Guidelines for Approximating Genomic Reliabilities of the Single-Step Genomic Model

Z. Liu, I. Strandén, J. Vandenplas, H. Eding, M. Lidauer, K. Haugaard, and P. M. VanRaden
Interbull Working Group on Genomic Reliability Calculation

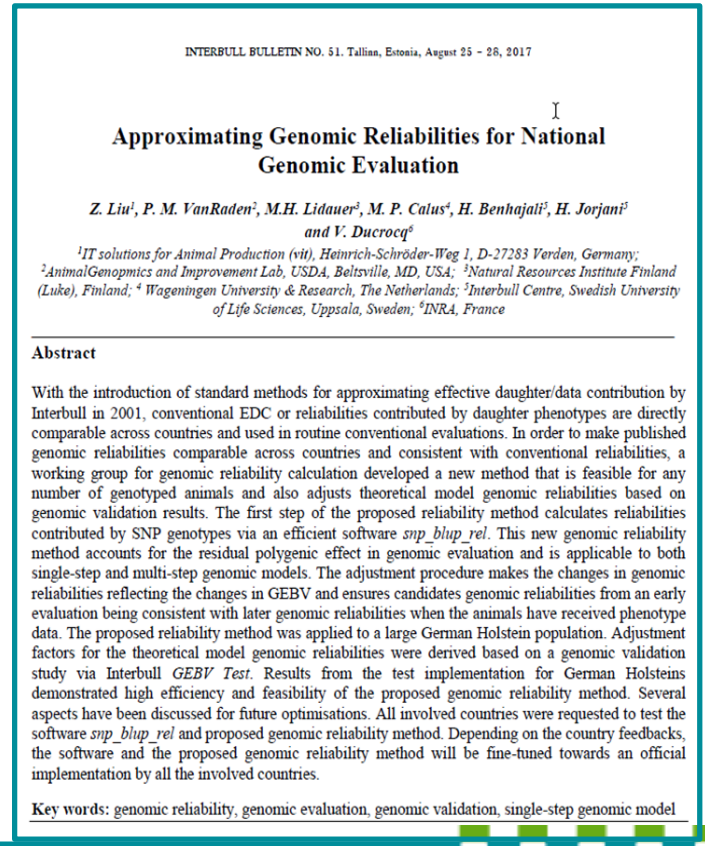
THE GLOBAL STANDARD
FOR LIVESTOCK DATA

Reliability approximation for conventional and genomic EBV

- Conventional reliability calculation for diverse genetic models
  - A single trait repeatability model (VanRaden and Wiggans, 1991)
  - A multi-trait animal model (Liu et al. 2002, 2004; Tier and Meyer 2004)
  - Proven to be fairly accurate and highly efficient for large populations

- Genomic reliability calculation methods
  - For a multi-step genomic model (Liu et al. 2010; Wiggans and VanRaden 2010)
  - For a single-step genomic model (Misztal et al. 2013)

- Interbull working group on genomic reliability calculation
  - Goal: Make national genomic reliabilities **comparable** across countries
  - A standardized genomic reliability method (Liu et al. 2017)
  - Applicable for the multi-step and single-step models
  - Large-scale female animal genotyping just started in some countries



INTERBULL BULLETIN NO. 51. Tallinn, Estonia, August 25 - 28, 2017

Approximating Genomic Reliabilities for National Genomic Evaluation

Z. Liu[1], P. M. VanRaden[2], M.H. Lidauer[3], M. P. Calus[4], H. Benhajali[5], H. Jorjani[5] and V. Ducrocq[6]

[1]IT solutions for Animal Production (vit), Heinrich-Schröder-Weg 1, D-27283 Verden, Germany; [2]AnimalGenopmics and Improvement Lab, USDA, Beltsville, MD, USA; [3]Natural Resources Institute Finland (Luke), Finland; [4] Wageningen University & Research, The Netherlands; [5]Interbull Centre, Swedish University of Life Sciences, Uppsala, Sweden; [6]INRA, France

