

A framework for power and sensitivity analyses for quantitative genetic studies of natural populations, and case studies in Soay sheep (*Ovis aries*)

M. B. MORRISSEY,* A. J. WILSON,† J. M. PEMBERTON† & M. M. FERGUSON*

*Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

†Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Keywords:

genetic correlation;
heritability;
maternal effect;
natural populations;
Ovis aries;
pedigree error;
power analysis;
quantitative genetics;
sensitivity analysis;
Soay sheep.

Abstract

Studies of the quantitative genetics of natural populations have contributed greatly to evolutionary biology in recent years. However, while pedigree data required are often uncertain (i.e. incomplete and partly erroneous) and limited, means to evaluate the effects of such uncertainties have not been developed. We have therefore developed a general framework for power and sensitivity analyses of such studies. We propose that researchers first generate a set of pedigree data that they wish to use in a quantitative genetic study, as well as data regarding errors that occur in that pedigree. This pedigree is then permuted using the data regarding errors to generate hypothetical 'true' and 'assumed' pedigrees that differ so as to mimic pedigree errors that might occur in the study system under consideration. Phenotypic data are then simulated across the true pedigree (according to user-defined genetic and environmental covariance structures), before being analysed with standard quantitative genetic techniques in conjunction with the 'assumed' pedigree data. To illustrate this approach, we conducted power and sensitivity analyses in a well-known study of Soay sheep (*Ovis aries*). We found that, although the estimation of simple genetic (co)variance structures is fairly robust to pedigree errors, some potentially serious biases were detected under more complex scenarios involving maternal effects. Power analyses also showed that this study system provides high power to detect heritabilities as low as about 0.09. Given this range of results, we suggest that such power and sensitivity analyses could greatly complement empirical studies, and we provide the computer program PEDANTICS to aid in their application.

Introduction

Quantitative genetic techniques are currently contributing greatly to our understanding of evolutionary processes in natural populations. Until recently, one of the major challenges to the application of quantitative genetic techniques in natural populations was the unbalanced nature of breeding designs that tend to result from natural mating systems (Lynch & Walsh, 1998). However, with the recent introduction of linear mixed models, primarily the 'animal model' (Kruuk, 2004), to

evolutionary genetics, robust application of quantitative genetic techniques in natural populations has become possible in a wider range of situations (e.g. Kruuk *et al.*, 2001; Coltman *et al.*, 2003). Currently, an unresolved issue concerns the application of these techniques in study systems where pedigrees cannot be resolved with absolute certainty (Charmantier & Réale, 2005). The primary goal of this work is to describe a general framework in which analyses might be conducted to assess the sensitivity of particular quantitative genetic analyses to pedigree errors. To illustrate this framework, we describe two case studies of a well-known system of Soay sheep (*Ovis aries*). We first examined the potential for biases in variance component estimates because of the rates of pedigree errors that occur in this study system. Second, we conducted a simple power analysis to

Correspondence: Michael B. Morrissey, Department of Integrative Biology, University of Guelph, Guelph, ON, Canada N1G 2W1.
Tel.: +1 519 824 4120 ext. 58596; fax: +1 519 767 1656;
e-mail: mmorris@uoguelph.ca

determine the power of this study system to detect heritability of a particular size.

All procedures for obtaining quantitative genetic estimates and predictions use the extent to which relatives resemble one another to determine the degree to which (co)variances of (between) traits are genetically determined. All these procedures assume that the underlying pedigree data are known without error. This assumption is often violated, both in animal breeding programmes (e.g. Gelderman *et al.*, 1986; Visscher *et al.*, 2002) and in analyses of natural populations (e.g. Charmantier & Réale, 2005). For estimating the genetic variance (or heritability) of a single trait, the most likely consequence of pedigree errors is systematic parameter underestimation. This is expected to occur because pedigree errors of any sort will reduce the extent to which relatives appear to resemble one another. By a similar argument, one would also expect estimates of additive genetic covariances between two traits to be downwardly biased. However, the expected effect of pedigree errors on estimated genetic correlations is less clear. Genetic correlations are calculated by scaling the additive genetic covariance between two traits to the additive genetic variances of the traits. Whether or not any bias will occur in estimates of genetic correlations will therefore depend on whether or not additive genetic variances and covariances are underestimated by the same proportions.

In addition to heritabilities and genetic correlations, pedigree errors may also affect other quantitative genetic parameters of interest to evolutionary biologists. For example, a number of recent studies have scrutinized the role that maternal genetic effects may play in shaping evolutionary trajectories (Wolf & Wade, 2001; McAdam *et al.*, 2002; Wilson *et al.*, 2005a). However, the impact of pedigree error on estimates of maternal genetic variance is rather difficult to predict *a priori*. For example, it seems intuitive that estimates of maternal genetic effects might be inflated if errors occur more often in paternal pedigree links than in maternal links. This is because individuals will appear to be more similar, for genetic reasons, to their mothers than to their assumed fathers. However, it could also be argued that, as maternal genetic effects are both maternally and paternally inherited (although only expressed in offspring phenotypes via genes acting in the mother), pedigree errors will downwardly bias estimated maternal genetic variance in the same way as additive genetic variance. Furthermore, as components of phenotypic variance are jointly estimated under an animal model, any bias in estimating maternal genetic variance may have consequences for other parameters. For example, if similarity among maternal half-siblings results from both maternal genetic and maternal environmental effects, then a downward bias in estimating the former may cause an upward bias in the latter (or vice versa).

Several studies have assessed the effects (or potential effects) of pedigree errors on quantitative genetic anal-

yses in various ways. The majority of these studies examine situations that might occur in animal breeding programmes, primarily in dairy cattle. Senneke *et al.* (2004) showed that heritability estimates for birth weight and weaning weight decreased as the rate of simulated pedigree errors increased. Interestingly, they also showed that pedigree errors caused an underestimation of maternal genetic effects, although no reason was given for this. Israel & Weller (2000) showed that pedigree errors affected predicted breeding values such that the rate of genetic improvement was reduced in a simulated cattle breeding program. While conditions in natural populations usually differ in important respects (e.g. sample sizes and pedigree structures) from those typical of animal breeding systems, only one comprehensive simulation study has been performed to date. Charmantier & Réale (2005) used simulations based on data from a natural blue tit (*Parus caeruleus*) population to determine the extent to which undetected extra-pair paternity in avian systems will affect heritability estimates. Their results showed that heritabilities will be underestimated as a result of pedigree errors arising from unrecognized extra-pair paternity. Fortunately, though, their results also indicated that this effect will be small (less than 15% underestimation of heritability) for the rates of extra-pair paternity (typically about 20%) that occur in most avian populations. Similarly, by accounting for all known instances of extra-pair paternity in a population of collared flycatchers (*Ficedula albicollis*), Merilä *et al.* (1998) found that heritability estimates for several traits increased, but not by enough to qualitatively change their results.

However, work on the effect of pedigree errors in natural populations has been limited in two important respects. First, it has been focused on avian systems and passerines in particular. No study to date has examined the impact of errors in pedigree structures typical of natural populations for other taxa. For example, a large proportion of quantitative genetic studies in the wild are based on ungulate populations where available data sets may have very different properties (i.e. small sibships and higher variance in male than in female reproductive success). Consequently, it may not be appropriate to generalize the results of passerine studies to quantitative genetic analyses of natural populations in general. Secondly, the range of conditions in study systems is enormous: patterns and the rates of pedigree errors vary, patterns and quantities of available data vary, the size of pedigrees varies, genetic architectures vary, the relative and absolute magnitudes of different types of genetic and nongenetic effects vary, and the uses of quantitative genetic estimates are very diverse. Thus, extrapolation between systems will be difficult even when many more studies have been conducted.

