



Validation of consistency of Mendelian sampling variance

A.-M. Tyrisevä,^{*1} W. F. Fikse,^{†2} E. A. Mäntysaari,^{*} J. Jakobsen,[‡] G. P. Aamand,[§] J. Dürr,[#] and M. H. Lidauer^{*}

^{*}Natural Resources Institute Finland, Green Technology, Biometrical Genetics, 31600 Jokioinen, Finland

[†]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden

[‡]Norwegian Association of Sheep and Goat Breeders, 1431 Ås, Norway

[§]NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus, Denmark

[#]Council on Dairy Cattle Breeding, Bowie, MD 20716

ABSTRACT

Experiences from international sire evaluation indicate that the multiple-trait across-country evaluation method is sensitive to changes in genetic variance over time. Top bulls from birth year classes with inflated genetic variance will benefit, hampering reliable ranking of bulls. However, none of the methods available today enable countries to validate their national evaluation models for heterogeneity of genetic variance. We describe a new validation method to fill this gap comprising the following steps: estimating within-year genetic variances using Mendelian sampling and its prediction error variance, fitting a weighted linear regression between the estimates and the years under study, identifying possible outliers, and defining a 95% empirical confidence interval for a possible trend in the estimates. We tested the specificity and sensitivity of the proposed validation method with simulated data using a real data structure. Moderate (M) and small (S) size populations were simulated under 3 scenarios: a control with homogeneous variance and 2 scenarios with yearly increases in phenotypic variance of 2 and 10%, respectively. Results showed that the new method was able to estimate genetic variance accurately enough to detect bias in genetic variance. Under the control scenario, the trend in genetic variance was practically zero in setting M. Testing cows with an average birth year class size of more than 43,000 in setting M showed that tolerance values are needed for both the trend and the outlier tests to detect only cases with a practical effect in larger data sets. Regardless of the magnitude (yearly increases in phenotypic variance of 2 or 10%) of the generated trend, it deviated statistically significantly from zero in all data replicates for both cows and bulls in setting M. In setting S with a mean of 27 bulls in a

year class, the sampling error and thus the probability of a false-positive result clearly increased. Still, overall estimated genetic variance was close to the parametric value. Only rather strong trends in genetic variance deviated statistically significantly from zero in setting S. Results also showed that the new method was sensitive to the quality of the approximated reliabilities of breeding values used in calculating the prediction error variance. Thus, we recommend that only animals with a reliability of Mendelian sampling higher than 0.1 be included in the test and that low heritability traits be analyzed using bull data sets only.

Key words: international sire evaluation, validation, trend, genetic variance

INTRODUCTION

Dairy cattle breeding is a global business, and the highest-ranking bulls are sold worldwide. Therefore, reliable estimation of the genetic merit of dairy bulls originating from different populations and production environments is of fundamental importance. International breeding values are currently obtained with a multiple-trait across-country evaluation (MACE) method, which uses deregressed breeding values from national genetic evaluations as observations (Schaeffer, 1994). The MACE method considers one biological trait at a time and accounts for genotype \times environment interactions by assuming observations from different countries as different but genetically correlated traits.

Experience has shown that the MACE method is sensitive to the quality of the national evaluations. Several studies have demonstrated that biased genetic trends and genetic variance trends in national evaluations affect the MACE evaluations. Top bulls from birth year classes with inflated genetic variances and bulls from countries with overestimated genetic trends in their national evaluation benefit from an upward bias in ranking (Weigel et al., 1996; Van Doormaal et al., 1999; Gengler et al., 2000; Miglior et al., 2002; Ducrocq et al., 2003).

Received May 30, 2017.

Accepted November 7, 2017.

¹Corresponding author: anna-maria.tyriseva@luke.fi

²Current address: Växa Sverige, 75105 Uppsala, Sweden.

To ensure unbiased international evaluation, all participating countries are required to validate their national evaluations. The validation methods (Boichard et al., 1995) test whether national evaluations give unbiased genetic trend estimates, but the homogeneity of genetic variance across years has so far not been tested. There is an increasing need to test for homogeneity of genetic variance, which is one of the basic assumptions of the evaluation.

The main cause of heterogeneity of variance (**HV**) is a scaling effect: variance changes with a change in the phenotypic mean (Meuwissen et al., 1996; Robert-Granié et al., 1999; Lidauer et al., 2008). This is observed as animal production levels increase over time or because production levels vary among herds, which may be due to factors such as different management practices and geographical conditions, such as smaller herd sizes in mountain areas with harsh environments versus high-input systems in lowlands (e.g., Lidauer et al., 2008). Today, more and more countries are combining their breeding populations and national genetic evaluations. This increases HV due to differences in production environments and management practices among cows included in the same genetic evaluation. Other reasons for HV are related to factors such as selection, use of different breeds and genetic groups in the same evaluation, and different mating practices (Meuwissen et al., 1996). Precorrection for heterogeneous variance may not be sufficient in more complex cases, which require more elaborate methods (Meuwissen et al., 1996; Robert-Granié et al., 1999; Lidauer et al., 2008, 2015). Currently, countries follow varying practices. Yearly standard deviations (**SD**) of EBV offer a simple option for studying trends in genetic variance because any changes in genetic variance are generally reflected in SD. However, restricted maximum likelihood estimation is a more preferable alternative because within-year SD of EBV are sensitive to factors such as temporal fluctuations in the average number of daughters per bull, number of bulls in different age classes, changes in the genetic structure of a population, and reliability of breeding values (e.g., Miglior et al., 1998; Van Doormaal et al., 1999; Miglior et al., 2002).

Sullivan (1999) suggested a restricted maximum likelihood method for estimating within-year genetic variances by deriving an equation based on Mendelian sampling (**MS**) and its prediction error variance (**PEV**). Using Sullivan's method, Miglior et al. (2002) outlined a new validation test for the International Bull Evaluation Service (Interbull). Because the computation of PEV from large data sets is not feasible, Fikse et al. (2003) proposed a procedure that uses approximated reliabilities of EBV of animals and their parents. Later,

Fikse et al. (2005) introduced a framework for obtaining lower and upper bounds of the tolerance interval; the test then boiled down to counting the number of years for which the across-year estimate of genetic variance was outside the tolerance interval. The empirical tolerance interval for a birth year class was obtained by bootstrapping data within that birth year class and finding the 0.025 and 0.975 quantiles for the lower and upper genetic variance estimates. Any inaccuracies in the approximation of PEV values were taken into account in the lower and upper genetic variance estimates by applying a multiplication factor of 0.99 (lower) and 1.02 (upper) to the approximated PEV. The sensitivity and specificity of the test were unknown, calling for further research. When the procedure was tested on field data, some inconclusive results were obtained. Testing with simulated data sets gave reliable results for large cow data sets, but the method failed to detect a generated trend in genetic variance for bulls from small birth year classes (Tyrisevä et al., 2011).

Further, Lidauer et al. (2007) developed a full model sampling method (**FMS**) to estimate within-year genetic variances. Although this FMS method and the method by Fikse et al. (2003) differ in their way of estimating the PEV of MS, they yield relatively similar results (Lidauer et al., 2007). The FMS approach requires simulation of new observations according to the model used in the national evaluation system and therefore is not easy to implement in a scheme with a wide variety of national evaluation models. Based on experiences from earlier studies, our aim was to develop and test a validation method for HV of MS that would be applicable for national evaluations.

