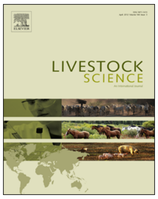
Livestock Science 181 (2015) 121–130

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



Review article

Strategies for comparing and combining different genetic and genomic evaluations: A review

J. Vandenplas [a,b,*], N. Gengler [a]

[a] Animal Science Unit, Gembloux Agro Bio-Tech, University of Liege, 5030 Gembloux, Belgium
[b] National Fund for Scientific Research, 1000 Brussels, Belgium

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form
22 August 2015
Accepted 19 September 2015

Keywords:
Conversion equation
Multiple across-country evaluation
Pseudo-records
Bayesian integration
Blending
Post-evaluation

ABSTRACT

Exchange of genetic material within and among national populations has increased rapidly with the development of artificial insemination and frozen embryos. This has increased the need to compare genetic evaluations across populations and ultimately to combine those evaluations for animals of interest. The combination of different sources of information became even more crucial with the development of genomic evaluation. This review summarizes different strategies and algorithms for solving issues related to comparison of methodology for genetic and genomic evaluations and their combination. Reviewed strategies and algorithms for genetic evaluations were categorized as either a post-evaluation or simultaneous combination approach. Post-evaluation approaches make external and internal estimates of genetic merit and their associated reliabilities comparable or combine them after performing external and internal evaluations. Simultaneous combination approaches combine external estimates of genetic merit and their associated reliabilities with internal phenotypic and pedigree data as interval evaluations are calculated. Several of the strategies developed for genetic evaluations were recently adapted for the context of genomic selection, and were mentioned in this paper.

Contents

---

[*] Corresponding author. Present address: Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands.
E-mail addresses: jeremie.vandenplas@wur.nl (J. Vandenplas), nicolas.gengler@ulg.ac.be (N. Gengler).

http://dx.doi.org/10.1016/j.livsci.2015.09.012
1871-1413/

1. Introduction

Traditionally phenotypic and pedigree data are recorded following official recording schemes for a population within well-defined borders, e.g., country borders. Because of that, they are considered "internal genetic evaluations". These internal genetic evaluations are characterized by their own scale, measurement units, and genetic bases. However, since the 1970s, technical developments such as frozen semen and embryos have increased exchange of genetic material worldwide and led to internationalization of breeding schemes and breeds. Within well-defined borders, the internal population might have a large proportion of genes from external populations. Such situations are observed for sport horses (Arnason, 2013) and dairy cattle (Philipsson, 1987).

Different issues may arise if genetic material is widespread outside borders, while genetic evaluations are calculated within borders. First, because imported (i.e., external) genetic material is usually strongly selected, internal genetic evaluations for external animals could be biased if external data used for their selection in external populations are ignored (VanRaden, 2012). Second, selection and importation of the most suitable external genetic material for internal breeding needs and goals requires the comparison of genetic evaluations. However, such comparisons usually are not possible between internal and external populations because of differences in scales, measurement units, and genetic bases for genetic evaluations (Weigel and Rekaya, 2000). One way to solve both issues is to use all phenotypic and pedigree data simultaneously (i.e., from all populations of interest) to get unbiased estimates of genetic merit (EGMs) and their associated reliabilities (RELs) through a joint genetic evaluation (e.g., Banos et al., 1992; Furre et al., 2013; Weigel and Rekaya, 2000). However, joint genetic evaluations usually cannot be calculated because data from the different populations are not available in the same dataset for various reasons, e.g., political roadblocks or data inconsistencies that prevent merges. In addition, even if data can be combined in the same dataset, joint genetic evaluations may not be possible because of computing or logistic problems (Powell and Sieber, 1992). Therefore, methods and algorithms were developed to solve both issues, and the objective of this study is to review those strategies.

Reviewed methods and algorithms were categorized as either a post-evaluation procedure or a simultaneous combination. For post-evaluation procedures, external EGMs (EGM_E) and their associated RELs (REL_E) are made comparable or combined with internal EGMs (EGM_I) and their associated RELs (REL_I) after evaluation calculation. For simultaneous combination, EGM_E and REL_E are merged with internal phenotypic and pedigree data as internal genetic evaluations are calculated. In addition, several previous strategies were adapted when the advent of genomic selection created a need to combine genomic information with phenotypic and pedigree data.

