting, it does provide a way of testing whether the relationship seen in the observed data is consistent with the pattern of missing data. This is possible because the lifespan of individuals with missing phenotype information can be compared with the phenotypes of their relatives. This possibility allows us to address the difficult empirical problem that arises when the invisible fraction is made up of individuals that die before a particular trait is expressed (Grafen 1988). In such situations, it generally has to be assumed that the trait and viability prior to trait expression are independent. However, although a causal relationship is precluded, associations may still exist through genetic and/or environmental correlations with traits that are expressed earlier in ontogeny. Such associations may help to explain the persistent problem of why some heritable traits appear to be under consistent directional selection but fail to evolve (Merila *et al.* 2001b).

It is surprising that the problem of missing data has received so little attention given that viability selection is central to evolutionary biology. One consequence of this is that most of the existing solutions to missing data exist outside of our field, particularly in the medical sciences (Little & Rubin 2002). A concerted effort is required to

bring these techniques into evolutionary biology and develop them for our own, sometimes unique, purposes. In particular, we need to extend quantitative genetic techniques to deal with censored and missing data, and for distributions that reflect the distribution of fitness components more accurately. These techniques are likely to come from the agricultural sciences where Bayesian implementations addressing some of these issues are starting to appear (Sorensen & Gianola 2002), or from ecology where missing data are often dealt with using auxiliary variables (e.g. Clark & Bjornstad 2004; King *et al.* 2006). The flexibility of these models should allow us to dispense with the practice of using PBVs as a shortcut for proper model building. This practice has probably led us to believe that missing traits are less of a problem than they really are.

Thanks go to Loeske Kruuk, Mark Rees, Dylan Childs, Alastair Wilson, Bill Hill, Darren Obbard and two anonymous referees for their comments on this manuscript. This work was funded by a Leverhulme Prize awarded to Loeske Kruuk.

REFERENCES

- Bennington, C. C. & McGraw, J. B. 1995 Phenotypic selection in an artificial population of *Impatiens pallida*—the importance of the invisible fraction. *Evolution* **49**, 317–324. (doi:10.2307/2410342)
- Blasco, A. 2001 The Bayesian controversy in animal breeding. *J. Anim. Sci.* **79**, 2023–2046.
- Blasco, A. & Piles, M. 2003 A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. *Genet. Sel. Evol.* **35**, 21–41. (doi:10.1051/gse:2002034)
- Charmantier, A., Kruuk, L. E. B., Blondel, J. & Lambrechts, M. M. 2004 Testing for microevolution in body size in three blue tit populations. *J. Evol. Biol.* **17**, 732–743. (doi:10.1111/j.1420-9101.2004.00734.x)
- Clark, J. S. & Bjornstad, C. N. 2004 Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology* **85**, 3140–3150. (doi:10.1890/03-0520)
- Coltman, D. W., O'Donoghue, P., Jorgenson, J. T., Hogg, J. T., Strobeck, C. & Festa-Bianchet, M. 2003 Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655–658. (doi:10.1038/nature02177)
- Curnow, R. N. 1961 Estimation of repeatability and heritability from records subject to culling. *Biometrics* **17**, 553–566. (doi:10.2307/2527855)
- Damgaard, L. H. & Korsgaard, I. R. 2006 A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait. *Genet. Sel. Evol.* **38**, 45–64. (doi:10.1051/gse:200502)
- Falconer, D. S. 1983 *Introduction to quantitative genetics*. London, UK: Longman Group.
- Faucett, C. L. & Thomas, D. C. 1996 Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Stat. Med.* **15**, 1663–1685. (doi:10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1)
- Fernando, R. L. & Gianola, D. 1990 Statistical inferences in populations undergoing selection or non-random mating. In *Advances in statistical methods for genetic improvement of livestock* (eds D. Gianola & K. Hammond), pp. 437–453. Berlin, Germany: Springer.
- Garant, D., Kruuk, L. E. B., McCleery, R. H. & Sheldon, B. C. 2004a Evolution in a changing environment: a case study with great tit fledging mass. *Am. Nat.* **164**, E115–E129. (doi:10.1086/424764)