MATERIALS AND METHODS

Estimation of Genetic Variance

A univariate animal model can be described as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where \mathbf{y} is a vector of records, \mathbf{b} is a vector of fixed effects, \mathbf{u} is a vector of random animal effects, and \mathbf{e} is a vector of random residuals. Incidence matrices \mathbf{X} and \mathbf{Z} relate the records to the appropriate effects. Further, we assumed that $\mathbf{u} \sim MVN(\boldsymbol{\mu}_u, \mathbf{A}\boldsymbol{\sigma}_u^2)$, with the expected value of breeding values defined as $\boldsymbol{\mu}_u$ and covariance among breeding values defined as $\mathbf{A}\boldsymbol{\sigma}_u^2$. The breeding value of animal i can be further expressed as $u_i = 1/2(u_s + u_d) + m_i$, where u_s and u_d are the breeding values of the sire and dam of animal i , and m_i is the MS of animal i . Accordingly, the relationship matrix \mathbf{A} can be

decomposed into $\mathbf{A} = \mathbf{LDL}'$, where \mathbf{L} traces the flow of genes from parents to offspring and \mathbf{D} is the diagonal matrix of the variance related to MS m (Mrode, 1996). Thus, assuming that parents s and d are unrelated and noninbred, the genetic variance can be formulated as

$$\text{var}(u_i) = \text{var}\left[\frac{1}{2}(u_s + u_d) + m_i\right] = \frac{1}{4}\sigma_u^2 + \frac{1}{4}\sigma_u^2 + \frac{1}{2}\sigma_u^2,$$

showing that the MS variance is one half of the genetic variance. Within-year genetic variance $\sigma_{u_j}^2$ can be estimated by using MS and its PEV, as shown by Sullivan (1999), which is applied to the validation method explained in this study:

$$\sigma_{u_j}^2 = \frac{\sum_{i=1}^{q_j} d_i \hat{m}_i^2}{q_j - \sum_{i=1}^{q_j} d_i PEV(\hat{m}_i)}, \quad [1]$$

where q_j is the number of animals in year j , d_i is the inverse of the proportion of genetic variance not explained by the known parents of animal i , \hat{m}_i^2 is the squared estimated MS of animal i , and $PEV(\hat{m}_i)$ is the PEV of MS. The effect of inbreeding, when applied, was accounted for by modifying the coefficient d_i according to the parents' inbreeding. As the exact values for $PEV(m_i)$ are difficult to obtain, Fikse et al. (2003) proposed to approximate $PEV(\hat{m}_i)$ following Misztal and Wiggans (1988) and Misztal et al. (1991). Given the reliabilities of EBV, p_{ani} values, where ani refers to the sire, the dam, or the animal itself, can be calculated by defining the information content embedded in sire s (p_s), dam d (p_d), and animal i (p_i), assuming that the parents are unrelated:

$$\begin{aligned} & \begin{bmatrix} 1.5\alpha + p_s & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5\alpha + p_d & -\alpha \\ -\alpha & -\alpha & 2\alpha + p_i \end{bmatrix}^{-1} \\ & = \begin{bmatrix} 1/(\alpha + b_s) & \bullet & \bullet \\ \bullet & 1/(\alpha + b_d) & \bullet \\ \bullet & \bullet & 1/(\alpha + b_i) \end{bmatrix} \end{aligned} \quad [2]$$

where α is the residual to genetic variance ratio, \bullet denotes elements not known (3 unknowns due to symmetry), and

$$b_{ani} = \alpha \times REL_{EBV_{ani}} / (1 - REL_{EBV_{ani}}),$$

where $REL_{EBV_{ani}}$ is EBV reliability. After solving the nonlinear equation to obtain the p_{ani} values, $PEV(\hat{m}_i)$ can be approximated as

$$\begin{aligned} PEV(\hat{m}_i) &= \begin{bmatrix} -0.5 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} 1.5\alpha + p_s & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5\alpha + p_d & -\alpha \\ -\alpha & -\alpha & 2\alpha + p_i \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} -0.5 \\ -0.5 \\ 1 \end{bmatrix} \sigma_e^2 = C_i \sigma_e^2, \end{aligned} \quad [3]$$

where C_i is the prediction error coefficient of MS and σ_e^2 is the residual variance (Fikse et al., 2005).

Statistical Test for the Trend

A weighted linear regression of within-year genetic variances ($\hat{\sigma}_{u_j}^2$) on year (x_j) was fitted to study the existence of a possible trend:

$$\hat{\sigma}_{u_j}^2 = \alpha + \beta x_j + e_j,$$

in which α and β are regression coefficients and e_j is a residual term. The number of animals in each year class was used as the weight. The trend is expressed as a percentage relative to genetic variance ($\beta / \bar{\sigma}_u^2 \times 100\%$), where $\bar{\sigma}_u^2$ is the estimated genetic variance averaged over years. An empirical 95% confidence interval for the trend (β) was calculated by bootstrapping 1,000 samples with case resampling within year classes. For each bootstrap sample, the sampled animals were used to calculate yearly variances, after which the above regression model was fitted and the sample $\hat{\beta}$ was estimated. A 95% confidence interval was obtained by defining 0.025 and 0.975 quantiles for the bootstrapped $\hat{\beta}$. If the confidence interval did not include zero, the trend was considered to deviate statistically significantly from zero.

Statistical Test for Outliers

We applied an outlier test to detect years with possible outlier estimates of genetic variance that did not fit the linear trend model and to find indications of a possible nonlinear trend. Residual terms obtained from the regression model fitted on the 1,000 bootstrap samples were used to detect outliers. A 95% confidence interval for each e_j was obtained by finding the 0.025 and 0.975 quantiles of the bootstrapped residuals. We

used a Bonferroni correction for the N independent tests, N being the total number of years included in the analyses. If the confidence interval did not include zero for some year, then the variance estimate for this year was considered a statistical outlier.

Sensitivity and Robustness of the Validation Method

A simulation study was carried out to assess the conditions necessary for the validation method to have the required sensitivity but also sufficient robustness. The aims were (1) to assess the accuracy of the estimated genetic variance by validating it against a reference method, (2) to assess the sensitivity to accuracy of approximation of MS reliabilities, (3) to evaluate the effect of inbreeding on the estimates, and (4) to determine the sensitivity of the final validation test (MS variance test; **MSVT**) to detect HV. The simulations (Figure 1) were performed under 3 scenarios: 1 scenario with homogeneous variance (control; **C**) and 2 scenarios with phenotypic variance increasing over time by either 2% (**T2**) or 10% (**T10**) per year. To test the performance of the method on scenarios closely resembling field data conditions, the simulated scenarios were based on 2 real data sets (moderate and small size) and a currently used evaluation model. For each tested scenario we simulated 20 independent data replicates.