2. Post-evaluation procedures

2.1. Conversion equations

The simplest approach for comparing EGM and associated REL for two populations (e.g., from two countries) is regression-based conversion equations that convert EGM_E to the scale, units of measurement, and genetic base used for EGM_I. For dairy cattle, the first conversion equations recommended by the International Dairy Federation in 1981 (Gravert, 1983) had the form $\mathbf{y}=a+b\mathbf{x}$, where $\mathbf{y}$ is the vector of EGM_I, $\mathbf{x}$ is the vector of EGM_E, a is the intercept, and b is the slope. The intercept can be considered to be the difference between the genetic bases for the two populations. This assumption is valid as long as the genetic bases are fixed or moving at the same rate (Philipsson, 1987). If RELs are equal in both populations (i.e., the same number of observations and the same heritability) and if no genotype × environment interaction exists (i.e., genetic correlation of 1 between the two populations), the slope is the ratio of the standard deviations for EGM_E and EGM_I (Philipsson, 1987; Powell and Sieber, 1992; Wilmink et al., 1986). The slope can be considered to be the relationship (or scaling factor) between the scales and definitions for EGM_E and EGM_I, e.g., estimated breeding value (EBV) in kilograms and predicted transmitting ability (PTA) in pounds (Powell, 1988; Powell and Sieber, 1992).

The conditions previously mentioned, such as genetic bases moving at the same rate or same heritability in the different populations, generally are not found in practice, however, and several methods (e.g., Goddard, 1985; Philipsson et al., 1986; Powell, 1988; Wilmink et al., 1986) were proposed to estimate the intercept and the slope. Those methods also take into account the REL associated with the EGM from each population, and the slope then also considers genetic correlations of <1 among populations because of different heritabilities and trait definitions (Philipsson, 1987). Methods were also proposed to approximate REL associated with converted EGM based on REL_E, genetic correlations among populations, and accuracy of conversion equations (Goddard, 1985; Powell et al., 1994). Genetic correlations of <1 and preferential treatment are primarily responsible for the non-reciprocity of conversion equations (Powell et al., 1994).

Philipsson et al. (1986) defined 4 desirable properties of conversion methods that would enable comparison of equations. Methods should (1) give unbiased estimates of both the intercept and the slope, (2) consider the difference in REL from each population, (3) allow for the possibility of a genetic correlation of <1 between true genetic merit in each population, and (4) minimize the variance of differences between converted EGM and true internal values of genetic merit. Because accuracy of conversion equations is influenced by preferential matings for external animals, preferential treatment of some animals and suitability of animals selected for estimation of the intercept and slope (Powell et al., 1994), Wickham and Philipsson (1990a) made recommendations for their estimation for dairy cattle. Bulls with data to be used should (1) be born ≤ 10 years before the youngest bull, (2) have daughters in ≥ 20 herds in each population, and (3) have an REL of $\geq 75\%$ in both populations. Furthermore, the most recent data should be used, and, if a sufficient number of bulls is available, bulls initially sampled in the exporting population (i.e., according to the gene flow) should be selected for the estimation of the intercept and slope. Wickham and Philipsson (1990a) also made recommendations for cases with non-random use of bulls, a correlation between EGMs of <0.75, <20 available bulls, or few commonly used bulls. These recommendations resulted in some internal populations with too few external animals with internal progeny tests to allow conversion. Therefore, conversions of EGM were calculated using a third population with enough progeny-tested animals in common with the external and internal populations (Wickham and Philipsson, 1990b).

Conversion equations are simple and easy to apply and provide results for use in internal populations. However, estimations of the intercept and slope are based primarily on a small number of animals with EGMs in the two populations (if they are available). In addition, conversion equations can make only two populations at a time directly comparable (usually only in one direction) and may not be accurate for animals with extremely high EGM (Banos and Sigurdsson, 1996). Furthermore, relationships among animals are not taken into account by conversion equations, and external information is not propagated to relatives, which leads to a loss in