

## Avoiding double counting in genomic reliability calculations

H. Ben Zaabza<sup>1,3\*</sup>, M. Taskinen<sup>1</sup>, E.A. Mäntysaari<sup>1</sup>, T. Pitkänen<sup>1</sup>, G.P. Aamand<sup>2</sup> and I. Strandén<sup>1</sup>

<sup>1</sup> Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland; <sup>2</sup> NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark; <sup>3</sup> Department of Animal and Veterinary Sciences, University of Vermont, Burlington, VT, 05405, USA.

\*Hafedh.Ben-Zaabza@uvm.edu

### Abstract

In this study, the performance of weights for multi-step reliability calculation was examined by 4 approaches: two Interbull methods, single-trait reverse reliability (ST-REV), and multiple-trait reverse reliability (MT-REV). We concentrated on weights for genotyped animals and used 305-d milk yield data in the first 3 lactations in Nordic (Danish, Finnish, and Swedish) Holstein dairy cattle. We compared reliabilities in a classical pedigree-based animal model to the back-calculated reliabilities. Correlations between the pedigree-based animal model reliabilities and the back-calculated reliabilities were high for all approaches ( $>0.90$ ). However, gains from 0.02 to 0.11 on correlations between the classical and the back-calculated reliabilities were achieved using MT-REV compared with ST-REV and the Interbull methods. The MT-REV outperformed the ST-REV (higher correlations and smaller MSE) mainly for traits with many missing records. However, ST-REV and MT-REV yielded similar results for traits with few missing phenotype records.

### Introduction

The total number of daughters of a bull has been used as a weighting factor in some routine genetic evaluations of dairy sires. However, the effective number of daughters does not account for factors such as the number of lactations of each daughter, daughter dam information, and contemporary group (CG). Interbull proposed a procedure to calculate weighting factors to use in international genetic evaluations (Interbull, July 2000 mimeo). The Interbull procedure consists of 2 steps: 1) calculating animal model reliability based on the animal's own performance, and 2) using the reliability of progeny to calculate a weight for each bull. Fikse and Banos (2001) developed new weighting factors based on daughter information for the application in MACE. In their study, they used the number of effective daughter contributions (EDC) instead of the number of daughters. Effective record contributions (ERC) can be obtained as a linear transformation of EDC. Taskinen et al. (2014) developed an iterative method for calculating ERC: so-called reverse reliability approximation. The authors applied their method for single-trait estimation. The types of weights used for de-regressed proofs (DRP) are important when either an approximate procedure is used to regress EBV or when individual reliabilities of final EBV are computed based on the prediction error variance of the model (Calus et al., 2016). In this study, we present and test a multiple-trait reverse reliability approximation method to calculate ERCs. Our objective was to compare the performance of the different methods for calculating EDC/ERC.

### Materials & Methods

**Interbull method 1.** Interbull method 1 (IB1) uses a sire model and contains 2 steps. The first step consists of calculating reliability based on the animal's own performance records. This step accounts for the most important factors affecting reliability, including the proportion of genetic

variance, repeatability of the trait, and the CG information. Step 2 calculates a weight for each bull using the calculated reliability of its progeny. EDC provides a measure of the precision of the daughter used to compute the bull's DRP. The formula for the computation of EDC, which includes the performance of the dam of daughter  $k$  of bull  $i$ , is:  $EDC_i = \sum_k \frac{\lambda r_{k(o)}^2}{4 - r_{k(o)}(1 + r_{dam(o)}^2)}$ , where the summation is over all  $k$  daughters of the bull,  $\lambda = \frac{4 - h^2}{h^2}$ ,  $r_{dam(o)}^2$  is the reliability of the dam's own performance, and  $r_{k(o)}^2$  is the reliability of animal  $k$ 's own performance computed as:

$$r_{k(o)}^2 = \frac{n_k h^2}{1 + (n_k - 1) r_k^2}$$

where  $r_k^2$  is the reliability of the animal's records and  $n_k$  is the number of lactations of daughter  $k$  of the sire, adjusted for the CG size.