Creating Data Sets Representing Moderate and Small Size Populations. Cows with protein yield test-day observations from Danish Holstein herds were sampled for yield traits from the joint Nordic test-day data, excluding 5% of the largest herds as a first editing

step. To create a data set representing a moderate-size national population (**M**), herds were required to have at least 6 heifers calving in a calving year for at least a 10-yr period starting from 1991. The sampled cows had to have at least 1 observation during the first 70 DIM in first parity. A maximum of 3 parities were included for a cow. Of the 5,253 herds fulfilling these requirements, we randomly sampled 2,000 herds, comprising 13 million protein yield test-day observations of 756,537 cows and covering a time interval of 22 yr. The pruned pedigree included 1.2 million animals. The average number of animals per birth year class was 43,605 for cows and 290 for bulls.

To study the performance of the validation method for a data set representing a small bull population, we created a small data set (**S**) by sampling all sires having at least 1,000 daughters in setting M and about 8% of sires with fewer than 1,000 daughters in setting M. All the daughters of these sires as well as the dams of these daughters were selected for the subsample, resulting in a similar daughter distribution as in setting M but with 30% less data. Thus, setting S comprised 9 million protein yield test-day observations of 483,410 cows over a 22-yr time interval. The pruned pedigree included 0.9 million animals. The average and minimum birth year class sizes for bulls were 27 and 16, respectively.

Generating Phenotypes. Data replicates for the simulation study were formed by replacing original field data observations with simulated protein yield observations. The simulated observations were based on the evaluation model and variance components used in the Nordic test-day model relevant for Danish protein yield

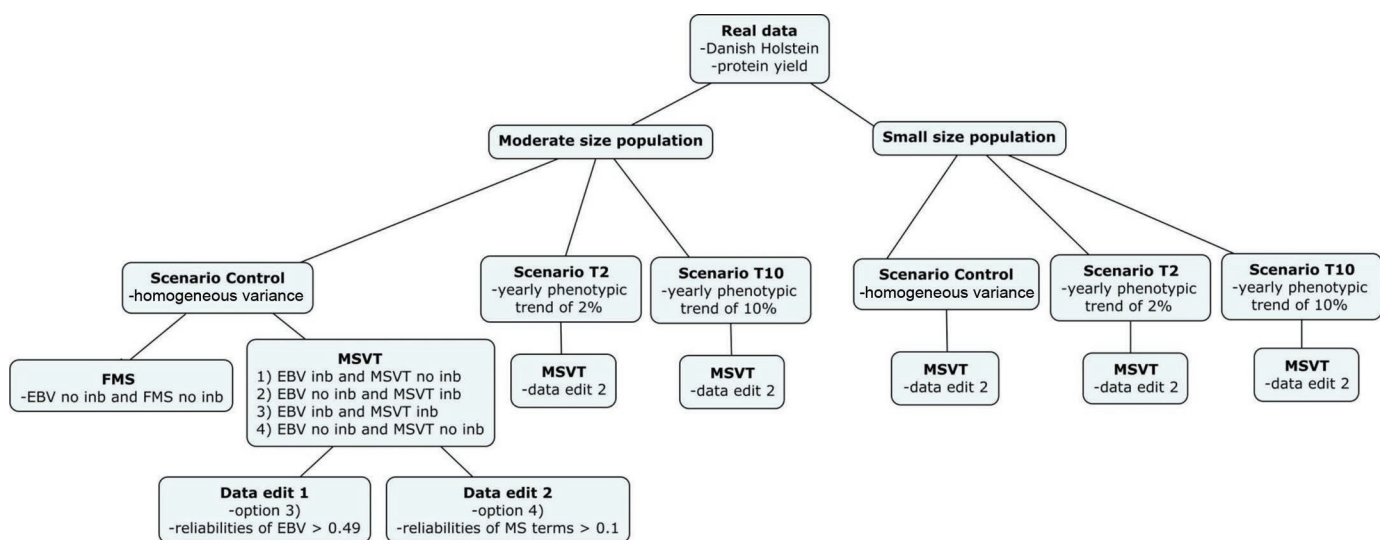


Figure 1. Design of the simulation study. Inb (no inb) refers to modeling (not modeling) of inbreeding in the prediction of breeding values and estimation of within-year genetic variances. For each tested scenario, 20 independent data replicates were simulated. FMS = full model sampling; MS = Mendelian sampling; MSVT = Mendelian sampling variance test. Color version available online.

observations. The applied multiple-trait random regression test-day model included fixed effects, a random herd test-day effect, a random regression on DIM nested within herd, covariance functions for nonadditive genetic and additive genetic animal effects, and a random residual effect, as explained in Lidauer et al. (2015). Observations (\tilde{y}) were formed by summing for each observation all the corresponding simulated model effects and adding a random error term (Lidauer et al., 2007, 2008). The simulated true fixed effects and genetic group effects were from preliminary evaluation runs for both M and S settings using original field data observations, whereas the simulated true values for all random factors in the model were generated from normal distributions by using the Cholesky decomposition of the variance-covariance matrices of the applied variance components. In simulating the true additive genetic effects, the covariance structure between animals was accounted for by utilizing the Cholesky decomposition of the numerator relationship matrix following Matilainen et al. (2012). Random residuals were drawn from a multivariate normal distribution assuming that residuals were uncorrelated across lactation and across parities. To accomplish a yearly increase in phenotypic variance over time for scenarios T2 and T10, we applied production month-specific scaling factors to the generated observations.

EBV and Their Reliabilities. Combined 305-d breeding values across the first 3 lactations were used for validation of the new method. These were formed by deriving 305-d breeding values from the cows' covariance functions for additive genetic effects for each lactation and by calculating the weighted sum over lactations using weights of 0.5, 0.3, and 0.2 for first, second, and third lactations, respectively (NAV, 2017). The genetic variance of the constructed 305-d combined protein yield was 247 kg², and the corresponding heritability was 0.35. We used MiX99 software in all steps of the simulations (MiX99 Development Team, 2016). Breeding value reliabilities were approximated according to Misztal and Wiggans (1988) and implemented in MiX99.

Assessing the Accuracy of Estimated Genetic Variance. To assess the accuracy of the variance estimates, we compared the estimates of within-year genetic variances obtained with the MSVT with those obtained with the FMS method (Lidauer et al., 2007). The latter is based on the idea of García-Cortés et al. (1995) to use resampling to calculate PEV, which is often impossible to calculate exactly. For each replicate of the simulation, we therefore resampled r independent data sets having the exact same structure as the original data set of that replicate. The true and estimated

MS deviations from each resampled data set were used to calculate PEV. Equation 1 can be reformulated as

$$\hat{\sigma}_{u_j}^2 = \frac{1}{q_j} \sum_{i=1}^{q_j} d_i [\hat{m}_i^2 + PEV(\hat{m}_i)].$$

Because $PEV(\hat{m}_i)$ can be further expressed as

$$PEV(\hat{m}_i) = 1 - r_{TI}^2 \frac{\sigma_u^2}{d_i},$$

where r_{TI}^2 is the reliability of MS, it can be shown that for sufficiently large data or a sufficient amount of resampled data sets,

$$\hat{\sigma}_{u_j}^2 = \frac{\sum_{i=1}^{q_j} d_i \hat{m}_i^2}{q_j} \frac{1}{r} \sum_{k=1}^r \frac{\sum_{i=1}^{q_j} d_i \tilde{m}_{ir}^2}{\sum_{i=1}^{q_j} d_i \hat{m}_{ir}^2}, \quad [4]$$

where \hat{m}_i is the MS estimate solved from the original data set of that replicate, \tilde{m}_{ir} is the simulated true MS of the resampled data set r , and \hat{m}_{ir} is its estimate solved from the resampled data set r (Lidauer et al., 2007). Comparisons of the methods were carried out only for setting M under scenario C (Figure 1) without accounting for inbreeding in the simulations; the effect of inbreeding was studied in a separate analysis presented in a later section. For PEV calculation in Equation 4, 5 resampled data sets were simulated in the same manner as explained in the section "Generating Phenotypes" for each of the 20 replicates, from which true breeding values and EBV were calculated to obtain the corresponding MS applied in Equation 4.