Similarly, power analyses of quantitative genetic studies in natural populations are rare, but are potentially

very useful. Quantitative genetic studies notoriously require large quantities of data, and furthermore these data are often very difficult and expensive to collect. Thus, a robust method of power analyses will likely greatly serve empiricists, particularly in the stages of planning a study or an analysis.

One approach to such power analyses that has been applied is the partial use of data from existing data sets (Quinn *et al.*, 2006). Quinn *et al.* sought to determine how much sampling effort was required to estimate heritability when data are to be collected for new phenotypic traits in long-term studies in avian systems. They approached the problem by manipulating the availability of existing phenotypic data in two existing long-term data sets. This approach has the benefit that data structures in the sensitivity analyses very closely mimic the structure of data in real analyses. The approach is not widely applicable, however, because of its requirement for an existing pedigree. Furthermore, genetic architectures cannot be manipulated, and results are dependent on the adequacy of the model with which the available phenotypic data were originally analysed. A framework for sensitivity analyses that could be applied in a variety of study systems and for a variety of different genetic architectures would clearly greatly complement empirical studies.

Here, our objective was to present a simulation-based framework for conducting power and sensitivity analyses to evaluate the impact of pedigree errors on quantitative genetic parameter estimation in natural populations. In this context, power analyses are attempts to determine the potential for a given data set to address a particular biological question; for example, whether or not a particular data set contains enough information for modest heritabilities to be detected. Sensitivity analyses are attempts to understand how robust we can expect our findings to be in the face of such uncertainties as partly erroneous data or model inadequacy. Such a framework will be most useful if simulation conditions can be tailored as closely as possible to the conditions of specific study systems. There are two basic simulation approaches that could be taken to this problem. First, we might adopt a 'forward' approach, simulating a 'true' pedigree before imposing errors in that pedigree to generate an 'observed' pedigree. The true pedigree is the pedigree structure across which phenotypic data are simulated, and the observed pedigree is the potentially erroneous structure that one might observe and use in a quantitative genetic analysis. Note that the true pedigree does not necessarily exactly reflect a pedigree that exists in nature. We use this term here to describe the true pedigree structure relative to the phenotypic data that are used in power and/or sensitivity analyses. This approach was taken by Charmantier & Réale (2005) in their analysis of the effects of unrecognized extra-pair paternity in an avian system. An alternative strategy is a 'reverse' approach in which we start with an observed

pedigree that contains errors from which we simulate a true pedigree, based on information about the types and rates of errors in the observed pedigree. The observed pedigree may be a pedigree that is available for a quantitative genetic analysis of real data. Phenotypic data are then simulated with user-determined variance components across the true pedigree. The impact of pedigree errors can then be assessed by comparison of estimates from quantitative genetic analyses of the simulated phenotypic data conducted using the original observed pedigree.

Both the forward and reverse approaches are likely to prove useful when conducting power and sensitivity analyses in natural populations. The forward approach will generally be most useful in the early stages of a quantitative genetic study. For example, the forward approach can allow investigators to effectively plan quantitative genetic studies by determining how much data and what type of data need to be collected. However, once such data (pedigree and phenotypic) have been collected, the reverse approach will probably allow investigators to most closely mimic the conditions of their particular study systems.

In this paper, we describe in detail how the framework that we advocate can be applied in many study systems, and we provide software, PEDANTICS, to aid its implementation. To illustrate this framework, we apply it to example power and sensitivity analyses of a well-studied system of Soay sheep (*O. aries*).

Methods

Framework for power and sensitivity analyses

Our framework for power and sensitivity analyses for quantitative genetic studies of natural populations can be summarized in seven steps and is depicted as a flow diagram in Fig. 1. The first three steps require that the investigator specifies: (1) a pedigree; (2) information about potential errors in that pedigree; and (3) the genetic architectures for traits of interest. The following three steps require that a computer generates: (4) separate true and observed pedigrees based on the information supplied by the investigator in steps 1 and 2; (5) simulated phenotypic data based on the true pedigree and parameters defining genetic architecture (from step 3); and (6) output files containing the newly simulated data. In the final step (7), the investigator uses the observed pedigree with the simulated phenotypes to conduct quantitative genetic analyses, and compares parameter estimates to the genetic architectures initially specified in the simulations. Note that in the forward approach, the true pedigree is the one supplied by the investigator in step 1, whereas in the reverse approach the observed pedigree (i.e. available structure based on observational work and/or molecular pedigree analysis) is the pedigree from step 1. Each of these steps and some

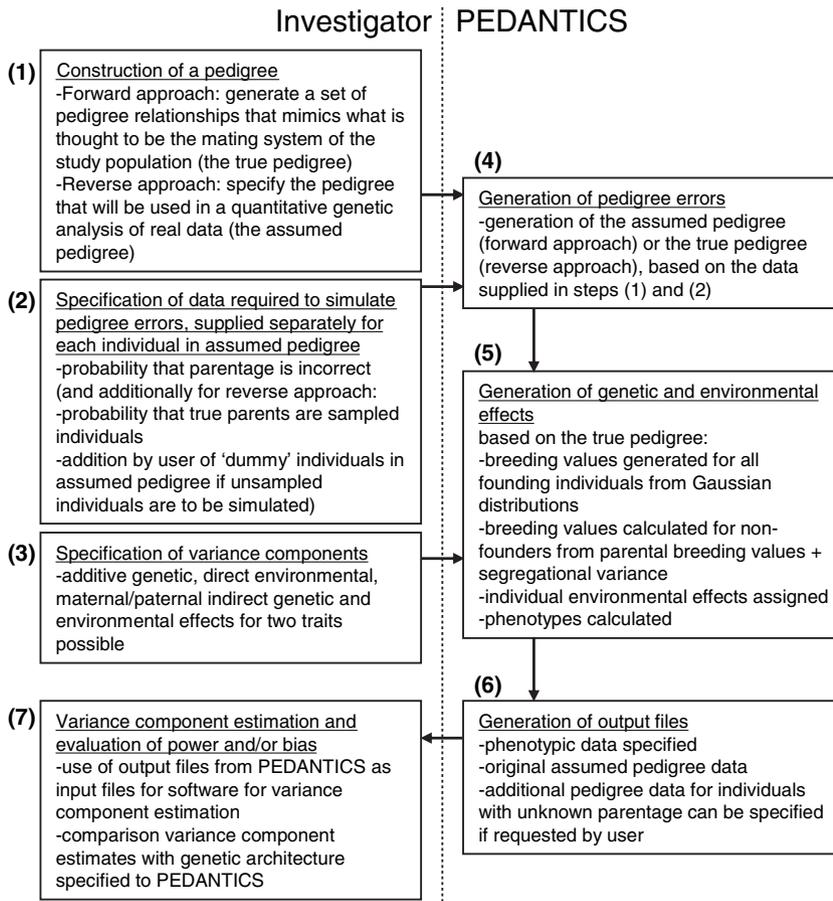


Fig. 1 Flow diagram of the proposed framework for power and sensitivity analyses of quantitative genetic analyses of natural populations. The right-hand side of the figure shows those steps that will generally require the aid of a computer, and that can currently be performed using the software PEDANTICS.

important considerations are discussed separately below (and are also described in Fig. 1). The steps in this framework that will generally require the assistance of a computer can be performed using our software, PEDANTICS. For the purposes of convenience and illustration, some aspects of these steps are described based on how they are performed by PEDANTICS.