- Garant, D., Sheldon, B. C. & Gustafsson, L. 2004b Climatic and temporal effects on the expression of secondary sexual characters: genetic and environmental components. *Evolution* **58**, 634–644.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. 2004 *Bayesian data analysis. Texts in statistical science*, 2nd edn. London, UK: Chapman & Hall.
- Gianola, D., Im, S. & Fernando, R. L. 1988 Prediction of breeding value under Henderson's selection model—a revisit. *J. Dairy Sci.* **71**, 2790–2798.
- Gianola, D., Fernando, R. L., Im, S. & Foulley, J. L. 1989 Likelihood estimation of quantitative genetic parameters when selection occurs—models and problems. *Genome* **31**, 768–777.
- Gienapp, P., Postma, E. & Visser, M. E. 2006 Why breeding time has not responded to selection for earlier breeding in a songbird population. *Evolution* **60**, 2381–2388.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. 2002 *ASREML user guide release 1.0*. Hemel Hempstead, UK: VSN International Ltd.
- Gimelfarb, A. & Willis, J. H. 1994 Linearity versus nonlinearity of offspring–parent regression—an experimental study of *Drosophila melanogaster*. *Genetics* **138**, 343–352.
- Goffinet, B. 1983 Selection on selected records. *Genet. Sel. Evol.* **15**, 91–97. (doi:10.1051/gse:19830107)
- Grafen, A. 1988 On the uses of data on lifetime reproductive success. In *Reproductive success* (ed. T. H. Clutton-Brock), pp. 454–471. Chicago, IL: University of Chicago Press.
- Gustafsson, L. 1986 Lifetime reproductive success and heritability—empirical support for Fisher's fundamental theorem. *Am. Nat.* **128**, 761–764. (doi:10.1086/284601)
- Hadfield, J. D., Burgess, M. D., Lord, A., Phillimore, A. B., Clegg, S. M. & Owens, I. P. F. 2006 Direct versus indirect sexual selection: genetic basis of colour, size and recruitment in a wild bird. *Proc. R. Soc. B* **273**, 1347–1353. (doi:10.1098/rspb.2005.3459)
- Hadfield, J. D., Nutall, A., Osorio, D. & Owens, I. P. F. 2007 Testing the phenotypic gambit: phenotypic, genetic and environmental correlations of colour. *J. Evol. Biol.* **20**, 549–557. (doi:10.1111/j.1420-9101.2006.01262.x)
- Henderson, C. R. 1975 Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447. (doi:10.2307/2529430)
- Henderson, C. R. 1990 Accounting for selection and mating biases in genetic evaluation. In *Advances in statistical methods for genetic improvement of livestock* (eds D. Gianola & K. Hammond), pp. 413–436. Berlin, Germany: Springer.
- Henderson, C. R., Kempthorne, O., Searle, S. R. & Vonkrosigk, C. M. 1959 The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218. (doi:10.2307/2527669)
- Heywood, J. S. 2005 An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection. *Evolution* **59**, 2287–2298.
- Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., Beerli, P. & Kingsolver, J. G. 2001 Strength and tempo of directional selection in the wild. *Proc. Natl Acad. Sci. USA* **98**, 9157–9160. (doi:10.1073/pnas.161281098)
- Im, S., Fernando, R. L. & Gianola, D. 1989 Likelihood inferences in animal breeding under selection—a missing-data theory view point. *Genet. Sel. Evol.* **21**, 399–414. (doi:10.1051/gse:19890401)
- Janzen, F. J. & Stern, H. S. 1998 Logistic regression for empirical studies of multivariate selection. *Evolution* **52**, 1564–1571. (doi:10.2307/2411330)