Sensitivity to Accuracy in Approximation of MS Reliabilities. The analyses carried out in this study and the reanalyses of the field data sets used for testing the first version of the MSVT (Fikse et al., 2003) indicated that the new method for genetic variance estimation was sensitive to the level of EBV reliability. Therefore, we conducted further analyses using different data edits under scenario C to study the sensitivity of within-year genetic variance estimation with respect to the level of approximated reliabilities (Figure 1). First, setting M including animals with EBV reliability values of 0.50 or higher (data edit 1) was compared with the results obtained using the complete data. This limit was chosen because although the overall average EBV reliability across the data for cows was 0.51, it was clearly below this average in the most challenging

birth year classes (1988 and 2009). When the test is applied in field conditions, however, data sets cannot be restricted by the level of EBV reliability without causing a possible selection bias. Therefore, instead of EBV reliability, we evaluated the reliability of MS as a criterion for excluding animals that would cause numerical instability in solving the nonlinear Equation 2. The MS reliability was derived as follows:

$$REL_{MS_i} = 1 - \frac{2PEV(\hat{m}_i)}{\sigma_u^2} = 1 - \frac{2C_i\sigma_e^2}{\sigma_u^2} = 1 - 2\alpha C_i,$$

where C is the quadratic form given in Equation 3. We tested different lower thresholds empirically for MS reliability and compared the results with those obtained on the complete data and those obtained with the FMS method on the complete data. Further, we studied the distribution of EBV reliabilities of the animals and their parents when the data were edited based on the animal's MS reliability value. These preliminary studies showed that an MS reliability of 0.1 was the lowest possible criterion for exclusion of animals causing numerical instability in the estimation. Hence, we selected it as the second data editing criterion (data edit 2).

The overall level of reliability for cows is well known to be lower than that for bulls. To determine the consequences of data edit 2 for cow data sets in the case of traits with low heritability (i.e., how many records a cow should have to obtain an MS reliability value of 0.1), we performed model calculations using GNU Octave (Eaton et al., 2016). A simple repeatability animal model, including a mean and random permanent environmental and animal effects, was fitted to the data. The data consisted of 1 sire having either 10 or 100 daughters, the dams of the daughters being unrelated and with only 1 offspring. The daughters and their dams had an equal number of records (from 1 to 10). In the case of only 1 record, the repeatability animal model was reduced to an animal model. We assumed a heritability of 0.1 and genetic variance of 1 for the trait under examination. A repeatability of 0.2 was assumed in the case of more than 1 observation. The MS reliability values for cows were calculated for each combination by using full inverse of C . The heritability of 0.1 was considered to describe an upper limit of the lowly heritable traits. The above model calculations were done to provide an indication if the validation test is applicable for cows in case of traits with a heritability of 0.1 or lower if data edit 2 is applied.

Effect of Inbreeding on Estimates of Genetic Variance. The effect of inbreeding was analyzed in setting M under scenario C (Figure 1). The inbreeding was taken into consideration in generating the phenotypes

for this part of the study. We examined the following alternatives: inbreeding was accounted for only in EBV prediction, inbreeding was accounted for only in estimation of within-year genetic variances, inbreeding was accounted for in both of these, or inbreeding was not modeled at all. For each alternative, 20 independent data replicates were simulated. The analyses of T2 and T10 and all analyses in setting S were carried out without modeling inbreeding, and the reliabilities of EBV were approximated in all cases without accounting for inbreeding.

Sensitivity of MSVT to Detect HV. For the MSVT, we selected bulls born in 1986 to 2006 and having daughters in at least 10 herds in both M and S settings. Cows born in 1988 to 2008 were included in the corresponding cow data sets. The validation tests were carried out for all 3 scenarios by applying data edit 2 (Figure 1). The analyses were performed with a program tailored specifically for this purpose (Tyrisevä et al., 2017). We excluded the last birth year (2009) in the complete data for cows following the recommendation by Tyrisevä et al. (2017) that each birth year class size should be at least half the average class size within the testing period.

RESULTS AND DISCUSSION

Accuracy of Estimates of Genetic Variance

The overall genetic variance for cows, averaged over the 20 replicates, was $245.4 \text{ kg}^2 \pm 0.10$ with the MSVT and $252.7 \text{ kg}^2 \pm 0.13$ with the FMS under scenario C (Figure 2). The corresponding genetic variance for bulls was $245.4 \text{ kg}^2 \pm 1.08$ with the MSVT and $253.5 \text{ kg}^2 \pm 1.20$ with the FMS (Figure 3). Thus, the FMS method gave a slightly higher overall level of genetic variance than the MSVT. The difference between average simulated genetic variance and the genetic variance (247 kg^2) applied to simulated breeding values was -0.7% with the MSVT and around 2.5% with the FMS. Hence, both estimation methods resulted in slightly biased estimates, but the bias was consistent across years and therefore the methods are deemed suitable for testing the existence of trends in variance.

Accuracy of Approximation of MS Reliabilities

With the MSVT method, the use of complete data and modeling of inbreeding resulted in notable peaks in genetic variance estimates for cows belonging to the first (1988) and last (2009) birth year classes, as reported by Tyrisevä et al. (2012) and illustrated in Figure 4 using results from 1 replicate. The same was not seen in estimates obtained with the FMS method (Tyrisevä et al.,

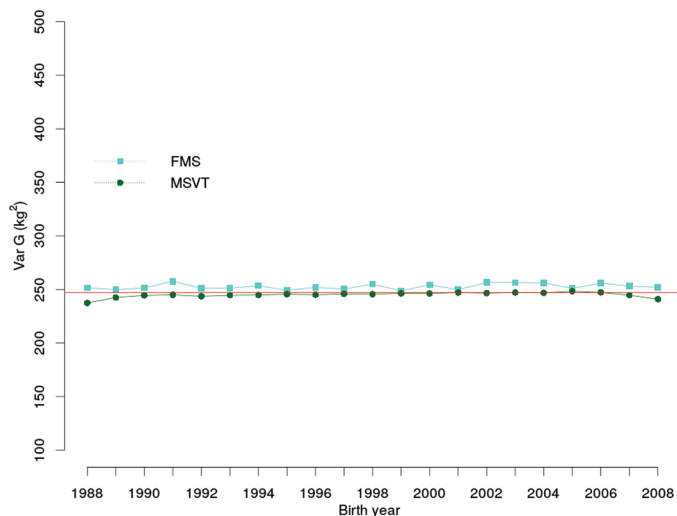


Figure 2. Within-year genetic variances (varG) for cows obtained with full model sampling (FMS) and Mendelian sampling variance test (MSVT). Results are averaged over the 20 replicates from scenario C (control). Only cows with Mendelian sampling reliability higher than 0.1 were included in the MSVT analyses. The FMS results were based on 5 FMS samples. The simulated value 247 kg² is marked with a straight (red) line. Color version available online.