(1) Pedigree specification

In this step, either the true pedigree or the observed pedigree is specified, depending on whether the forward or reverse approach is being applied, respectively. Both pedigrees can be specified via individual records with the identity of each individual's mother and father. Note that in some situations, a parentage analysis is inherently problematic, and observed pedigrees may be based on sibship reconstruction, therefore containing only a single generation (or cohort). In such cases, a set of sibs can still be specified based on their relationships to their parents, with those parents designated as unsampled individuals. In the forward approach, the specified pedigree must be complete, i.e. all parental links must be specified. In the reverse approach not all pedigree links need be specified, although any unspec-

ified links involving nonfounding individuals will necessarily be simulated in step 4 and supplemental information must therefore be provided at that stage. To allow simulation of a true pedigree in the presence of null assignments or erroneous assignments (or both), cohort (or birth year) designations will generally be required (discussed below). Finally, the investigator must specify some individuals in the pedigree to be designated as founders, sampled from a base population. In an animal model, it will be the variance components in this base population that are ultimately estimated. Designation of founders is also necessary because these will be the first individuals to which a computer will have to assign simulated genetic and environmental effects on phenotype.

(2) Information regarding pedigree errors

In some cases, pedigree data specified in step 1 may be both complete and assumed to be perfect, for example in some power analyses. More generally, information regarding pedigree errors is required and is best specified on a per-record basis. For each individual in the observed pedigree, probabilities that the assumed mother is not the true mother and that the assumed

father is the not true father must be specified. These values may typically be the same for all individuals in a pedigree. However, flexibility to allow error rates to vary will also be useful, for example, when the conditions of a study vary over time. If erroneous assignments have non-zero probabilities, then information about potential assumed parents (forward approach) or true parents (reverse approach) must also be provided. To allow sets of potential parents to be determined, the sex, as well as the first and last cohorts to which each individual might have contributed offspring, must be specified in each individual's record.

In many field situations, sampling of the populations will be incomplete. In this case, it is possible that the true parents of any given individual have not been sampled. To handle this scenario, unsampled individuals are added to the pedigree by the investigator. The first and last cohorts to which these unsampled individuals might have contributed must again be specified. In the forward approach, some individuals will have these unsampled individuals specified as their true parents. To specify the assumed pedigree, any such individuals will also either have to have a probability of 1 of having an unsampled parent in the observed pedigree or must have a probability of 1 of having an erroneously assigned parent.

In the reverse approach, for those individuals with missing parental links, the probability that their true parent (mother and/or father) was an unsampled individual is included. For individuals with specified parentage in the observed pedigree, the probability, given that an erroneous parentage assignment occurs, that the true mother and/or father are/is taken from among the unsampled individuals must be specified.

(3) Genetic (co)variance structures for phenotypic trait(s) of interest

In principle, the framework we advocate is applicable to almost any specified genetic architecture. Currently, the software PEDANTICS can accommodate direct additive genetic, environmental, and maternal and paternal indirect genetic and environmental effects simultaneously for two traits.

In PEDANTICS, each genetic or environmental effect is specified using a variance–covariance matrix [V1, V2, COV12]. In terms of heritabilities and genetic correlations, two traits, each with heritabilities of 0.2, with a genetic correlation of 0.5, and total phenotypic variances of unity with neither environmental covariance nor indirect effects would be specified by the following values: direct additive genetic [0.2, 0.2, 0.1] and direct environmental [0.8, 0.8, 0].

In real study systems, phenotypic data will rarely be available for all traits for all individuals. Therefore, individual records must also include information regarding whether or not phenotypic data will ultimately need to be specified for each trait.

(4) Simulation of a true (reverse approach) or assumed (forward approach) pedigree

The software PEDANTICS performs this step one record at a time. In the forward approach, two steps are required to generate the assumed pedigree. First, when an individual is to have an erroneously assigned parent, one is assigned by picking a parent (of the correct sex) from among all individuals that might have been the parent (based on the cohort of the offspring and the first and last cohorts to which all the other individuals in the pedigree might have contributed). Second, if an individual is to have a null parentage assignment in the assumed pedigree (the user specifies a probability of any given individual having an assigned parent), the record is deleted from the assumed pedigree.

In the reverse approach, for each individual that does not have a specified parent, or for which an erroneous parentage assignment is to be generated, one is assigned by picking a parent (of the correct sex) from among all individuals that might have been the parent (based on the cohort of the offspring and the first and last cohorts to which all the other individuals in the pedigree might have contributed). All such possible parents are currently considered equally likely, lest many more parameters need be specified. If the possibility of unsampled parents is considered, the different probabilities specified in step 2 for whether or not the true parent was in the pool of sampled individuals will be accommodated. However, within the two pools of appropriate parents (i.e. sampled vs. unsampled), the probability of a particular individual being chosen as the true parent follows a uniform distribution.

(5) Simulation of phenotypic values

First, the breeding values (direct effects or maternal or paternal indirect effects) of founding individuals are assigned from Gaussian distributions with (co)variances as specified in step 3. The computer then assigns breeding values to all other individuals in the pedigree. Only when breeding values have been assigned to a specific individual's parents, can an individual's own breeding values be generated. PEDANTICS simulates breeding values of all nonfounding individuals from Gaussian distributions with mean equal to the mean of the parental breeding values and variance equal to half of the specified population variance for the component being simulated, i.e.

$$\mathbf{a}_i = \mathbf{m}_i + C(\mathbf{G}/2) \cdot \mathbf{N}$$

where \mathbf{a}_i is the vector of breeding values of the focal individual, \mathbf{m}_i is the vector of mid-parent values, $C(\mathbf{G}/2)$ is the Cholesky decomposition of the matrix describing the segregational variance (i.e. \mathbf{G} is the variance–covariance matrix), and \mathbf{N} is a vector of standard random normal deviates. Mean expected breeding values of individuals are thus calculated ignoring inbreeding

among the parents (inbreeding depression) and the segregational variance is calculated ignoring the effects of inbreeding among grandparents (reduced segregational variance among the offspring of inbred individuals). Environmental effects are simulated for each individual based on data specified by the investigator in step 3. When both breeding values and individual environmental effects have been simulated, an individual's phenotype is calculated by their summation. When parental indirect genetic effects are specified by the investigator in step 3, the breeding values of the parents which are associated with the indirect effects are included in the summation to calculate individual phenotypes.

(6) Generation of output files

PEDANTICS generates output files that the investigator can use as data files for subsequent estimation of variance components or for other quantitative genetic analyses. A phenotypic data file is generated in which phenotypes are reported for all sampled individuals and for all traits as specified in step 3. If requested by the investigator, PEDANTICS will include maternal and/or paternal identities in the phenotypic data file, as these may be required for the estimation of indirect effects.