**Interbull method 2.** Interbull method 2 (IB2) is an extension of Interbull method 1 (IB1), where the animal model is used instead of the sire model. Thus,  $EDC_i = \sum_k \frac{\lambda r_{k(o)}^2}{1 - r_{k(o)}^2}$  and  $\lambda = \frac{1 - h^2}{h^2}$ .

**Single-trait reverse reliability estimation.** Single-trait ERC calculation is based on so-called reverse reliability estimation (Taskinen et al. 2014), approximating ERC from a given reliability value by reversing the Harris and Johnson (1998) method. The original algorithm estimates animal model reliability using ERC. The reverse reliability estimation estimates  $\widehat{ERC}$  that gives the same prediction error variance PEV as the original PEV\*, where the asterisk stands for the initial value of prediction error variance calculated by the Tier and Meyer (2004) method when only a subset of (genotyped) animal reliabilities and their full pedigree are available.

**Multi-trait reverse reliability estimation.** This approach reverses the Tier and Meyer (2004) method. In the original Tier and Meyer (2004) algorithm, animal model reliabilities are approximated using ERC. In the reverse estimation approach, approximate  $\widehat{ERC}$  are calculated that give the same prediction error variances PEV as the original PEV\*. When the PEV\* is from the full data with records on all animals, but ERC is solved for only a subset of (genotyped) animals, the approach automatically includes the information from the non-genotyped relatives of the genotyped animals. The calculation of  $\widehat{ERC}$  involves the accumulation of information from progeny and parents using pedigree information as in the original Tier and Meyer approach. In the Tier and Meyer reliability estimation, each animal will receive a  $n_t$  by  $n_t$  prediction covariance matrix having the diagonals (PEV) used in the estimation of the reliabilities. However, in the reverse reliability estimation, with a given  $ERC^k$  for all animals, we estimate  $PEV^k$ , and if it does not correspond to  $PEV^*$ , we search for a new  $ERC^{k+1}$  that will give  $PEV^{k+1}$  closer to  $PEV^*$ . The approach needs to be repeated iteratively. The new  $ERC^{[k+1]}$  was solved animal-wise from a non-linear equation:

$$\left\{ \left( \mathbf{G}_i^{-1} + \mathbf{E}(\mathbf{c}) \mathbf{R}_0^{-1} \mathbf{E}(\mathbf{c}) \right)^{-1} \right\}_{jj} = \{\mathbf{y}\}_j,$$

where  $\{\mathbf{y}\}_j$  are the original PEV\* of animal  $i$  and trait  $j$ ,  $\mathbf{G}_i^{-1}$  contains  $\mathbf{G}_0^{-1}$  and contributions from animal's offspring and ancestors, and  $\mathbf{R}_0$  is the between-trait residual variance-covariance matrix recorded for animal  $i$ . Animals may have missing observations for some traits, leading to missing ERC and a missing ERC pattern. For these ERC "patterns", submatrices of  $\mathbf{R}_0$  were used by omitting rows and columns corresponding to the missing observations. The ERC values are in the diagonal  $\mathbf{E}(\mathbf{c})$  matrix,  $\{\mathbf{E}(\mathbf{c})\}_{jj} = \sqrt{\mathbf{c}_j}$ , where  $\mathbf{c}_j$  is the ERC for trait  $j$  to be solved.

We implemented a Newton-Raphson based algorithm for solving the  $ERC_{rev}$ . See Ben Zaabza et al. (2022) for more details.

**Performance of different methods for calculating EDC/ERC.** The obtained ERC can be checked by back-calculating the reliability for animals by using an animal model. The model only contains an intercept and a genetic effect. The observations can have any value, and ERC are used as weights. The reliabilities for the model should be approximately the same as those used as input in ERC calculation.

**Data.** We applied the methods in Nordic (Danish, Finnish, and Swedish) Holstein Dairy cattle data, which had 8.10 million first, 5.97 million second, and 3.72 million third lactation 305-d milk yield records from 8.28 million Nordic Holstein dairy cows. The pedigree comprised up to 11 million animals, of which 274,145 were genotyped for 46,342 SNP markers.

## Results and Discussion

The correlations between approximate animal model reliabilities and back-calculated reliabilities are given in Table 1.