2012). Birth year classes 1988 and 2009 with notable peaks differed in several ways from other birth year classes. The average number of observations for cows born in 1988 (2009) was zero (1) compared with 14 in the other classes. Further, the average EBV reliability for cows born in 1988 (2009) was 0.37 (0.42), whereas

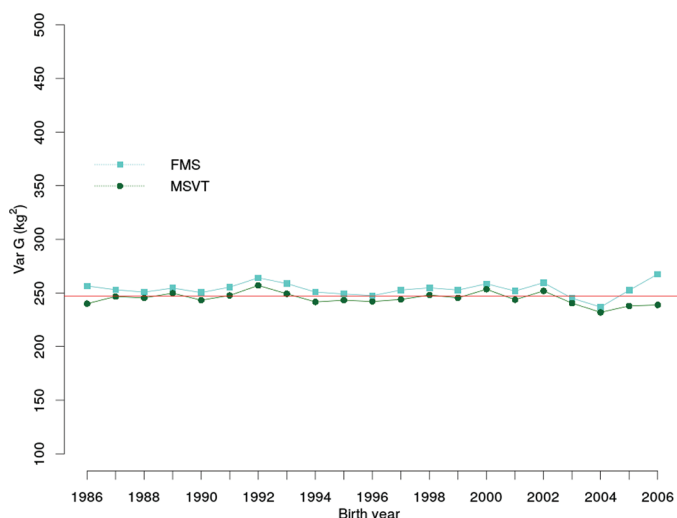


Figure 3. Within-year genetic variances (varG) for bulls obtained with full model sampling (FMS) and Mendelian sampling variance test (MSVT). Results are averaged over 20 replicates from scenario C (control). Only bulls with Mendelian sampling reliability higher than 0.1 were included in the MSVT analyses. The FMS results were based on 5 FMS samples. The simulated value 247 kg² is marked with a straight (red) line. Color version available online.

it was 0.51 in the other classes. Also, the third quartiles of the EBV reliabilities in the outermost classes were low: 0.46 (1988) and 0.45 (2009). As expected, the sires of the cows in these 2 birth year classes also received slightly lower EBV reliabilities than those in the other classes. Moreover, the dams of cows born in 1988 had the lowest EBV reliability values on average compared with the dams of cows in all other birth year classes. When all cows with EBV reliability lower than 0.50 were removed from the data, both peaks disappeared and the overall level of genetic variance approached that obtained with the FMS method (Figure 4). No such peaks in genetic variance estimates were detected for bull data, as also reported previously by Tyrisevä et al. (2012). The above results indicate that the FMS method is clearly more robust than the MSVT to data with low reliability values.

When EBV reliability is used as the editing criterion, animals with many lactations or many progeny are more likely to be included in the analysis (data edit 1). Thus, to avoid a possible selection bias, we decided to study the usability of MS reliability instead of EBV reliability as a criterion. Empirical testing suggested that an MS reliability of 0.1 was the lowest possible value with no effect on the estimates. After editing the data accordingly (data edit 2), the EBV reliabilities of the qualifying animals ranged from 0.25 to 0.82. Thus,

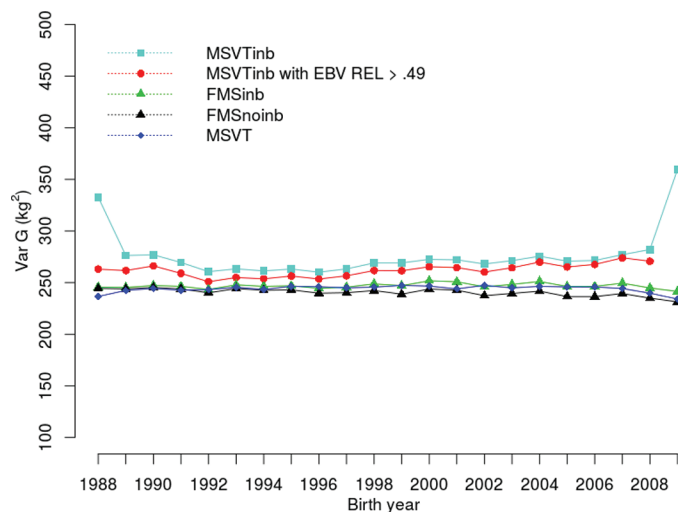


Figure 4. Effect of the data edits and modeling of inbreeding on estimates of within-year genetic variances (varG) for cows. Results are from 1 replicate under scenario C (control). Inb (noInb) refers to modeling (not modeling) of inbreeding in the prediction of breeding values and estimation of within-year genetic variances. EBV REL > 0.49 refers to data edit 1, in which only animals with EBV reliability equal to or greater than 0.50 were included. Mendelian sampling variance test (MSVT) alone refers to the approach proposed as the final validation test, in which inbreeding was not modeled and only animals with Mendelian sampling reliability higher than 0.1 were analyzed. FMS = full model sampling. Color version available online.

although low reliability values still existed, the peaks observed with unedited data were no longer detected. When animals were excluded from the analyses based on the criterion of 0.1 MS reliability and inbreeding was not modeled, we obtained estimates very close to those obtained with the FMS method (Figure 4). Fikse et al. (2003) further observed that the accuracy of genetic variance estimated with the MSVT was associated with the PEV level, and this, in turn, was affected by the level of the approximated EBV reliabilities. After excluding records with incomplete information and animals with MS reliability values of 0.1 or lower, the amount of bull data records declined by 0.3% and cow data records declined by 26%.

To illustrate the consequences of applying data edit 2 on cow data sets with lowly heritable traits, we performed model calculations of MS reliability values by varying the number of records for cows and for 2 alternative paternal half-sib group sizes, assuming a heritability of 0.1 (Table 1). Results showed that 3 observations were needed for a cow to attain an MS reliability value of 0.1. Thus, a large number of cows would not reach the reliability level needed for lowly heritable traits. We therefore suggest that only bulls be used for analysis of traits having low heritability.

Effect of Inbreeding

The inbreeding coefficient for both cows and bulls in setting M under scenario C was around 1% in the earlier years and increased to 4.5% in the later years. Figure 5 summarizes the effect of inbreeding under scenario C, averaged over the 20 replicates. When inbreeding was accounted for in both EBV prediction and genetic variance estimation, the overall estimated genetic variance was 247 kg², which is exactly the simulated variance, but it was lower (236.4 kg²) when inbreeding was

modeled only in EBV prediction and higher (257.1 kg²) when inbreeding was accounted for only in genetic variance estimation.

The trends in genetic variance were lowest (−0.31%) when inbreeding was accounted for only in EBV prediction and highest (0.24%) when inbreeding was accounted for only in genetic variance estimation. The trends were closest to zero when inbreeding was accounted for in both the prediction of EBV and the estimation of genetic variance (0.06%) and turned negative when inbreeding was not accounted for at all (−0.13%). The effect of leaving inbreeding unaccounted for was still tolerable. Thus, it appears that ignoring inbreeding in the estimation of genetic variance trends—as is the case in many national evaluation models—is not a crucial issue.