A pedigree data file is also prepared. In the reverse approach, this pedigree file will be equivalent to the observed pedigree that was supplied to PEDANTICS. In the forward approach, this structure may differ from the original pedigree because of the simulation of erroneous assignments. If requested by the investigator, PEDANTICS can include simulated parentage records for all non-founding individuals, therefore allowing null assignment rates of zero to be simulated.

The phenotypic and pedigree data files generated by PEDANTICS are intended for use with the software ASREML (VSN International, Hemel Hempstead, UK), as this software currently appears to be the most flexible software for quantitative genetic analyses of natural populations. However, data files used by ASREML are similar in format to those used by other analytical programmes. Investigators may also want to perform further types of power and sensitivity analyses not considered here and PEDANTICS therefore generates a master data file with much more data are included. This master file can be modified for use by any software and contains identities of true and assumed parents, phenotypic data for all individuals and, for each trait, the individual level genetic and environmental effects that sum to the phenotype.

(7) Analysis and evaluation

Finally, the output of PEDANTICS must be analysed and evaluated by the investigator. First, the investigator will probably want to analyse replicate sets of output files from PEDANTICS, using quantitative genetic models appropriate to their intended study design (e.g. the animal model, ANOVA, parent-offspring regression, etc.).

The comparison of the results of such analyses with the genetic architectures initially specified (in step 3) will allow a great variety of power and sensitivity analyses to be performed of quantitative genetic analyses of natural populations.

Case study 1: sensitivity to pedigree errors of the St Kilda Soay sheep study system

Observed pedigree

The Soay sheep study system, described thoroughly in Clutton-Brock & Pemberton (2004), has been the subject of a long-term, multigenerational study and has been the subject of several quantitative genetic analyses (e.g. Milner *et al.*, 2000; Wilson *et al.*, 2005a; Wilson *et al.*, 2005b). This system therefore represents a suitable case study for illustrating our framework for power and sensitivity analyses.

In the Soay sheep system, the pedigree structure has been reconstructed using both observational (for maternities) and genotypic (for paternities) data (Pemberton *et al.*, 1999). Maternities based on observation of behaviour have been confirmed to be very accurate using molecular methods (Pemberton *et al.*, 1999). The paternal links that have been reconstructed from molecular data are made with varying levels of confidence. Here, we use as our observed pedigree a structure described by Wilson *et al.* (2005a) comprising 5535 animals with maternity known for 2861 individuals. Paternity data have been generated by maximum-likelihood-based categorical parentage assignment using the program CERVUS (Marshall *et al.*, 1998). Paternity is assigned at the 95% pedigree-wide confidence level for 920 individuals, at the 80% confidence level for 1846 individuals, and 2695 individuals have been assigned a most likely father based on microsatellite data. We therefore consider three observed pedigree structures that contain all known dams but differ in the confidence and rates at which paternities are assigned. We refer to the pedigrees based on sires assigned at the 95% and 80% confidence levels as the 95% pedigree and the 80% pedigree respectively. Similarly, we refer to the pedigree based on all known dams and all sires based on the most likely assignments as the ML pedigree. Most likely fathers are designated based solely on microsatellite data. They are the most likely individuals among the candidate males; but these assignments do not necessarily have enough statistical support for inclusion in either the 95% or 80% confidence pedigrees. All individuals born (or believed to have been born) prior to 1980 were designated as founders. Additionally, all individuals believed to have been immigrants to the study area were designated as founders.

Considerations regarding simulation of pedigree errors

Comparison of observational and molecular data has shown that maternal identification is very accurate in the

Soay sheep (Pemberton *et al.*, 1999) and we therefore assume that erroneous parentage assignments occur only in paternal links in the Soay sheep pedigree. The confidence levels for pedigree reconstruction describe the expected pedigree-wide paternity error rate, rather than thresholds for confidence in individual assignments. We therefore assume that paternal erroneous assignments occur at the rates of 0.05 and 0.2 in the 95% and 80% pedigrees respectively. The rate of erroneous assignments in the ML pedigree is difficult to estimate but was assumed to be 50% for the purposes of our simulations. This figure is somewhat arbitrary but seems appropriate as it is used primarily for illustrative purposes.

In addition to erroneous paternity assignment, it is also the case that parental information is incomplete, and the level of missing information varies across the pedigree structure used. Excluding founders (for which parents are, by definition, unknown), maternal null assignments occur at a rate of 0.11 in all three available observed pedigree structures, whereas paternal null assignments occur at the rates of 0.47, 0.32 and 0.21 in the 95%, 80% and ML pedigrees respectively. Note that a trade-off occurs between the rates of null and erroneous parentage assignments, and the effects of these two types of pedigree uncertainty are therefore confounded. See Appendix for an example of procedures to separately investigate the effects of these two types of uncertainty.

Because extensive pedigree data have already been collected in the Soay sheep study system, we adopted the reverse approach for the generation of the true and assumed pedigree. To generate the true pedigree from the available pedigree data we used available birth and death data to determine which individuals were potential mothers and fathers of lambs in each cohort. We assumed that the first cohort to which each individual could contribute was the cohort following that individual's birth. We assumed that the last cohort to which a male could have contributed lambs was the cohort following the last mating season during which the individual was known to have been alive. We assumed that the last cohort to which a female could have contributed was the cohort following the last spring in which the individual was known to have been alive. Any individuals that were never resighted following capture at birth were assumed to have died in or near infancy, such that they could not have contributed any offspring to future cohorts.

It is also believed that a small number of unsampled males gain some paternity in the study system, introducing the possibility that relatedness among sampled individuals can occur via unsampled fathers. We simulated five unsampled males born every year. Each of these males potentially contributed lambs to the six cohorts following its birth. When either null assignments (as specified by the observed pedigree) or paternal

erroneous assignments (generated probabilistically) occurred, we assumed that the true father was an unsampled individual 30% of the time.

Variance components

We simulated two types of genetic architectures. The first genetic architecture involved two continuous phenotypic traits, each with additive genetic variances of 0.2, with an additive genetic covariance of 0.1, and total phenotypic variances of 1 with no environmental covariance and no indirect effects. This generated a genetic architecture for the two traits with heritabilities of 0.2 and a genetic correlation of 0.5. In reality, phenotypic data were not available for all individuals for all traits. To mimic realistic levels of missing data in our analyses of the Soay pedigrees we only simulated phenotypic records for individuals that had in reality been captured and measured at birth (trait 1) and in the first August of their life (trait 2).

The second genetic architecture was simulated for a single trait with a direct additive genetic effect of 0.2, an indirect effect of the maternal environment of 0.2, an indirect effect of maternal additive genes of 0.2 and a total phenotypic variance of 1. For this genetic architecture, phenotypic records were only simulated for those individuals that had phenotypic measurements made at birth.

The magnitudes of these variance components were chosen based on three criteria. First, these effects are all large enough to be of practical significance in many evolutionary studies. Second, these variance components are large enough that investigators might reasonably hope to be able to detect them given feasible sample sizes. Third, these variance components are not so large that their detection, when they exist, may be regarded as a foregone conclusion.