Abstract

With the introduction of standard methods for approximating effective daughter/data contribution by Interbull in 2001, conventional EDC or reliabilities contributed by daughter phenotypes are directly comparable across countries and used in routine conventional evaluations. In order to make published genomic reliabilities comparable across countries and consistent with conventional reliabilities, a working group for genomic reliability calculation developed a new method that is feasible for any number of genotyped animals and also adjusts theoretical model genomic reliabilities based on genomic validation results. The first step of the proposed reliability method calculates reliabilities contributed by SNP genotypes via an efficient software *snp_blup_rel*. This new genomic reliability method accounts for the residual polygenic effect in genomic evaluation and is applicable to both single-step and multi-step genomic models. The adjustment procedure makes the changes in genomic reliabilities reflecting the changes in GEBV and ensures candidates genomic reliabilities from an early evaluation being consistent with later genomic reliabilities when the animals have received phenotype data. The proposed reliability method was applied to a large German Holstein population. Adjustment factors for the theoretical model genomic reliabilities were derived based on a genomic validation study via Interbull *GEBV Test*. Results from the test implementation for German Holsteins demonstrated high efficiency and feasibility of the proposed genomic reliability method. Several aspects have been discussed for future optimisations. All involved countries were requested to test the software *snp_blup_rel* and proposed genomic reliability method. Depending on the country feedbacks, the software and the proposed genomic reliability method will be fine-tuned towards an official implementation by all the involved countries.

Key words: genomic reliability, genomic evaluation, genomic validation, single-step genomic model

Features of the Interbull genomic reliability method (1)

- Main features of the Interbull genomic reliability method (Liu et al. 2017, Interbull Bulletin 51)
  - Genotype data treated as an additional source of information contributing to total reliability
  - Keep using **traditional reliability methods for conventional part** of single-step model
    - a random regression test-day model for milk production and somatic cell scores
    - a maternal-effect animal model for calving traits
    - a multi-parity, multi-trait animal model for fertility traits
  - Include young animals and all genotyped animals

Features of the Interbull genomic reliability method (2)

- Main features of the Interbull genomic reliability method

  - Consider genomic relationship among **ALL** genotyped animals
  - DGV reliabilities of young candidates depend on relationship to reference animals
  - **No longer approximating DGV reliabilities** for young candidates (Liu et al. 2010, Wiggans & VanRaden, 2010)

  - By applying the SNP BLUP model via the efficient Software **snp_blup_rel** (Strandén et al., LUKE, Finland)
  - Allow to run multiple ‘single’ traits in parallel
  - High efficiency of the software for extremely large data set with > 1.5 million genotyped animals
  - Newest version v0.99 vs current v0.88 (Ismo Strandén and Zengting Liu, April 2024)
    - DEU 25 conformation traits: 1.32 mio genotyped animals, 386,062 reference animals
    - Peak RAM reduced by 56%
    - CPU time reduced by 45%

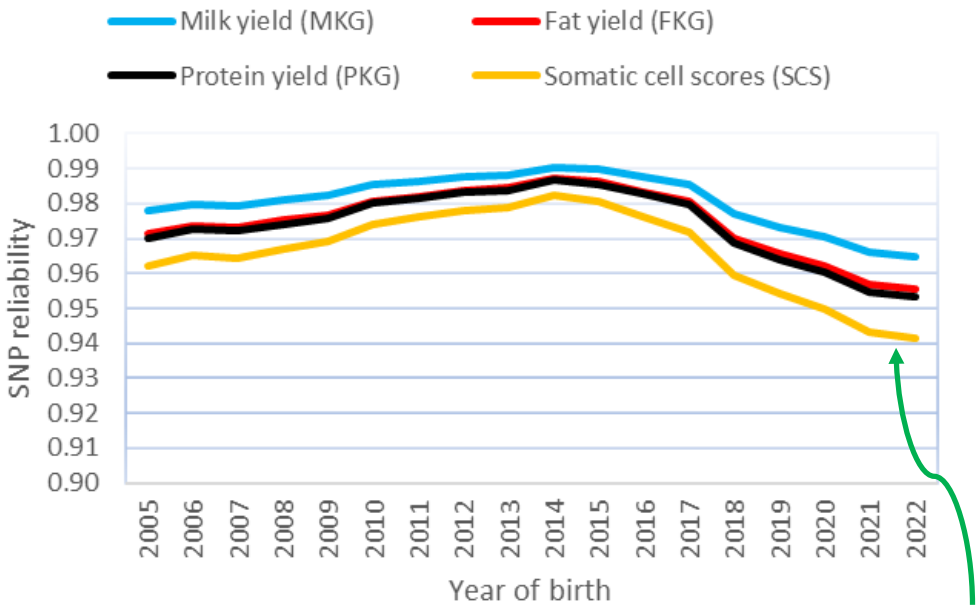
Data materials for single-step genomic reliability calculation