Sensitivity of the Validation Method to Detect HV

Moderate-Size Populations. For cows, the observed trend in genetic variance estimated in setting M under scenario C ranged from 0.02 to 0.13%, the average being 0.07% (Table 2). Although the trend was close to zero, it deviated statistically significantly from zero in 11 of the 20 replicates. We detected a statistically significant outlier in 8 out of the 20 replicates; there were 0.45 outliers on average per replicate. For bulls, the average trend in genetic variance was −0.13%

Table 1. Effect of parental data and the cow’s own data on the cow’s Mendelian sampling reliability value (MS REL) assuming 0.1 heritability

No. of records/cow	MS REL	
	10 Daughters/sire	100 Daughters/sire
1	0.049	0.051
2	0.083	0.087
3	0.108	0.114
4	0.127	0.134
5	0.143	0.151
6	0.156	0.165
7	0.167	0.176
8	0.176	0.186
9	0.183	0.194
10	0.190	0.201

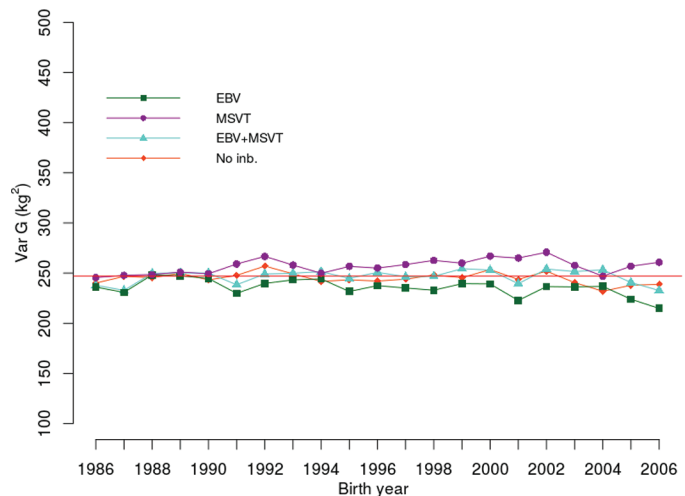


Figure 5. Effect of inbreeding on within-year genetic variances (varG) for bulls. Results are from 20 data replicates under scenario C (control). EBV refers to modeling of inbreeding only in breeding value prediction, Mendelian sampling variance test (MSVT) refers to modeling of inbreeding only in genetic variance estimation, EBV+MSVT refers to modeling of inbreeding in both breeding value prediction and genetic value estimation, and No inb indicates that no inbreeding was modeled. Color version available online.

Table 2. Results of the Mendelian sampling variance test applied to a moderate-size cow population under the control with simulated homogeneous variance (scenario C) and under scenarios with a simulated increase of 2% (T2) and 10% (T10) in phenotypic variance¹

Replicate	Scenario C			Scenario T2			Scenario T10		
	Trend (%)	H1	H2	Trend (%)	H1	H2	Trend (%)	H1	H2
1	0.05	NS	0	1.56	SS	1	4.46	SS	2
2	0.08	SS	0	1.60	SS	1	4.50	SS	1
3	0.08	SS	1	1.60	SS	1	4.49	SS	4
4	0.03	NS	0	1.55	SS	0	4.46	SS	2
5	0.06	NS	1	1.57	SS	1	4.48	SS	2
6	0.11	SS	1	1.63	SS	1	4.53	SS	3
7	0.07	SS	0	1.59	SS	1	4.49	SS	4
8	0.02	NS	0	1.54	SS	0	4.44	SS	1
9	0.09	SS	2	1.61	SS	2	4.52	SS	4
10	0.02	NS	0	1.53	SS	0	4.44	SS	2
11	0.11	SS	1	1.63	SS	1	4.53	SS	3
12	0.13	SS	1	1.64	SS	1	4.53	SS	3
13	0.07	SS	0	1.59	SS	0	4.50	SS	2
14	0.09	SS	1	1.61	SS	2	4.51	SS	4
15	0.04	NS	0	1.56	SS	1	4.46	SS	5
16	0.09	SS	1	1.61	SS	2	4.50	SS	3
17	0.07	SS	0	1.58	SS	1	4.48	SS	2
18	0.02	NS	0	1.54	SS	0	4.44	SS	3
19	0.04	NS	0	1.56	SS	0	4.47	SS	4
20	0.05	NS	0	1.57	SS	0	4.47	SS	6
Mean	0.07	11/20	0.45	1.58	20/20	0.80	4.49	20/20	3.0

¹H1 refers to the statistical test result for a trend (NS = nonsignificant; SS = significant), and H2 refers to the number of statistically significant outliers.

(Table 3), ranging from -0.71 to 0.43% among replicates. The slight overall negative trend was caused by inbreeding that was not accounted for. The trend for

bulls deviated statistically significantly from zero in 1 replicate only, whereas a statistical outlier was observed in 2 replicates.

Table 3. Results of the Mendelian sampling variance test applied to a moderate-size bull population under the control with simulated homogeneous variance (scenario C) and under scenarios with a simulated increase of 2% (T2) and 10% (T10) in phenotypic variance¹

Replicate	Scenario C			Scenario T2			Scenario T10		
	Trend (%)	H1	H2	Trend (%)	H1	H2	Trend (%)	H1	H2
1	-0.24	NS	0	1.17	SS	0	3.84	SS	0
2	-0.17	NS	0	1.28	SS	0	3.97	SS	0
3	-0.32	NS	0	1.10	SS	0	3.78	SS	0
4	-0.01	NS	0	1.38	SS	0	3.98	SS	0
5	-0.30	NS	1	1.10	SS	1	3.74	SS	1
6	0.14	NS	0	1.55	SS	0	4.15	SS	0
7	-0.14	NS	0	1.25	SS	0	3.83	SS	0
8	-0.17	NS	0	1.29	SS	0	4.01	SS	0
9	0.43	NS	0	1.88	SS	0	4.57	SS	0
10	0.37	NS	0	1.75	SS	0	4.30	SS	0
11	-0.34	NS	0	1.09	SS	0	3.75	SS	0
12	0.37	NS	0	1.77	SS	0	4.34	SS	0
13	-0.34	NS	0	1.09	SS	0	3.77	SS	0
14	-0.32	NS	0	1.12	SS	0	3.82	SS	0
15	0.21	NS	0	1.68	SS	0	4.42	SS	0
16	0.24	NS	0	1.70	SS	0	4.42	SS	0
17	-0.34	NS	0	1.11	SS	0	3.82	SS	0
18	-0.71	SS	1	0.73	SS	0	3.43	SS	0
19	-0.43	NS	0	0.98	SS	0	3.63	SS	0
20	-0.52	NS	0	0.91	SS	0	3.60	SS	0
Mean	-0.13	1/20	0.1	1.30	20/20	0.05	3.96	20/20	0.05

¹H1 refers to the statistical test result for a trend (NS = nonsignificant; SS = significant), and H2 refers to the number of statistically significant outliers.