Analytical models

For simulations involving the first assumed genetic architecture, bivariate analyses were conducted using an animal model of the form:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}_1 \mathbf{a} + \mathbf{e}$$

where \mathbf{y} is the matrix containing individual phenotypes at both traits, $\boldsymbol{\mu}$ contains the population means for each trait, \mathbf{a} is a matrix of direct additive genetic effects and \mathbf{e} is a matrix of residual errors. \mathbf{Z}_1 is an incidence matrix relating the individual effects in \mathbf{a} to phenotypes in \mathbf{y} .

For the second architecture in which maternal effects on a single trait were simulated, we used a univariate animal model of the form:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}_1 \mathbf{a} + \mathbf{Z}_2 \mathbf{m}_e + \mathbf{Z}_2 \mathbf{m}_a + \mathbf{e}$$

where \mathbf{m}_e and \mathbf{m}_a are the matrices of maternal permanent environmental and maternal additive genetic effects, respectively, and \mathbf{Z}_2 is an incidence matrix relating individuals to maternal identities. All models

were solved using the default REML algorithms in ASREML. In the analyses of maternal genetic effects, phenotypic records were discarded for all individuals without known maternity.

Interpretation of results

We illustrate our results with figures showing the mean estimates of each quantitative genetic parameter. Performing 1000 replicates of multiple simulation scenarios is time consuming and may therefore not be feasible for many investigators. We therefore bootstrapped 500 samples from our pools of 1000 simulation results with both $n = 10$ and 50, where n represents the number of simulation results, from the available pool of 1000 simulations, of which each of the 500 samples was composed. Thus, when $n = 10$, we are mimicking the situation where power and/or sensitivity analyses are conducted based on 10 replicate simulations, rather than the 1000 replicates that we used in our example case studies. From these samples we calculated the means of each variance component estimate. Across all samples, we calculated the 95% quantiles of the variance component estimates.

Case study 2: power of the St Kilda Soay sheep study system to detect nonzero heritability

Pedigree and variance components

For simplicity, we used only the pedigree with intermediate levels of error, i.e. the 80% pedigree. In one set of simulation scenarios, we simulated phenotypic data for all individuals that had phenotypic records for birth weight ($n = 2606$). In a second set of simulations, we simulated phenotypic records only for those individuals that had records for body weight at age 3 ($n = 692$). In the simulations where data mimicked data availability for birth weight, we simulated heritabilities of 0, 0.02, 0.04, 0.06, 0.08, 0.1, 0.15, 0.2, 0.3, 0.4 and 0.5. In the simulations where the data set mimicked data availability for age 3 body weight, we simulated heritabilities between 0 and 0.5 at intervals of 0.05.

Analytical model

For the data generated in these power analyses, univariate analyses were conducted using an animal model of the form

$$\mathbf{y} = \mu + \mathbf{Z}_1\mathbf{a} + \mathbf{e}$$

where \mathbf{y} is the vector containing individual phenotypes at both traits, μ contains the population mean, \mathbf{a} is a matrix of direct additive genetic effects and \mathbf{e} is a vector of residual errors. \mathbf{Z}_1 is an incidence matrix relating the individual effects in \mathbf{a} to phenotypes in \mathbf{y} .

Interpretation of results

As in our sensitivity analyses, we performed 1000 replicates of each simulation scenario and resampled

the replicate data to mimic analyses based on fewer replicates. As in the sensitivity analysis, we also calculated 95% quantiles of the power estimates derived from 500 samples with $n = 50$ and 10.

Results

Case study 1: sensitivity to pedigree errors of the St Kilda Soay sheep study system

Under the simple genetic architecture, we found no evidence of bias in estimates of genetic correlation, and while we did observe a downward bias in h^2 this was only substantial (approx. 35%) in the ML pedigree (Fig. 2a). By contrast, when a more complex genetic architecture was simulated, more evidence for an estimation bias was seen. For example, estimates of maternal genetic effects were systematically downwardly biased by about 25% (Fig. 2b) in simulations based on the three available pedigrees. Substantial upward biases of 50% and 25%

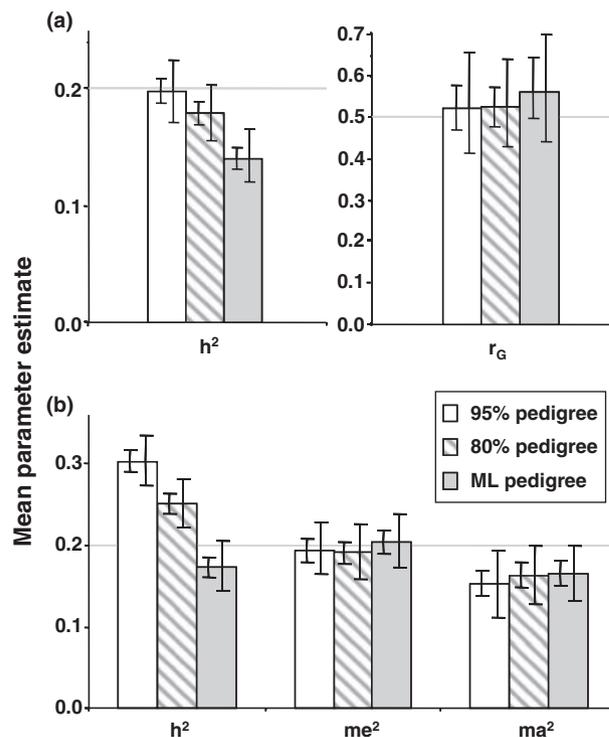


Fig. 2 Biases in quantitative genetic parameter estimates in the three available pedigrees in Soay sheep (*Ovis aries*). Horizontal grey lines represent the true parameter values and bar heights represent the mean of parameter estimates generated from 1000 replicate simulations. Error bars indicate the 95% quantiles of means calculated from 500 samples of the original 1000 simulations, thus describing the variability that might be expected if these estimates were based on fewer replicate simulations. The left- and right-hand error bars represent results generated by resampling with $n = 50$ and 10 respectively.

were detected in the estimates of heritability in the 95% and 80% pedigrees respectively (Fig. 2b).

Estimates of heritability (in both genetic architectures) and of effects of the maternal environment and of maternal genes based on 50 and 10 samples had 95% quantiles of about ± 0.015 and 0.025 respectively. 95% quantiles of the heritability estimates based on 50 and 10 samples were about ± 0.055 and 0.14 respectively.

Case study 2: power of the St Kilda Soay sheep study system to detect nonzero heritability

Sigmoidal relationships occur between true heritability and the power to detect heritability in the scenarios that we analysed (Fig. 3). When we simulated phenotypic records for all individuals with known birth weight, high power (i.e. ≥ 0.8) was detected for true heritabilities over about 0.09. When we simulated phenotypic data for