- King, R., Brooks, S. P., Morgan, B. J. T. & Coulson, T. 2006 Factors influencing soay sheep survival: a Bayesian analysis. *Biometrics* **62**, 211–220. (doi:10.1111/j.1541-0420.2005.00404.x)
- Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P. & Beerli, P. 2001 The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261. (doi:10.1086/319193)
- Kirkpatrick, M., Lofsvold, D. & Bulmer, M. 1990 Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* **124**, 979–993.
- Kruuk, L. E. B., Clutton-Brock, T. H., Slate, J., Pemberton, J. M., Brotherstone, S. & Guinness, F. E. 2000 Heritability of fitness in a wild mammal population. *Proc. Natl Acad. Sci. USA* **97**, 698–703. (doi:10.1073/pnas.97.2.698)
- Kruuk, L. E. B., Merila, J. & Sheldon, B. C. 2001 Phenotypic selection on a heritable size trait revisited. *Am. Nat.* **158**, 557–571. (doi:10.1086/323585)
- Kruuk, L. E. B., Slate, J., Pemberton, J. M., Brotherstone, S., Guinness, F. & Clutton-Brock, T. 2002 Antler size in red deer: heritability and selection but no evolution. *Evolution* **56**, 1683–1695.
- Lande, R. 1979 Quantitative genetic analysis of multivariate evolution, applied to the brain: body size allometry. *Evolution* **33**, 402–416. (doi:10.2307/2407630)
- Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226. (doi:10.2307/2408842)
- Lawless, J. F. 2003 *Statistical models and methods for lifetime data. Wiley series in probability and statistics*, 2nd edn. New York, NY: Wiley.
- Lebreton, J. D., Burnham, K. P., Clobert, J. & Anderson, D. R. 1992 Modeling survival and testing biological hypotheses using marked animals—a unified approach with case-studies. *Ecol. Monogr.* **62**, 67–118. (doi:10.2307/2937171)
- Little, R. J. A. 1995 Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Stat. Assoc.* **90**, 1112–1121. (doi:10.2307/2291350)
- Little, R. J. A. & Rubin, D. B. 2002 *Statistical analysis with missing data*, 2nd edn. Hoboken, NJ: Wiley-InterScience.
- Lush, J. L. & Shrode, R. R. 1950 Changes in milk production with age and milking frequency. *J. Dairy Sci.* **33**, 338–357.
- Lynch, M. & Arnold, S. J. 1988 The measurement of selection on size and growth. In *Size-structured populations* (eds B. Ebenmann & L. Persson), pp. 47–58. Berlin, Germany: Springer.
- Lynch, M. & Walsh, B. 1998 *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer.
- Manly, B. F. 1985 *The statistics of natural selection*. London, UK: Chapman & Hall.
- Merila, J., Kruuk, L. E. B. & Sheldon, B. C. 2001a Natural selection on the genetical component of variance in body condition in a wild bird population. *J. Evol. Biol.* **14**, 918–929. (doi:10.1046/j.1420-9101.2001.00353.x)
- Merila, J., Kruuk, L. E. B. & Sheldon, B. C. 2001b Cryptic evolution in a wild bird population. *Nature* **412**, 76–79. (doi:10.1038/35083580)
- Metcalfe, C. J. E. & Pavard, S. 2007 Why evolutionary biologists should be demographers. *Trends Ecol. Evol.* **22**, 205–212. (doi:10.1016/j.tree.2006.12.001)
- Meyer, K. & Thompson, R. 1984 Bias in variance and covariance component estimators due to selection on a correlated trait. *Zeitschrift Fur Tierzucht Und Zuchtungsbiologie-J. Anim. Breed. Genet.* **101**, 33–50.
- Mitchell-Olds, T. & Shaw, R. G. 1987 Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* **41**, 1149–1161. (doi:10.2307/2409084)

