Impact of incomplete pedigree data and independent culling level pre-selection on the genetic evaluation of livestock: A simulation study on lamb growth

Samir Id-Lahoucine, Joaquim Casellas*

Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

ABSTRACT

Incomplete pedigree data due to multiple-sire mating systems is an issue for extensive sheep livestock production systems that leads to inefficiency in genetic selection programs. Although paternity testing technologies can deal with this challenge, their costs prevent the systematic testing of all offspring born, often being restricted to the replacement breeding stock (i.e., a non-random sample of individuals). This may result in important biases during genetic evaluations for traits expressed early in life, where animals can be pre-selected on the basis of their phenotypic performance, and poor performing individuals (most of them with poor genetic merit), are preferentially discarded. The potential bias due to the joint impact of this pre-selection and the specific cohort of lambs tested for missing paternity data was evaluated on simulated lamb growth data. Genetic evaluations where performed on different scenarios depending on the pre-selection criterion for the replacement breeding stock (> 150, > 200, > 250 or > 275 g/d) and availability of pedigree data. The results suggested a relevant impact on the ability of genetic evaluation models to capture the additive genetic variance (simulated heritability, $h^2=0.2$), and h^2 moved from slight ($h^2=0.225 \pm 0.004$) to severe overestimations ($h^2=0.618 \pm 0.002$) when pre-selection criterion rose from 150 g/d to 275 g/d and with sire data restricted to the replacement breeding stock. This impact was attenuated when sire data was unknown for all individuals ($h^2 < 0.25$), low levels of pre-selection, or when recovering sire data for 10–20% of discarded lambs (additional increases marginally attenuated the bias and improvements were almost absent from ~50% of the lambs). The complete loss of sire data impaired genetic evaluations and revealed moderate-to-low accuracies (r_a) for predicted breeding values. Scenarios with both moderate and high pre-selection criteria on lamb growth reported remarkable reductions on r_a when sire data was only available for replacement individuals. When sire data was also available for a percentage of discarded lambs, r_a increased, although the marginal benefit was almost negligible when paternity testing was applied to at least half of the discarded lambs. The expected genetic gain exhibited a similar behavior. As a whole, if sire data is only available for replacement individuals, pre-selection criterion must be minimal to avoid relevant biases during genetic evaluation. If not, the statistical performance of genetic evaluation procedures without sire data was similar or even better than the one obtained with sire data restricted to the replacement breeding stock.

1. Introduction

Breeding schemes have been implemented in most of the predominant livestock species worldwide (Everett et al., 1983; Hunton, 1990; Rathje, 2000), although their impact on small ruminant species still remains limited (FAO, 2007). This shortage in genetic selection programs may be due to both the idiosyncrasy of the sector or the lack of reliable pedigree data (Altarriba et al., 1998); this is of special relevance under extensive management systems where the use of multiple-sire mating systems prevents paternity assignment, unless

determined by appropriate DNA testing technologies. These tests have been proposed as a reliable alternative, although economic restrictions preclude their systematic implementation in most sheep industries. If used, they are generally restricted to the replacement breeding stock, a non-random subset of offspring which may cause important biases during genetic evaluation. When complete pedigree data is restricted to a specific subset of the progeny (and remaining individuals lack or have partial pedigree data), assumptions inherent to best linear unbiased prediction (**BLUP**; Henderson, 1973) models are violated and this may lead to relevant biases in predicted breeding values (Stock and Distl,

---

2010). If biased, predicted breeding values distort the subsequent genetic gain of any population under selection, impairing its profitability and wasting both time and efforts (Casellas and Piedrafita, 2015).

If genetic selection occurs, even if unconscious, ram- and ewe-lambs cannot be viewed as a random sample from offspring, but a subset of higher-genetic merit individuals for one or more traits of interest. This bias may reach a maximum for highly heritable traits expressed early in life such as live weight or growth. Indeed, lamb growth-related traits for meat-type production systems are very relevant (Santos et al., 2015) and poor-growth lambs are systematically pruned as selection candidates, even when the flock is not formally involved in a genetic selection program. Within this context, any replacement breeding stock must be viewed as a non-random sample of individuals where shepherds have already discarded low-performance lambs by applying, often unconsciously, the independent-culling-levels selection (Ercanbrack and Knight, 1998). In order to prevent further biases in genetic evaluations, we must be especially cautious when missing pedigree data is not distributed at random but conditioned to the selection process (Goffinet, 1987; Im et al., 1989).

Focusing in lamb growth as example, the main objective of this research was to evaluate the impact of two different potential sources of biases on the genetic evaluation of livestock when using BLUP models, i.e., (1) the amount of non-random missing pedigree data, and (2) the magnitude of independent-culling-levels selection. Our ultimate goal was to provide a set of rules to guarantee reliable genetic evaluations for current livestock industry.

2. Material and methods

Animal Care and Use Committee approval was not obtained for this study because analyses were performed on simulated data sets. Neither real animals nor biological tissues from alive animals were involved in this research.