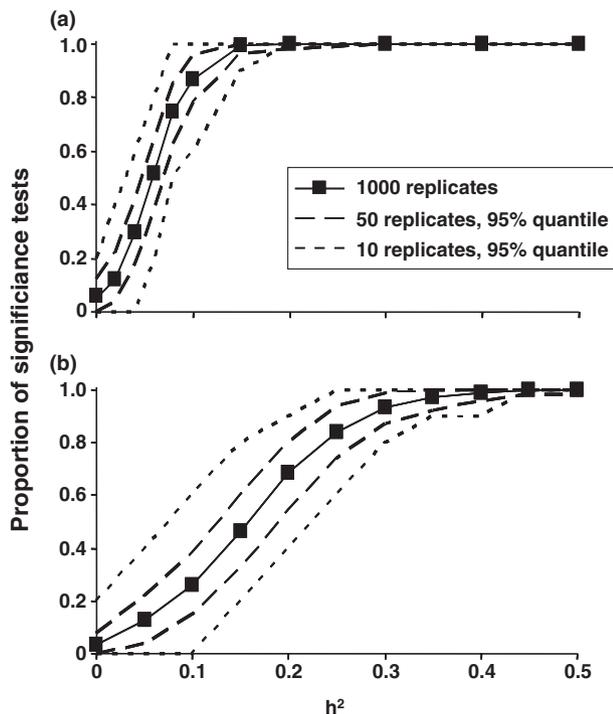


Fig. 3 Power of quantitative genetic analyses in Soay sheep (*Ovis aries*) to detect heritability in models involving (a) direct genetic effects on two correlated traits, and (b) direct (h^2), maternal environmental (m_c^2) and maternal genetic (m_a^2) effects. Data were generated for all individuals with phenotypic records for (a) birth weight ($n = 2606$) and for (b) body weight at age 3 ($n = 692$). Solid lines represent the probability of obtaining a significant test result based on 1000 replicate simulations. Dashed and dotted lines indicate the 95% quantiles of means calculated from 500 samples of the original 1000 simulations, thus describing the variability that might be expected if these estimates were based on fewer replicate simulations. The dashed and dotted lines represent results generated by resampling with $n = 50$ and 10 respectively.

individuals with data for body weight at age 3, power was reduced as expected, although true heritability values greater than 0.24 still allowed for high statistical power. These values were calculated from linear interpolation between true heritability values for predicted power values above and below 0.8 in each simulation scenario.

Results from power analyses based on many fewer than 1000 replicate simulations were most variable when power levels were intermediate (i.e. about 0.5; Fig. 3). For example, when heritability of 0.15 was simulated and records were generated for individuals with data for body weight at age 3, the predicted power based on 1000 replicate simulations was only 0.46. The 95% quantiles of estimates based on 50 and 10 replicates ranged 0.32–0.62 and 0.2–0.8 respectively.

Discussion

In the following discussion, we first consider the results of the Soay sheep case studies, before presenting a more general discussion of the proposed simulation approach to power and sensitivity analyses. Although we have focused here on the impact of pedigree errors and power to detect heritability, the methods we propose may also allow exploration of further issues relating to quantitative genetic analyses of natural populations. We therefore conclude by drawing attention to these.

Case studies in Soay sheep

Under simple genetic architectures, our simulations of pedigree error did not lead to biases large enough to hinder the biological interpretation of variance component estimates. Those biases that were detected occurred only in the estimation of heritability, whereas estimates of genetic correlation appeared to be largely unaffected by pedigree errors. Furthermore, quantitative genetic analyses of natural populations that have relied on likelihood-based parentage assignment have used pedigrees reconstructed based on the 80% confidence level or higher (e.g. Kruuk *et al.*, 2002; Wilson *et al.*, 2005a). The practice of avoiding the use of pedigrees based on less than 80% confidence on parental assignments seems justified. The greatest biases that we detected in analyses based on such pedigrees were in the order of a 25% downward bias in heritability estimates (Fig. 2). This corroborates the findings of Charmantier & Réale (2005), who came to a similar conclusion following a study of the effects of pedigree errors arising from extra-pair paternity in an avian study system. We might thus be able to cautiously conclude that, for the rates of pedigree errors that evolutionary biologists tend to tolerate when conducting quantitative genetic analyses, biases in estimates of direct heritability may generally be small relative to other types and sources of error.

Nevertheless, the effects of pedigree errors in our case study became more apparent when a more complex genetic architecture was simulated. When direct additive genetic and environmental effects, as well as maternal genetic and environmental effects, were simulated, more biases were detected, and some of these biases were larger in magnitude. For example, in the 95% pedigree, an upward bias of almost 50% was detected in the estimate of heritability, whereas estimates of the maternal genetic effect were downwardly estimated by about 20%. Again, these biases may not always greatly impact the biological interpretation of real data. However, if true heritability is zero, then parameter overestimation can clearly (wrongly) change the interpretation with respect to, for example, whether or not a trait is likely to respond to selection. Interestingly, the 95% pedigree has the lowest rate of erroneous paternity assignment, but also has the greatest rate of null assignment. Both of these may be viewed as sources of pedigree error that may be traded-off against each other. See Appendix for an investigation of the mechanism by which pedigree errors cause these patterns of biases in estimates with this particular genetic architecture.

Our sample power analyses provide examples of how power analyses might generally complement quantitative genetic studies (Fig. 3). The simulations based on data availability for phenotypes measured at birth demonstrated that this system can be very powerful for estimating heritability. An investigator might conclude that studies relying on the ability to detect fairly small heritability for traits expressed early in life are well advised in this system. Conversely, such studies that rely on powerful estimation of small heritability values for traits expressed late in life should be approached with caution.

Applications of the framework for power and sensitivity analyses

Applications of the approach described in this paper will probably differ somewhat from the example analysis that we have presented here. For simplicity, we only simulated variance components with values of 0.2 (and 0.5 for genetic correlations) in our sensitivity analyses, but we simulated a large range of heritabilities in our power analyses. Empirical studies will probably be best supported by these procedures if variance components of several magnitudes are simulated, perhaps corresponding to heritabilities of 0.05, 0.2 and 0.5.

From a practical point of view, an investigator will also need to consider the number of replicate simulations to perform. Although more is always better, it is clear that it will rarely be necessary to perform 1000 replicates of any simulation scenario. In our sensitivity analyses, the effect of sampling error resulting from $n = 10$ replicate simulations would not have changed this study's conclusions (Fig. 2). In our power analysis, $n = 10$ replicate simulations were also adequate for most purposes (Fig. 3),

except when heritabilities were modest and the data set was not powerful (Fig. 3b when $0.05 < h^2 < 0.25$). It would not be reasonable to attempt to provide specific guidelines regarding generally appropriate levels of replication for power and sensitivity analyses. Individual investigators will have to make their own decisions regarding levels of replication that are appropriate both for their data sets and for the biological interpretation of their analyses.

In our case studies, we have focused on the potential impact of pedigree errors for the estimation of quantitative genetic variance components in the wild, and on power to detect heritability. However, it should be noted that the simulation framework we propose (and the software provided) is readily adapted to exploration of other issues. Quantitative genetic studies of natural populations are notoriously costly in time and effort and power analyses may help optimize study designs. For example, given limited resources and the goal of estimating a trait heritability, should one invest more in field work (i.e. the amount of phenotypic data) or in molecular pedigree analysis (i.e. reduced null assignment and erroneous assignment rates in the pedigree)? *PEDANTICS* can easily be used to explore the implications of such decisions, or to perform power analyses in a more general sense by comparing pedigrees of different sizes, or with different levels of missing phenotypic data.

Our approach can also be applied to evaluations of model adequacy. For example, it is well known that common environmental effects (including maternal effects) can inflate estimates of heritability, if not adequately modelled (Kruuk & Hadfield, 2007). The methods described here can easily be extended to evaluate such situations. The investigator needs only to use *PEDANTICS* to simulate phenotypic data (step 3) across a specified pedigree using a particular genetic architecture (e.g. including a maternal effect) and later (step 7) fit a variety of analytical models (e.g. with and without maternal effects) to see how each performs.