For cows, a yearly increase of 2% in phenotypic variance (scenario T2) was observed as an average yearly increase of 1.58% in genetic variance (Table 2). The observed trends ranged from 1.53 to 1.64% and deviated statistically significantly from zero in all replicates. Statistically significant outliers were found in 13 replicates out of the 20; thus, their number increased under scenario T2 compared with the 8 outliers observed under scenario C. The corresponding trend in genetic variance for bulls under scenario T2 was 1.30% on average, ranging from 0.73 to 1.88% (Table 3). This trend also deviated statistically significantly from zero in all data replicates. A statistically significant outlier was found only in 1 bull data replicate.

For cows, generating a 10% yearly increase in phenotypic variance (scenario T10) resulted in an average yearly increase of 4.49% in genetic variance (Table 2). As in scenario T2, the trend estimates varied over a narrow range: from 4.44 to 4.53%. A statistically significant genetic variance trend was observed in all the cow data replicates. Further, there was a clear increase in the average number of outliers (3) under scenario T10 compared with scenario C. The overall genetic trend for bulls (3.96%) under scenario T10 was a bit lower than for cows (Table 3, T10). There was also more variation in the trend estimates for bulls (from 3.43 to 4.57%) than for cows. The trend in genetic variance among bulls was statistically significant in all data replicates, but in contrast to cows, only 1 statistical outlier was found in 1 data replicate.

Although number of significant outliers observed under scenario C was in a good agreement with the expected number of false positives, number of significant outliers clearly exceeded the expected false-positive rate in scenarios with the increasing trend in genetic variance. This is most likely caused by the way the genetic variance was simulated. Under the assumption that HV is attributed in proportion to all model effects, an increase in variance was introduced by scaling the variance of simulated observations. However, in the prediction of breeding values, HV was apparently not assigned in proportion to all effects, which caused an increase in between-years variation of genetic variance and, by this, an increase in number of outliers. Moreover, the model for prediction of breeding values assumed homogeneous (genetic) variance, causing the estimated breeding values to show less variance heterogeneity than the true breeding values, which explains why the obtained trend in genetic variance was on average smaller than would be expected.

Our results show that the MSVT is able to detect a trend in genetic variance rather accurately from a moderate-size population, provided that there is one in the data. However, the results also illustrate the impor-

ance of applying tolerance values for larger data sets, as is the policy for other Interbull validation tests (International Bull Evaluation Service, 2017). Field data sets can be huge, comprising hundreds of thousands of animals in a single year class. Hence, the statistical power of the method is high to detect very small deviations from a zero trend in genetic variance. Accordingly, a single estimate of within-year genetic variance that deviates from its expectation is often likely to be identified as a statistically significant outlier in large data sets. Therefore, tolerance values are necessary for both the trend and the outlier tests to enable detecting only those cases that have a real practical effect.

Based on our preliminary studies, a yearly increase of 2 to 4% in genetic variance may be a tolerable trend (Tyrisevä et al., 2012). For testing single outliers, an acceptance interval of $\bar{\sigma}_u^2 \pm 0.10\bar{\sigma}_u^2$ of the average estimated genetic variance was used. This was based on considering a 5% hypothetical standard error for the estimated variances and defining estimates that deviate more than 2 times the standard error (i.e., 10%) as failed outliers. The equality corresponds roughly to variances estimated from 800 observations; an approximated estimate of the standard error of $\bar{\sigma}_u^2$ is $\sqrt{2/(n-1)\bar{\sigma}_u^2}$.

Usability of the Validation Method in Small Populations. Under scenario C, the average trend in genetic variance for bulls in setting S was -0.27% (Table 4). The most extreme trend estimates among the 20 replicates were -1.78 and 1.95% . Although this range is clearly wider than for bulls in setting M (Table 3), none of the estimates differed statistically from zero (Table 4). One to 2 statistical outliers were found in 6 replicates out of the 20, and again, their proportion was higher than for bulls in setting M (2/20). Year-to-year variation in the estimates for bulls clearly increased in setting S compared with setting M. Nevertheless, the overall level of genetic variance estimated from replicates of the small bull population was still close to the parametric value ($249.4 \text{ kg}^2 \pm 3.89$ vs. 247 kg^2).

The average trend in estimated genetic variance under scenario T2 was 1.07% , ranging from -0.55 to 3.28% (Table 4). The trend deviated statistically significantly from zero in only 3 out of the 20 replicates. Under scenario T10, the average trend in genetic variance was 3.51% , ranging from 1.72 to 5.66% (Table 4). In total, a statistically significant trend in genetic variance was found in 19 of the 20 replicates. No increase was observed in the number of statistical outliers in either scenario T2 or scenario T10 compared with scenario C.

The above results in setting S show an increase in the sampling error and, hence, an increase in the probability of obtaining a false-positive result in small popula-

tions. However, the overall estimated genetic variance was still close to the simulated value. When no heterogeneity was simulated, no trend in genetic variance was detected. On the other hand, the generated trend had to be rather strong to differ statistically significantly from zero.

CONCLUSIONS

In this study we proposed a method for validating the homogeneity of MS variance in national genetic evaluation models. The method comprises the following steps: estimating within-year genetic variances, fitting a weighted linear regression model between the estimates and the years under study, identifying possible outliers, and defining a 95% empirical confidence interval for a possible trend. A simulation study was conducted to examine the specificity and sensitivity of the validation method. Results showed that the overall genetic variance estimate obtained by the proposed method was close to the value used in simulating the breeding values. The results were also in good accordance with the reference method. When no heterogeneity was introduced in the generated data representing a moderate-size population (setting M), the overall trend in genetic variance was practically zero. For bulls, we observed statistically significant deviation from a zero trend in only 5% of the replicates and a statistically

significant outlier in 10% of the replicates. However, for cows, a practical tolerance value proved necessary for both the trend and the outlier tests. Further, the MSVT method was able to detect a genetic variance trend in all cases in setting M regardless of the magnitude of the generated trend. In a small population (setting S), the MS validation test performed rather well, although only strong trends in genetic variance can be expected to reach statistical significance. Our results also indicate that the new method is sensitive to the quality of the approximated reliabilities of breeding values used to calculate the PEV. Therefore, we recommend that only animals with MS reliabilities higher than 0.1 be considered for the validation method and that low heritability traits be analyzed using bull data sets only.