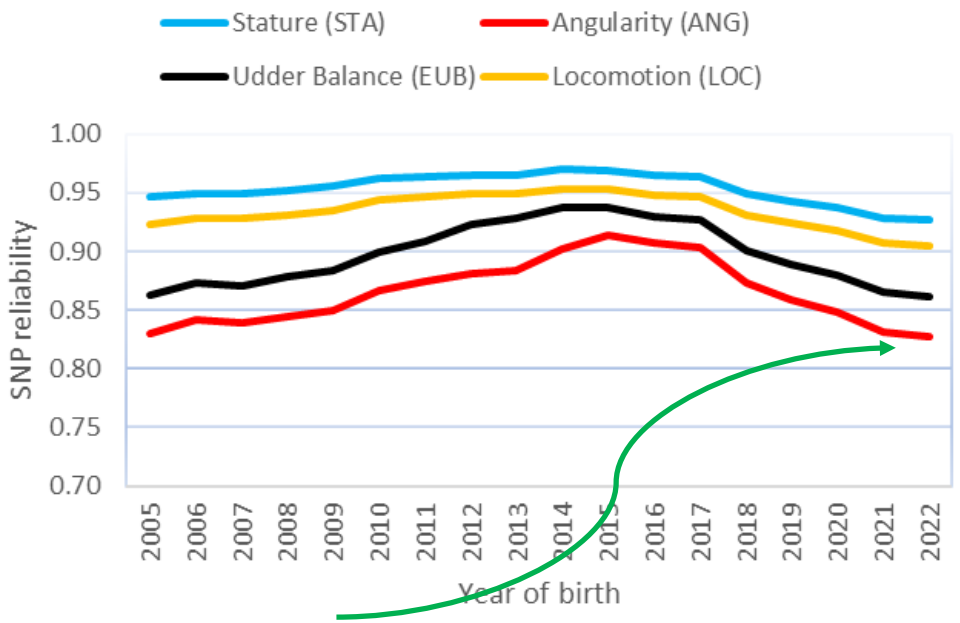
April 2023 evaluation Frequency of	Test-day traits 4 traits, e.g. protein yield	Conformation traits 25 traits in 3 sub-groups
Genotyped Holstein animals	1,318,780 (1,138,039 females and 180,741 males)	
Cows and bulls with phenotypes	13,528,444	3,144,366
--- Phenotypic records	263,673,267	3,144,366
Genotyped or phenotyped animals	14,402,662	4,131,336
Animals in pedigree	21,850,276	10,048,593
Reference animals (cows & bulls)	524,187	386,062

Genomic reference populations for April 2023 evaluation

	Protein yield PKG	Stature STA	Locomotion LOC	Angularity ANG	Udder balance EUB
Reference cows	478,588	357,365	349,083	198,170	305,122
Reference bulls	45,591	28,635	27,696	27,748	27,205
Total	524,179	386,000	376,779	225,918	332,327

Theoretical DGV reliabilities for genotyped German Holstein AI bulls in April 2023 single-step evaluation