**Table 1. Correlation (r), maximum difference (max), and mean-squared error (MSE) between the original reliabilities and the back-calculated reliabilities using the Interbull method 1 (IB1), Interbull method 2 (IB2), single-trait reverse reliability estimation (ST\_REV), and multi-trait reverse reliability estimation (MT\_REV).**

Methods	305-d milk lactation	r	Max	MSE	b <sub>1</sub>	b <sub>0</sub>
IB1	1	0.969	0.87	0.0990	0.814	-0.002
	2	0.949	0.86	0.0194	0.737	-0.006
	3	0.937	0.87	0.0214	0.714	-0.008
IB2	1	0.921	0.99	0.0187	0.766	0.011
	2	0.900	0.99	0.0298	0.689	0.006
	3	0.883	0.99	0.0321	0.665	0.004
ST_REV	1	0.977	0.59	0.0081	0.849	0.029
	2	0.965	0.63	0.0089	0.791	0.031
	3	0.958	0.63	0.0107	0.775	0.027
MT_REV	1	0.993	0.32	0.0012	0.935	0.005
	2	0.993	0.34	0.0018	0.910	0.007
	3	0.993	0.33	0.0018	0.909	0.005

Intercept (b<sub>0</sub>) and slope (b<sub>1</sub>) of regression of original reliabilities on back-calculated reliabilities.

For Interbull method 1, correlations ranged from 0.969 for the first lactation to 0.937 for the third lactation. For Interbull method 2, however, they were between 0.921 for the first lactation

and 0.883 for the third lactation. The correlations between reliabilities from ST-REV and back-calculated reliabilities ranged from 0.977 for first lactation to 0.958 for third lactation. An improvement of 0.02 was observed in the correlations between approximate animal model reliabilities and back-calculated reliabilities from MT-REV as compared to the ST-REV method. With Interbull method 1(2), MSE ranged from 0.099(0.0187) for the first lactation to 0.0214(0.0321) for the third lactation. With ST-REV (MT-REV), MSE ranged from 0.0081(0.0012) to 0.0107(0.0018), indicating that MT-REV outperformed ST-REV, especially for lactations 2 and 3. This could be expected because there were more missing observations in second and third lactations. The MT-REV also outperformed the Interbull method 1 and 2. This can be explained by the fact that the MT-REV method allows to account for information from non-genotyped animals using the pedigree relationship structure and the multi-trait covariance structure. Quite contrary, when both the sire and its daughter are genotyped, the use of the Interbull methods will lead to a double counting of the daughter record information. The MT-REV attempts to account for that double counting of information.

### Conclusions

The performances of four methods for calculating EDC/ERC were compared in 305-d milk yield data in the first 3 lactations in Nordic Holstein dairy cattle. Our results indicate that ERC can be approximated satisfactorily for genotyped animals using a multiple-trait reverse reliability method, which outperformed the single-trait reverse reliability method mainly for traits with many missing records.

### References

- Ben Zaabza H., Taskinen M., Mäntysaari E.A., Pitkänen T., Aamand G.P., et al. (2022) J. Dairy Sci. In press.
- Calus M.P.L., Vandenplas J., ten Napel J., and Veerkamp R.F. (2016) J. Dairy Sci. 99(8): 6403-6419. DOI: [10.3168/jds.2016-11028](https://doi.org/10.3168/jds.2016-11028)
- Fikse W.F., and Banos G. (2001) J. Dairy Sci. 84(7): 1759-1767. DOI: [10.3168/jds.S0022-0302\(01\)74611-5](https://doi.org/10.3168/jds.S0022-0302(01)74611-5)
- Harris B., and Johnson D. (1998) J. Dairy Sci. 81: 2723-2728. DOI: [10.3168/jds.S0022-0302\(98\)75829-1](https://doi.org/10.3168/jds.S0022-0302(98)75829-1).
- [Interbull, July 2000 mimeo. Available at https://wiki.interbull.org/public/CoP\\_AppendixIV?action=SlideShow](https://wiki.interbull.org/public/CoP_AppendixIV?action=SlideShow)
- Taskinen M., Mäntysaari E.A., Aamand G.P., and Strandén I. (2014) Proc. of the 10th WCGALP, Vancouver, Canada.
- Tier B., and Meyer K. (2004) J. Animal Breed. Genet. 121:77-89.