ACKNOWLEDGMENTS

This study was conducted as a joint project by Natural Resources Institute Finland (Jokioinen, Finland), Interbull Centre (Uppsala, Sweden), Swedish University of Agricultural Sciences (Uppsala, Sweden), and NAV Nordic Cattle Genetic Evaluation (Aarhus, Denmark), all of which are gratefully acknowledged for their financial support. The project resulted in a tailored program for testing homogeneity of genetic variance (Tyrisevä et al., 2017), which can be acquired from the Interbull

Table 4. Results of the Mendelian sampling variance test applied to a small bull population under the control with simulated homogeneous variance (scenario C) and under scenarios with a simulated increase of 2% (T2) and 10% (T10) in phenotypic variance¹

Replicate	Scenario C			Scenario T2			Scenario T10		
	Trend (%)	H1	H2	Trend (%)	H1	H2	Trend (%)	H1	H2
1	-1.43	NS	1	0.07	NS	1	2.89	SS	1
2	-0.80	NS	1	0.46	NS	1	2.79	SS	1
3	0.39	NS	0	1.77	NS	0	4.28	SS	0
4	-0.88	NS	0	0.33	NS	0	2.55	SS	0
5	-0.40	NS	0	1.09	NS	0	3.85	SS	0
6	-0.86	NS	0	0.40	NS	0	2.73	SS	0
7	-1.04	NS	2	0.34	NS	2	2.87	SS	2
8	0.02	NS	1	1.41	NS	1	3.97	SS	1
9	-0.23	NS	0	1.08	NS	0	3.47	SS	0
10	1.09	NS	1	2.37	SS	1	4.70	SS	1
11	-1.08	NS	0	0.34	NS	0	3.04	SS	0
12	-1.78	NS	0	-0.55	NS	0	1.72	NS	0
13	1.95	NS	0	3.28	SS	0	5.66	SS	0
14	-1.13	NS	0	0.29	NS	0	2.92	SS	0
15	0.22	NS	0	1.62	NS	0	4.20	SS	0
16	-0.52	NS	2	0.69	NS	2	2.90	SS	2
17	0.80	NS	0	2.08	SS	0	4.40	SS	0
18	-0.81	NS	0	0.45	NS	0	2.74	SS	0
19	0.78	NS	0	2.14	NS	0	4.58	SS	0
20	0.40	NS	0	1.63	NS	0	3.88	SS	0
Mean	-0.27	0/20	0.4	1.07	3/20	0.4	3.51	19/20	0.4

¹H1 refers to the statistical test result for a trend (NS = nonsignificant; SS = significant), and H2 refers to the number of statistically significant outliers.

Centre. We also thank NAV for providing the Nordic Holstein test-day data for testing. Finally, we thank the anonymous reviewers for their excellent contribution in improving the manuscript.

REFERENCES

- Boichard, D., B. Bonaiti, A. Barbat, and S. Mattalia. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. *J. Dairy Sci.* 78:431–437.
- Ducrocq, V., I. Delaunay, D. Boichard, and S. Mattalia. 2003. A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations. *Interbull Bull.* 30:101–111.
- Eaton, J. W., D. Bateman, S. Hauberg, and R. Wehbring. 2016. GNU Octave. Version 4.2.0. Accessed Jan. 4, 2017. <http://www.gnu.org/software/octave/octave.pdf>.
- Fikse, W. F., L. Klei, Z. Liu, and P. G. Sullivan. 2003. Procedure for validation of trends in genetic variance. *Interbull Bull.* 31:30–36.
- Fikse, W. F., Z. Liu, and P. G. Sullivan. 2005. Tolerance values for validation of trends in genetic variances over time. *Interbull Bull.* 33:200–203.
- García-Cortés, L. A., C. Moreno, L. Varona, and J. Altarriba. 1995. Estimation of prediction-error variances by resampling. *J. Anim. Breed. Genet.* 112:176–182.
- Gengler, N., T. Dusseldorf, G. R. Wiggans, J. R. Wright, and T. Druet. 2000. Joint estimation of variances and effects in the US Jersey type evaluation system. *Interbull Bull.* 26:34–39.
- International Bull Evaluation Service. 2017. Code of practice. Appendix III. Accessed Jan. 4, 2017. <http://www.interbull.org/ib/codeofpractice>.
- Lidauer, M., K. Vuori, I. Strandén, and E. A. Mäntysaari. 2007. Experiences with Interbull test IV: Estimation of genetic variance. *Interbull Bull.* 37:69–72.
- Lidauer, M. H., R. Emmerling, and E. A. Mäntysaari. 2008. Multiplicative random regression model for heterogeneous variance adjustment in genetic evaluation for milk yield in Simmental. *J. Anim. Breed. Genet.* 125:147–159.
- Lidauer, M. H., J. Pösö, J. Pedersen, J. Lassen, P. Madsen, E. A. Mäntysaari, U. S. Nielsen, J.-Å. Eriksson, K. Johansson, T. Pitkänen, I. Strandén, and G. P. Aamand. 2015. Across-country test-day model evaluations for Holstein, Nordic Red Cattle and Jersey. *J. Dairy Sci.* 98:1296–1309.
- Matilainen, K., E. A. Mäntysaari, M. H. Lidauer, I. Strandén, and R. Thompson. 2012. Employing a Monte Carlo algorithm in expectation maximization restricted maximum likelihood estimation of the linear mixed model. *J. Anim. Breed. Genet.* 129:457–468.
- Meuwissen, T. H. E., G. De Jong, and B. Engel. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci.* 79:310–316.
- MiX99 Development Team. 2016. MiX99: A software package for solving large mixed model equations. Release XI/2016. Accessed Mar. 31, 2017. <http://www.luke.fi/mix99>.
- Miglior, F., P. G. Sullivan, and B. J. van Doormaal. 2002. Preliminary analysis of Mendelian sampling terms for genetic evaluation validation. *Interbull Bull.* 29:183–187.
- Miglior, F., K. A. Weigel, and G. Banos. 1998. Impact of heterogeneity of variance over time on international comparisons using a simulation approach. *Interbull Bull.* 17:40–45.
- Misztal, I., T. J. Lawlor, T. H. Short, and G. R. Wiggans. 1991. Continuous genetic evaluation of Holsteins for type. *J. Dairy Sci.* 74:2001–2009.
- Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71(Suppl. 2):27–32.
- Mrode, R. A. 1996. Genetic covariance between relatives. Pages 24–36 in *Linear Models for the Prediction of Animal Breeding Values*. CAB International, Wallingford, UK.
- NAV. 2017. NAV routine genetic evaluation of dairy cattle—Data and genetic models. 4th ed. Accessed Mar. 31, 2017. http://www.nordicebv.info/wp-content/uploads/2017/03/NAV-routine-genetic-evaluation-122016_FINAL.pdf.
- Robert-Granié, C., B. Bonaiti, D. Boichard, and A. Barbat. 1999. Accounting for variance heterogeneity in French dairy cattle genetic evaluation. *Livest. Prod. Sci.* 60:343–357.
- Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678.
- Sullivan, P. G. 1999. REML estimation of heterogeneous sire (co)variances for MACE. *Interbull Bull.* 22:146–148.
- Tyrisevä, A.-M., W. F. Fikse, and M. H. Lidauer. 2017. User Guide for Mendelian Sampling Variance Test. Version 2.6. Natural Resources Institute Finland (Luke), Jokioinen, Finland.
- Tyrisevä, A.-M., E. A. Mäntysaari, F. Fikse, and M. H. Lidauer. 2011. Simulation study on Mendelian sampling variance tests. *Interbull Bull.* 44:57–61.
- Tyrisevä, A.-M., E. A. Mäntysaari, J. Jakobsen, G. P. Aamand, J. Dürr, W. F. Fikse, and M. H. Lidauer. 2012. Validation of consistency of Mendelian sampling variance in national evaluation models. *Interbull Bull.* 46:97–102.
- Van Doormaal, B. J., G. J. Kistemaker, and P. G. Sullivan. 1999. Heterogeneous variances of Canadian bull EBVs over time. *Interbull Bull.* 22:141–148.
- Weigel, K. A., G. Banos, and A. Sigurdsson. 1996. International sire evaluations and conversions in upgrading populations with changing means and variances. *Interbull Bull.* 14:21–25.