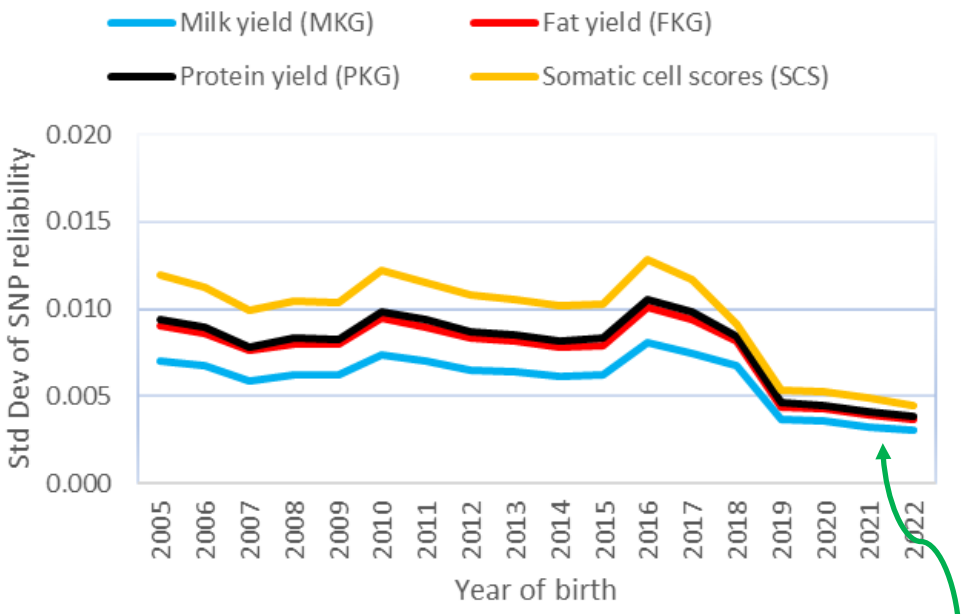


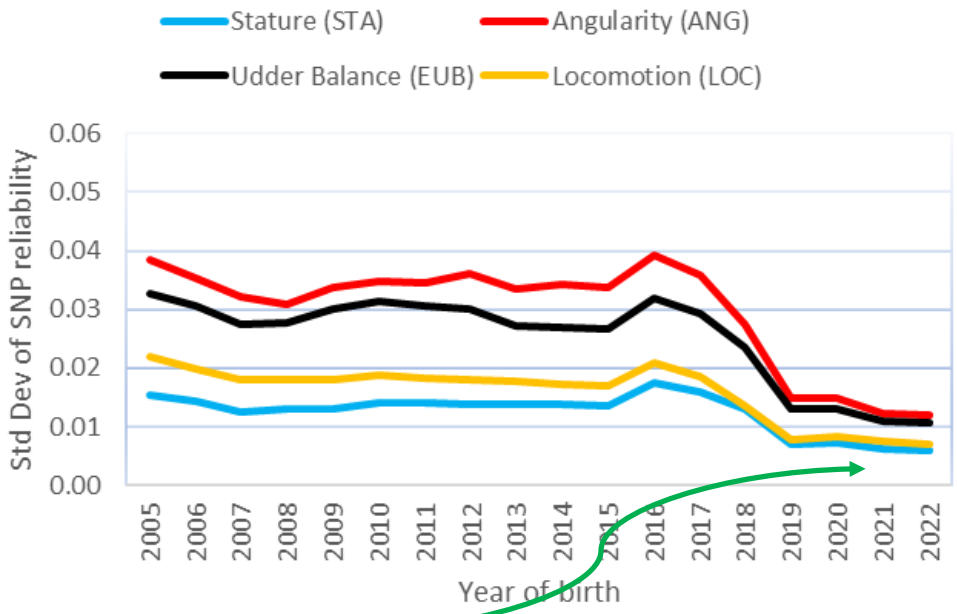


Very high theoretical DGV reliabilities for youngest AI bulls

Important to adjust genomic reliabilities based on genomic validation results

Variation in theoretical DGV reliabilities for genotyped DEU Holstein AI bulls in April 2023 single-step evaluation





Small variation of theoretical DGV reliabilities within youngest birth years

1. Less critical, if variability in individual DGV reliabilities is ignored
2. A constant genomic EDC gain may give a reasonable approximation
3. Avoid the time-consuming part of theoretical DGV reliability calculation in routine evaluation

GREL method optimization, modification and changes since 2017 (I)

- Variation in genomic reliabilities among young animals becomes smaller
  - Due to a high number of genotyped animals and more complete ancestry in reference population
  - Level of genomic reliabilities for young animals more important to ascertain
  - A constant of genomic EDC gain by genotype data to be determined