Although we have demonstrated this framework for analyses in a large mammal study system, the approach is not taxonomically limited. For example, it would be equally applicable to avian study systems in which pedigree errors are likely to occur primarily as a result of extra-pair paternity. Typically, the rate of null assignments in such studies will be 0 (the observed pedigree will be complete), and only paternal erroneous assignments will be simulated. Furthermore, *PEDANTICS* can accommodate pedigrees of monoecious organisms, allowing power and sensitivity analyses to be conducted in study systems of hermaphroditic animals and monoecious plants.

Finally, the utility of *PEDANTICS* and its associated methodologies need not be limited to considerations of estimates of variance components. Breeding values and other individual effects are also of great interest and use to evolutionary biologists (e.g. Kruuk *et al.*, 2002;

Coltman *et al.*, 2003). The prediction of these effects is expected to be fraught with the same problems as the estimation of population-level parameters. PEDANTICS reports these individual-level effects to the investigator, and therefore all of the approaches described here can be applied at the individual level as well. Although not supported by PEDANTICS, the approaches described here should be amenable to supporting other methods of dissecting quantitative genetic inheritance, such as the detection and characterization of quantitative trait loci. PEDANTICS is available from AJW's website. The distributed software includes a manual and example input files.

Summary

We have presented a framework for the power and sensitivity analysis of quantitative genetic studies of natural populations. This framework is both generally applicable and easily implemented. Our application of this framework in a well-known study system has yielded results that are both of practical importance and that have highlighted the utility of our approach. Specifically, our results have corroborated previous findings that pedigree errors, at least at the rates at which they are normally tolerated, do not bias estimates of heritability to an extent that biological interpretation of such estimates is affected. However, we did demonstrate that pedigree errors can have substantial effects on variance component estimation when more complex genetic architectures are modelled. We hope that the availability of our framework for power and sensitivity analyses will produce two major benefits for the field of evolutionary quantitative genetics. First, we hope that the collection and interpretation of empirical data regarding the quantitative genetics of natural populations will be both aided and promoted. Second, we hope that the use of this framework will in future allow a more thorough evaluation of the effects of various uncertainties in the quantitative genetics of natural populations, based on results from a variety of systems.

Acknowledgments

Loeske Kruuk and three anonymous reviewers provided very helpful comments on earlier versions of this manuscript. MBM was supported by a CGS scholarship from the Natural Sciences and Engineering Council of Canada (NSERC) while the work described here was conducted. AJW was supported by a Leverhulme Trust project grant and by a Natural Environment Research Council post-doctoral fellowship. The St Kilda Soay sheep project and pedigree construction for the Soay population has been supported by the UK Natural Environment Research Council and the Wellcome Trust and takes place with the permission of the National Trust for Scotland, Scottish Natural Heritage and with the help of NTS, QinetiQ and Eurest staff. MMF was supported by an NSERC Discovery Grant.

References

- Charmantier, A. & Réale, D. 2005. How do misassigned paternities affect the estimation of heritability in the wild? *Mol. Ecol.* **14**: 2839–2850.
- Clutton-Brock, T. & Pemberton, J.M. (eds) 2004. *Soay sheep dynamics and selection in an island population*. Cambridge University Press, Cambridge.
- 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.
- Gelderman, H., Pieper, U. & Weber, W.E. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. *J. Anim. Sci.* **63**: 1759–1768.
- Israel, C. & Weller, J.I. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *J. Dairy Sci.* **83**: 181–187.
- Kruuk, L.E.B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. *Philos. Trans. R. Soc. Lond.* **359**: 873–890.
- Kruuk, L.E.B. & Hadfield, J. 2007. How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* doi: 10.1111/j.1420-9101.2007.01377.x.
- Kruuk, L.E.B., Merilä, J. & Sheldon, B.C. 2001. Phenotypic selection on a heritable size trait revisited. *Am. Nat.* **158**: 557–571.
- 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.
- Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, MA.
- Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**: 639–655.
- McAdam, A.G., Boutin, S., Réale, D. & Berteaux, D. 2002. Maternal effects and the potential for evolution in a natural population of animals. *Evolution* **56**: 846–851.
- Merilä, J., Sheldon, B.C. & Ellegren, H. 1998. Quantitative genetics of sexual size dimorphism in the collared flycatcher. *Evolution* **52**: 870–876.
- Milner, J.M., Pemberton, J.M., Brotherstone, S. & Albon, S.D. 2000. Estimating variance components and heritabilities on the wild: a case study using the "animal model" approach. *J. Evol. Biol.* **13**: 804–813.
- Pemberton, J.M., Coltman, D.W., Smith, J.A. & Pilkington, J. 1999. Molecular analysis of a promiscuous, fluctuating mating system. *Biol. J. Linn. Soc.* **68**: 289–301.
- Quinn, J.L., Charmantier, A., Garant, D. & Sheldon, B.C. 2006. Data depth, data completeness, and their influence on quantitative genetic estimation in two contrasting bird populations. *J. Evol. Biol.* **12**: 994–1002.
- Senneke, S.L., MacNeil, M.D. & Van Vleck, L.D. 2004. Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. *J. Anim. Sci.* **82**: 2307–2312.
- Visser, P.M., Woolliams, J.A., Smith, D. & Williams, J.L. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J. Dairy Sci.* **85**: 2368–2375.

- Wilson, A.J., Coltman, D.W., Pemberton, J.M., Overall, A.D.J., Byrne, K.A. & Kruuk, L.E.B. 2005a. Maternal genetic effects set the potential for evolution in a free-living vertebrate population. *J. Evol. Biol.* **18**: 405–414.
- Wilson, A.J., Pilkington, J.G., Pemberton, J.M., Coltman, D.W., Overall, A.D.J., Byrne, K.A. & Kruuk, L.E.B. 2005b. Selection on mothers and offspring: whose phenotype is it and does it matter? *Evolution* **59**: 451–463.
- Wolf, J.B. & Wade, M.J. 2001. On the assignment of fitness to parents and offspring: whose fitness is it and when does it matter? *J. Evol. Biol.* **14**: 347–356.

Received 17 April 2007; revised 26 June 2007; accepted 26 June 2007

Appendix: Separation of the effects of null assignments and of erroneous assignments

The approaches that we have advocated for power and sensitivity analyses can be extended to separately investigate the effects of null parentage assignments and of erroneous parentage assignments. The understanding of the causes of these effects is important because investigators will often have to decide between analysing pedigrees that either contain many null assignments or that contain many erroneous assignments. Ultimately, these considerations require that we consider null parentage assignments as a type of pedigree error. This is reasonable, because failure to assign a parentage will result in failure to recognize relatedness between potentially phenotypically similar individuals.

PEDANTICS can be used to separate the effects of null and erroneous parentage assignments. Null assignment rates can be reduced to zero by requesting PEDANTICS to generate parentage records for all individuals, regardless

of whether or not parentage records existed in the pedigree that was supplied by the investigator. For simplicity, such analyses will generally require that unsampled individuals are assumed not to exist. Similarly, the rates of erroneous parentage assignments can be reduced to zero by specifying error rates of zero for all individuals in the input data to PEDANTICS. Importantly, these manipulations can be made independently of one another.