- Mrode, R. A. 1996 *Linear models for the prediction of animal breeding values*. Wallingford, UK: CAB International.
- Nussey, D. H., Wilson, A. J., Morris, A., Pemberton, J., Clutton-Brock, T. & Kruuk, L. E. B. 2008 Testing for genetic trade-offs between early- and late-life reproduction in a wild red deer population. *Proc. R. Soc. B* **275**, 745–750. (doi:10.1098/rspb.2007.0986)
- Ouweltjes, W., Schaeffer, L. R. & Kennedy, B. W. 1988 Sensitivity of methods of variance component estimation to culling type of selection. *J. Dairy Sci.* **71**, 773–779.
- Patterson, H. D. & Thompson, R. 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554. (doi:10.1093/biomet/58.3.545)
- Pearson, K. 1903 Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. R. Soc. A* **200**, 1–66. (doi:10.1098/rsta.1903.0001)
- Phillips, P. C. & Arnold, S. J. 1989 Visualizing multivariate selection. *Evolution* **43**, 1209–1222. (doi:10.2307/2409357)
- Pletcher, S. D. & Geyer, C. J. 1999 The genetic analysis of age-dependent traits: modeling the character process. *Genetics* **153**, 825–835.
- Postma, E. 2006 Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. *J. Evol. Biol.* **19**, 309–320. (doi:10.1111/j.1420-9101.2005.01007.x)
- Price, G. R. 1970 Selection and covariance. *Nature* **227**, 520–521. (doi:10.1038/227520a0)
- Price, T., Kirkpatrick, M. & Arnold, S. J. 1988 Directional selection and the evolution of breeding date in birds. *Science* **240**, 798–799. (doi:10.1126/science.3363360)
- Rausher, M. D. 1992 The measurement of selection on quantitative traits—biases due to environmental covariances between traits and fitness. *Evolution* **46**, 616–626. (doi:10.2307/2409632)
- Rice, S. H. 2000 The evolution of developmental interactions: epistasis, canalization and integration. In *Epistasis and the evolutionary process* (eds J. B. Wolf, E. D. Brodie & M. J. Wade), pp. 82–98. New York, NY: Oxford University Press.
- Rice, S. H. 2002 A general population genetic theory for the evolution of developmental interactions. *Proc. Natl Acad. Sci. USA* **99**, 15 518–15 523. (doi:10.1073/pnas.202620999)
- Rice, S. H. 2004a Developmental associations between traits: covariance and beyond. *Genetics* **166**, 513–526. (doi:10.1534/genetics.166.1.513)
- Rice, S. H. 2004b *Evolutionary theory: mathematical and conceptual foundations*. Sunderland, MA: Sinauer.
- Richards, F. J. 1959 A flexible growth function for empirical use. *J. Exp. Bot.* **10**, 290–300. (doi:10.1093/jxb/10.2.290)
- Robertson, A. 1966 A mathematical model of culling process in dairy cattle. *Anim. Prod.* **8**, 95–108.
- Rothschild, M. F., Henderson, C. R. & Quaas, R. L. 1979 Effects of selection on variances and covariances of simulated 1st and 2nd lactations. *J. Dairy Sci.* **62**, 996–1002.
- Rubin, D. B. 1976 Inference and missing data. *Biometrika* **63**, 581–590. (doi:10.1093/biomet/63.3.581)
- Schaeffer, L. R. 1987 Estimation of variance components under a selection model. *J. Dairy Sci.* **70**, 661–671.
- Schaeffer, L. R., Schenkel, F. S. & Fries, L. A. 1998 Selection bias on animal model evaluation. In *Sixth World Congress on Genetics Applied to Livestock Production, University of New England, Armidale, NSW*, pp. 501–508.
- Scheiner, S. M., Donohue, K., Dorn, L. A., Mazer, S. J. & Wolfe, L. M. 2002 Reducing environmental bias when measuring natural selection. *Evolution* **56**, 2156–2167.
- Schluchter, M. D., Greene, T. & Beck, G. J. 2001 Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal disease. *Stat. Med.* **20**, 989–1007. (doi:10.1002/sim.720)
- Sheldon, B. C., Kruuk, L. E. B. & Merila, J. 2003 Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution* **57**, 406–420.
- Sorensen, D. & Gianola, D. 2002 *Likelihood, Bayesian and MCMC methods in quantitative genetics. Statistics for biology and health*. New York, NY: Springer.
- Sorensen, D. A., Wang, C. S., Jensen, J. & Gianola, D. 1994 Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet. Sel. Evol.* **26**, 333–360. (doi:10.1051/gse:19940403)
- Stinchcombe, J. R., Rutter, M. T., Burdick, D. S., Tiffin, P., Rausher, M. D. & Mauricio, R. 2002 Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Am. Nat.* **160**, 511–523. (doi:10.1086/342069)
- Thompson, R. 1973 Estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* **29**, 527–550. (doi:10.2307/2529174)
- Tsiatis, A. A. & Davidian, M. 2004 Joint modeling of longitudinal and time-to-event data: an overview. *Stat. Sin.* **14**, 809–834.
- VanTassel, C. P. & VanVleck, L. D. 1996 Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. *J. Anim. Sci.* **74**, 2586–2597.
- van Tienderen, P. H. & de Jong, G. 1994 A general model of the relation between phenotypic selection and genetic response. *J. Evol. Biol.* **7**, 1–12. (doi:10.1046/j.1420-9101.1994.7010001.x)
- Vonesh, E. F., Greene, T. & Schluchter, M. D. 2006 Shared parameter models for the joint analysis of longitudinal data and event times. *Stat. Med.* **25**, 143–163. (doi:10.1002/sim.2249)
- Wei, X. M. & Borralho, N. M. G. 1998 Use of individual tree mixed models to account for mortality and selective thinning when estimating base population genetic parameters. *For. Sci.* **44**, 246–253.
- Willis, J. H., Coyne, J. A. & Kirkpatrick, M. 1991 Can one predict the evolution of quantitative characters without genetics. *Evolution* **45**, 441–444. (doi:10.2307/2409678)
- Winn, A. A. 2004 Natural selection, evolvability and bias due to environmental covariance in the field in an annual plant. *J. Evol. Biol.* **17**, 1073–1083. (doi:10.1111/j.1420-9101.2004.00740.x)
- Wu, M. C. & Carroll, R. J. 1988 Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188. (doi:10.2307/2531905)
- Zwann, B. J. 1999 The evolutionary genetics of ageing and longevity. *Heredity* **82**, 589–597. (doi:10.1046/j.1365-2540.1999.00544.x)