2.1. Simulation of lamb growth data

This research focused on lamb growth, defined as the average daily gain (g/d) between birth and slaughter (or selection as replacement breeding stock). This trait is of special relevance for the sheep industry (Snowder and Van Vleck, 2003; Santos et al., 2015) and its early expression in selection candidates may lead to a variable degree of lamb selection by independent culling levels in sheep flocks (Ercanbrack and Knight, 1998). Simulations used a sheep flock of 1000 ewes with 1:30 male-to-female ratio. Animals were unselected and unrelated in the founder generation, and they were assumed to be randomly sampled from a conceptually infinite population. Each ewe was randomly mated to a single sire, lambed once per year and provided 1.5 lambs per birth. Lamb sex was assigned at random with equal probability. Annual replacement was assumed to be 20%, and the flock evolved during five years to accumulate both phenotypic and pedigree data. For each year, both male- and female-replacement individuals were randomly selected from the set of individuals satisfying the independent culling level criterion (i.e., pre-selection criterion) in terms of lamb growth (see below).

For each lamb, growth (Y_{ijkl}) was simulated under the following hierarchical model,

$$Y_{ijklm} = \mu + SX_i + EA_j + NB_k + p_l + a_m + e_{ijklm}$$

where μ was the population mean (250 g/d), SX_i was the sex of the lamb with 2 levels (i.e., male or female), EA_j was the age of the ewe with 6 levels (< 3, 3, 4, 5, 6, and > 6 years), and NB_k was the birth type with 2 levels (i.e., single and twins). All these systematic effects were simulated by sampling from a uniform distribution between −10 and 10. Random sources of variation accounted for the maternal permanent environmental effect of the ewe (p_l) and the additive genetic merit

(a_m) of the lamb, whereas e_{ijklm} was a random residual. Note that p_l, a_m (for the founders) and e_{ijklm} were assumed to be drawn from independent Gaussian distributions with mean 0 and variance equal to 500, 500 and 1,500, respectively. The genetic merit of non-founder animals was generated as the average of parents' breeding value plus the Mendelian sampling term; i.e., a random value obtained from a Gaussian distribution with mean 0 and variance equal to 250. Variance components assumed in these simulations lead to a heritability of 0.2, agreeing with the estimates provided by Mousa et al. (1999) and Bromley et al. (2000) on lamb growth.

2.2. Simulation scenarios

This research focused on two different potential sources of bias on the genetic evaluation of the breeding stock by BLUP models, (1) the amount of non-random missing pedigree data, and (2) the effect of selection by independent culling levels on lamb growth. For pedigree data, dam was assumed to be known for all non-founder individuals whereas sire was (1) unknown (**US**), (2) known for all replacement animals and unknown for slaughtered lambs (**KS0**), or (3) known for all replacement animals and a percentage of slaughtered lambs (**KS**). For this last scenario, 10%, 20%, 30%, 50%, 70% or 100% (i.e., **KS10** to **KS100**) of harvested lambs were selected at random and pedigree sire information was made available for further genetic evaluation; sire was unknown for the remaining slaughtered lambs.

Replacement breeding stock were selected at random from the subset of individuals satisfying the minimum growth (i.e., selection by independent culling level). More specifically, four different thresholds were assumed and evaluated during the simulation process, 150, 200, 250 and 275 g/d. On average, those thresholds discarded 2%, 14%, 47% and 66% of lambs as further replacement candidates, respectively. Genetic evaluations for populations simulated under each threshold for lamb growth were performed after five years of data collection, whereas only two thresholds (i.e., 200 and 250 g/d) were also evaluated after ten years of data collection. For each simulation scenario, 100 replicates were generated and analyzed.

2.3. Statistical methods for genetic evaluation

Genetic evaluations were performed through BLUP with animal models as developed by Henderson (1973). The analytical model was,

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{p} + \mathbf{Z}_2\mathbf{a} + \mathbf{e},$$

where $\mathbf{y}$ was the vector of phenotypic records, $\mathbf{b}$ was the vector of systematic effects, $\mathbf{p}$ was the vector of maternal permanent environmental effects, $\mathbf{a}$ was the vector of additive genetic effects, $\mathbf{e}$ was the vector of residual terms, and $\mathbf{X}$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ were appropriate incidence matrices. Note that this analytical model accounted for the same effects used during the simulation process (see above), with lamb sex, ewe age, and birth type as systematic sources of variation, and $\mathbf{a}$ and $\mathbf{p}$ as random effects. This model was implemented under a standard Bayesian approach where lamb growth data was assumed to be distributed under the following multivariate normal process,

$$p(\mathbf{y} \mid \mathbf{b}, \mathbf{p}, \mathbf{a}, \sigma_e^2) = MVN(\mathbf{Xb} + \mathbf{Z}_1\mathbf{p} + \mathbf{Z}_2\mathbf{a}, \mathbf{I}\sigma_e^2),$$

where $\mathbf{I}$ was an identity matrix with dimension equal to the number of records in vector $\mathbf{y}$, and σ_e^2 was the residual variance. Permanent and genetic effects were also assumed to be a priori distributed under appropriate normal processes as follows,

$$p(\mathbf{p}|\sigma_p^2) = MVN(\mathbf{0}, \mathbf{I}\sigma_p^2) \text{ and } p(\mathbf{a}|\mathbf{A}, \sigma_a^2) = MVN(\mathbf{0}, \mathbf{A}\sigma_a^2),$$

where σ_p^2 and σ_a^2 were permanent environmental and additive genetic variances, and $\mathbf{A}$ was the numerator relationship matrix (Wright, 1922). Year-by-year genetic groups were accounted for into $\mathbf{A}$ for individuals with unknown sire as described by Westell et al. (1988). Flat priors were assumed for the remaining parameters of the model