We used the simulation scenarios summarized in Table A1 to separate the effects of null and erroneous parentage assignments in our sensitivity analyses of the Soay sheep study system. Briefly, we performed simulations that would allow us to compare scenarios in which no null assignments occurred, in which null assignments occurred only in maternal links, only in paternal links, and in both (scenarios 1–4). We then performed simulations where null assignments occurred at the rates that they occur in the 95%, 80% and ML pedigrees (scenarios 4–6). We then varied the rates of erroneous paternity assignments, both in the absence (scenarios 1 and 7–9) and in the presence (scenarios 5 and 10–12) of null assignments. To facilitate comparisons in scenarios with no null assignments, we did not simulate the presence of any unsampled individuals. We simulated both bivariate and univariate maternal genetic architectures in each of these scenarios. See the main text for descriptions of the magnitudes of these variance components and for descriptions of the models with which these data were analysed. We report biases here as units of h^2 , m_c^2 , m_a^2 or r_G above or below the true simulated values.

The type of pedigree error influenced whether or not biases occurred in variance component estimates. Null assignments had no effects on estimates of heritability

Table A1 Combinations of parameters used as input for PEDANTICS in sensitivity analyses of variance component estimates in Soay sheep (*Ovis aries*).

Scenario	Dams in observed pedigree	Sires in observed pedigree	Null assignment rate*		Erroneous assignment rate	
			Maternal	Paternal	Maternal	Paternal
1	All known + rest simulated	All known in ML pedigree + rest simulated	0	0	0	0
2	All known + rest simulated	Only those known in ML pedigree	0.11	0	0	0
3	Only known	All known in ML pedigree + rest simulated	0	0.21	0	0
4	Only known	Only those known in ML pedigree	0.11	0.21	0	0
5	Only known	Only those known in 80% pedigree	0.11	0.32	0	0
6	Only known	Only those known in 95% pedigree	0.11	0.47	0	0
7	All known + rest simulated	All known in ML pedigree + rest simulated	0	0	0	0.05
8	All known + rest simulated	All known in ML pedigree + rest simulated	0	0	0	0.20
9	All known + rest simulated	All known in ML pedigree + rest simulated	0	0	0	0.50
10	Only known	Only those known in 80% pedigree	0.11	0.32	0	0.05
11	Only known	Only those known in 80% pedigree	0.11	0.32	0	0.20
12	Only known	Only those known in 80% pedigree	0.11	0.32	0	0.50

These simulations are designed to separate the effects of null and erroneous parentage assignments.

*Note that the null assignment rates are not directly specified as input for PEDANTICS because the reverse approach is employed. Rather, the figures quoted here are the values that occur in the various possible input pedigrees. The null assignment rate is the rate of missing parentage among nonfounding individuals.

Table A2 Biases in parameter estimates resulting from pedigree errors Soay sheep (*Ovis aries*) because of null and erroneous parentage assignments.

Scenario	Null assignment rates		Erroneous assignment rates		Bias, bivariate model		Bias, maternal model		
	Maternal	Paternal	Maternal	Paternal	h^2	r_G	h^2	m_c^2	m_a^2
(a) Varying null assignments between zero rates and rates that occur in ML pedigree:									
1	0	0	0	0	0.00	+0.01	0.00	0.00	0.00
2	0	0.21	0	0	0.00	+0.01	0.00	0.00	0.00
3	0.11	0.00	0	0	0.00	+0.01	+0.08	-0.02	-0.03
4	0.11	0.21	0	0	0.00	+0.01	+0.08	-0.01	-0.03
(b) Varying null assignments between rates that occur in the different pedigrees:									
4	0.11	0.21	0	0	0.00	+0.01	+0.08	-0.01	-0.03
5	0.11	0.32	0	0	0.00	+0.01	+0.09	-0.01	-0.04
6	0.11	0.47	0	0	0.00	+0.02	+0.11	-0.01	-0.05
(c) Varying erroneous assignment rates in the absence of null assignments:									
1	0	0	0	0	0.00	+0.01	0.00	0.00	0.00
7	0	0	0	0.05	-0.01	+0.01	-0.01	0.00	0.00
8	0	0	0	0.2	-0.02	+0.03	-0.04	+0.01	0.00
9	0	0	0	0.5	-0.05	+0.07	-0.1	+0.03	-0.01
(d) Varying erroneous assignment rates in the presence of null assignments:									
5	0.11	0.32	0	0	0.00	+0.01	+0.09	-0.01	-0.04
10	0.11	0.32	0	0.05	-0.01	+0.01	+0.08	-0.01	-0.04
11	0.11	0.32	0	0.2	-0.02	+0.02	+0.05	0.00	-0.04
12	0.11	0.32	0	0.5	-0.05	+0.06	0.00	0.00	-0.03

All true values of h^2 , m_c^2 and m_a^2 are 0.2 and all true values of r_G are 0.5. Biases are stated as the mean difference between estimates generated from 1000 replicate simulations and the true values.

and genetic correlation in our case study (Table A2a,b). However, erroneous assignments did result in a downward bias of heritability (Table A2c,d). When the two types of errors were simulated simultaneously (Fig. 2a), the biases in heritability estimates were similar to those observed when only erroneous assignments were simulated (Table A2c). This indicates that the two types of pedigree errors did not interact in their effects on estimates of heritability. Neither null nor erroneous assignments had substantial effects on estimates of genetic correlation in our system. Note that a small upward bias did occur in the estimates of genetic correlation at the highest rates of erroneous assignment (Fig. 2a and Table A2c,d); but this bias was small compared to the variability that occurs in the estimation of the genetic correlation. This bias of about 0.05 units is only in the order of about one-third of the standard error (data not shown) of these estimates. Note also that the underlying genetic covariances were underestimated (data not shown). The lack of a substantial bias in genetic correlations occurred because genetic variances and covariances were underestimated in similar proportions.

When direct additive genetic and environmental effects, as well as maternal genetic and environmental effects were simulated, more biases were detected, and these biases were larger in magnitude (Fig. 2b and Table A2). When null assignments occurred in maternal pedigree links, estimates of heritability were upwardly biased, whereas estimates of the maternal genetic effects were downwardly biased (Table A2a,b). The bias in the

estimates of the effect of maternal genes probably occurs because the estimation of maternal genetic effects requires knowledge of grandparental identities. Such data become very sparse when maternal links are incomplete and the similarities between mothers and offspring that occur because of these effects become most parsimoniously explained by direct genetic effects.

Erroneous assignments in paternal pedigree links did not appear to have important and consistent influences on estimates of maternal genetic and environmental effects (Table A2c,d). These errors did, however, influence estimates of heritability in a manner similar to their influence when only direct genetic effects were simulated (Fig. 2 and Table A2c,d). When null and erroneous assignments were varied simultaneously at the rates at which they occur in the study system, the same trends held (Fig. 2b). Estimates of heritability were overestimated when maternal null assignments occurred in conjunction with low rates of paternal erroneous assignments (Fig. 2b). In these simulations, maternal genetic effects were underestimated (main text Fig. 2b), as expected because of the results when null assignments were simulated (Table A2a,b). Ultimately, these results indicate that erroneous assignments in paternal pedigree links are of little concern for the consideration of maternal genetic and environmental effects, at least under the conditions that we simulated and in this study system. Clearly, the greatest concern regarding pedigree errors and maternal effects in this study system is the effect of maternal null assignments.