- Separation of the GREL steps between routine single-step evaluation and genomic validation
  - Too long computation for DGV reliabilities due to millions of genotyped animals
    - Large-scale female genotyping just started back in 2016/2017
    - Even for the highly efficient software ***snp_blup_rel*** (Luke, Finland)

- Two separate Guidelines for routine genomic evaluation and deriving genomic EDC gain

GREL method optimization, modification and changes since 2017 (II)

- A SNP BLUP model without a residual polygenic effect for calculating exact reliabilities of DGV
  - A posterior consideration of residual polygenic variance: REL_DGV = (1-k)*REL_SNP
    - GEBV reliability as a weighted function of REL_DGV and conventional reliability A22
    - Too many reference cows with negative genomic EDC gain particularly for traits with low heritability
  - A SNP BLUP model with RPG requires more computing time than available (Ben Zaabza, et al. 2020)

- Reducing SNP markers by selecting equidistant markers for faster calculation of DGV reliabilities
  - At least 15,000 SNP markers were shown to be needed
  - Similar effect may be achieved by the adjustment of genomic reliability

- Propagated genomic reliabilities for non-genotyped relatives of genotyped animals following the concept of genotype confidence (Eding, 2022)
  - Instead of using a fixed value of upper limit

- The new GEBV test software 2024A (Sullivan, 08.05.2024) provides for validation bulls:
  - $var(\hat{\mathbf{u}}_{\mathrm{E}} - \hat{\mathbf{u}}_{\mathrm{L}})$

Two Guidelines for single-step genomic reliability calculation and adjustment

Guideline for GREL Calculation

Guidelines for Approximating Genomic Reliabilities of the Single-Step Genomic Model

Z. Liu, I. Strandén, J. Vandenplas, H. Eding, M. Lidauer, K. Haugaard, and P. M. VanRaden

Interbull Working Group on Genomic Reliability Calculation

A genomic reliability method (Liu et al., 2017) developed by the Interbull Working Group approximates reliabilities of estimated genomic breeding values (GEBV) for a multi-step or a single-step genomic model. Several modifications and improvements have been made thereafter. This document describes technical details of the calculation of genomic reliabilities (GREL) of the single-step genomic model.

The Interbull GREL method assumes that Interbull member countries applies an accurate method to calculating pedigree-based conventional reliabilities, by ignoring genotype data, for either a single-trait repeatability model (VanRaden and Wiggans, 1991) or a multi-trait animal model (Liu et al. 2004; Tier and Meyer 2004) such as a random regression test-day model for milk production traits or a maternal-effect model for calving traits. Besides animals with own phenotypic records, genotyped animals without own phenotypic records must also be included in the calculation of the conventional reliabilities.

Required data for approximating genomic reliabilities using the Interbull GREL method are:

1) A pedigree file which is used for the single-step genomic evaluation of an evaluated trait or a linear index of evaluated traits. The pedigree file must be sorted from the oldest to the youngest animals (or in the opposite order) and should include both genotyped and ungenotyped animals,
2) An estimate of the heritability (h^2) of the evaluated trait or index of interest,
3) Pedigree-based conventional reliability values of all animals in the pedigree file, including genotyped animals without own phenotypic records, for the evaluated trait or index of the evaluated traits, and
4) Genomic effective daughter contribution (EDC) gain (φ_c) for the evaluated trait or index of the evaluated traits, which was derived by the countries following the Interbull GREL procedure (see Appendix for the *Guidelines for Deriving Genomic Effective Daughter Contribution Gain*).

Technical steps for calculating the final GREL for genotyped and ungenotyped animals are given below:

1. *Propagation of genomic information of the genotyped animals to their non-genotyped relatives*
   In the propagation process the trait-specific constant of the genomic EDC gain φ_c is treated as weight on genotypic data for each of the genotyped animals to approximate genomic reliabilities of their non-genotyped relatives. The propagation involves two steps (VanRaden and Wiggans, 1991; Liu et al. 2004): 1) accumulating progeny contribution by passing the genomic information φ_c of the genotyped animals to their non-genotyped

Guideline for GREL Adjustment

Appendix: Guidelines for Deriving Genomic Effective Daughter Contribution Gain

Z. Liu, I. Strandén, J. Vandenplas, H. Eding, M. Lidauer, K. Haugaard, and P. M. VanRaden

Interbull Working Group on Genomic Reliability Calculation

The Interbull genomic reliability method (Liu et al., 2017) has been optimised to make the genomic reliability calculation feasible for routine single-step genomic evaluations with millions of genotyped animals (see the *Guidelines for Approximating Genomic Reliabilities of the Single-Step Model*). A parameter, called hereafter genomic effective daughter contribution gain (φ_c) and required by the Interbull genomic reliability method, must be derived for every trait evaluated by the Interbull member countries.

Conventional reliability values are assumed to be reasonably accurate using an accurate reliability method for a single-trait model like VanRaden and Wiggans (1991) and a multi-trait model like Liu et al. (2004) or Tier and Meyer (2004).

Genomic breeding values (GEBV) of a single-step evaluation using the full phenotypic, genotypic and pedigree data as well as GEBV of an early single-step evaluation using a sub-set of the phenotypic data are needed. According to VanRaden and O'Connell (2018), following data are required for deriving the genomic EDC gain parameter φ_c:

1) A pedigree file (PED_{full}) that is used for single-step evaluation using the full phenotypic and genotypic data. This pedigree file should also include genotyped animals without own phenotypic records;
2) An extracted pedigree file containing only genotyped animals and their ancestors (PED_{geno});
3) Heritability value (h^2) of the evaluated trait or a linear index of breeding values of evaluated traits and variance ratio of the animal model $\lambda = \frac{1-h^2}{h^2}$;
4) Conventional reliability values of all animals, including genotyped animals without own phenotypic records;
5) A file containing effective daughter contribution (EDC) of genotyped bulls and/or effective record contribution (ERC) of genotyped cows. When a genotyped cow with phenotypic records and her sire are both genotyped, her sire's EDC must be adjusted for her contribution to avoid a double counting of her own phenotype information. Interbull proposed an adjustment method for EDC of bulls and technical details of the EDC adjustment are given in Interbull (2018);
6) A list of genotyped animals for the single-step evaluation;
7) A file of allele frequencies for all SNP markers used in the genomic evaluation;
8) A SNP genotype file for all the genotyped animals containing ID of the animals and genotype string of all the SNP markers;
9) A list of validation bulls for *Interbull GEBV test* (Mäntysaari et al. 2010); and
10) GEBV of the validation bulls from the single-step evaluation with the full data set and from the early evaluation with the truncated subset of data.

Approved by Interbull Steering Committee

When a genomic EDC constant needs to be updated / GREL to be adjusted

- A constant value for genomic EDC gain is used for all genotyped animals
  - Properly determine the level of genomic reliability especially for young candidates
  - Ignores differences in DGV reliabilities among genotyped animals
    - Small difference for large reference population
  - For smaller populations like Jersey or a new trait
    - Low number of validation bulls
    - Special rules may be needed

- Updating the genomic EDC gain parameter whenever a GEBV test is required
  - Implementation of a new model
  - Introduction of major model changes
  - Routine genomic validation every 2 years

- Apply the same rules of data truncation of the GEBV test for the updating

Too high genomic reliabilities in case of inflation $b_1 < 1$?

- GEBV test based on validation bulls

$$\hat{\mathbf{u}}_{\mathrm{L}} = b_0 + b_1 \hat{\mathbf{u}}_{\mathrm{E}} + \varepsilon$$

- Regression and correlation coefficients: $b_1 = r \frac{\sqrt{var(\hat{\mathbf{u}}_{\mathrm{L}})}}{\sqrt{var(\hat{\mathbf{u}}_{\mathrm{E}})}}$ => $var(\hat{\mathbf{u}}_{\mathrm{L}}) = \frac{b_1^2}{r^2} var(\hat{\mathbf{u}}_{\mathrm{E}})$
- GREL adjustment according to GEBV from the full/later and truncated/early evaluations

$var(\hat{\mathbf{u}}_{\mathrm{E}} - \hat{\mathbf{u}}_{\mathrm{L}}) = var(\hat{\mathbf{u}}_{\mathrm{L}}) - (2b_1 - 1) var(\hat{\mathbf{u}}_{\mathrm{E}})$ versus $var(\hat{\mathbf{u}}_{\mathrm{L}}) - var(\hat{\mathbf{u}}_{\mathrm{E}})$

- $b_1 = 1$ suggesting no under- or overestimation, $var(\hat{\mathbf{u}}_{\mathrm{E}} - \hat{\mathbf{u}}_{\mathrm{L}}) = var(\hat{\mathbf{u}}_{\mathrm{L}}) - var(\hat{\mathbf{u}}_{\mathrm{E}})$
- $b_1 < 1$ suggesting $var(\hat{\mathbf{u}}_{\mathrm{E}})$ too high, $var(\hat{\mathbf{u}}_{\mathrm{E}} - \hat{\mathbf{u}}_{\mathrm{L}}) > var(\hat{\mathbf{u}}_{\mathrm{L}}) - var(\hat{\mathbf{u}}_{\mathrm{E}})$, expected average reliability of early evaluation $E(\mathfrak{R}_E)$ would be lower than the case of $b_1 = 1$
- $b_1 > 1$ suggesting $var(\hat{\mathbf{u}}_{\mathrm{E}})$ too low, $var(\hat{\mathbf{u}}_{\mathrm{E}} - \hat{\mathbf{u}}_{\mathrm{L}}) < var(\hat{\mathbf{u}}_{\mathrm{L}}) - var(\hat{\mathbf{u}}_{\mathrm{E}})$, expected average reliability of early evaluation $E(\mathfrak{R}_E)$ would be higher than the case of $b_1 = 1$

- No, the adjusted genomic reliabilities will NOT be too high, when $b_1 < 1$

Implementation issues to consider

- Applicability for small genotyped / reference populations
  - For a new trait, like dry matter intake
  - For a small breed, like Jersey
  - Next project for the WG
- Same GREL adjustment for all sub-traits of a trait group?
  - Sub-traits have similar data and heritability values
- A multi-breed genomic evaluation system
  - e.g. Holstein + Jersey
  - Breed-specific genomic reliability adjustment
- Interbull Mendelian Sampling variance test for the single-step model
  - Use genomic reliabilities instead of conventional reliabilities

Future research topics

- Extension to a SNP BLUP model with a residual polygenic effect (Ben Zaabza, et al. 2020; 2021)
  - A Monte Carlo sampling based approach
  - To further improve the computational efficiency
- Multi-trait models for modelling genomic information at ALL the steps
  - Currently, a single trait model is assumed for the genomic information
- Fewer steps for single-step genomic reliability approximation
  - Conventional reliabilities using a multi-trait or single-trait model
  - Divide the whole population into genotyped and non-genotyped sub-populations
    - Quantify ‘added value’ of genotyping based on reference population
  - Propagate the genomic information gain to the non-genotyped relatives
  - Adjust the theoretical genomic reliability levels using a truncated evaluation
- Merge the steps of conventional reliability calculation with those of genomic reliability
  - Explore the structure of LHS of the MME of the single-step model

Summary and Conclusions (I)

- The two GREL Guidelines for routine evaluation and validation
  - Developed by WG and approved by the Steering Committee
  - Separate Guidelines for routine evaluation and for deriving GREL adjustment via GEBV test
  - Feasible for large populations with millions of genotyped animals in routine evaluation
  - Ensure a realistic level of genomic reliabilities of young candidates
- DEU implemented the Guidelines in all 10 trait groups
  - Using data from a full evaluation April 2023 and a truncated data set
- Update genomic EDC gain parameter whenever a GEBV test is required
  - Adjusting genomic reliabilities is linked to the GEBV test
- Implementation issues like for small populations to be considered
- Future R&D projects for further optimization of the Interbull GREL method

Summary and Conclusions (II)

- All member countries are encouraged to apply the Interbull GREL method
- The Interbull GREL WG can provide technical support
- We are one step closer to our goal:
  - To make genomic reliabilities comparable across countries
  - To make genomic reliabilities comparable between traits within country
  - To make genomic reliabilities comparable between candidates and bulls with many daughters

Thank you for your attention!

IT-Solutions for Animal Production