Estimation Of The Proportion Of Variation Accounted For By DNA Tests. II. Phenotypic Variance

R.M. Thallman[*], K. J. Hanford[†], S. D. Kachman[†], L. A. Kuehn[*], R. L. Quaas[‡], R. J. Tempelman[§], R. L. Fernando[**], and E. J. Pollak[*].

Introduction

Predictions of merit from DNA tests have as much potential for application in marker-assisted management (**MAM)** as they have in marker-assisted selection (**MAS**), although at present, most tests are marketed for use in MAS. There are two fundamental differences between these applications. First, in MAS, the objective is to improve breeding value (the additive component of genetic merit), whereas in MAM, the objective is to predict the phenotype, which also includes non-additive components of the genotype. Second, in MAS, pedigree and breed composition are typically known. Consequently, it is often feasible to statistically partition the additive genetic from the residual components of phenotype. However, in MAM, pedigree and breed composition are often unknown. Fortunately, there is no need to partition the additive genetic component from the remainder of phenotype for MAM.

The proportion of variation accounted for by a DNA test is a useful metric with which to quantitatively evaluate the merit of commercial DNA tests (Van Eenennaam *et al*. (2009)). This paper evaluates two estimators of the proportion of phenotypic variation accounted for by a DNA test ($\boldsymbol{R_p^2}$).

Materials and Methods

We assume throughout that DNA test results will be presented in the form of molecular genetic values (**MGV**), which are continuous values computed from DNA test results that predict the total genetic merit of animals based only on the DNA test results.

Two estimators of R_p^2 were computed and evaluated. The first estimator is based on a multiple trait (**MT**) model in which the observed trait and MGV are each included as separate traits. It is the squared phenotypic correlation estimate divided by the broad sense heritability of the MGV ($\mathbf{h^2_{tm}}$). It is generally difficult to estimate h^2_{tm} directly, but the repeatability of the MGV from independently collected tissue samples may provide a useful estimate of it. Laboratory errors in determining genotypes, missing genotypes, and sample

* USDA-ARS U.S. Meat Animal Research Center, Clay Center, NE, USA
† University of Nebraska, Lincoln, NE, USA
‡ Cornell University, Ithaca, NY, USA
§ Michigan State University, East Lansing, MI, USA
** Iowa State University, Ames, IA, USA

identification errors should be the only causes of differences between h^2_{tm} and one. A practical solution could be to assume h^2_{tm} to be 0.99 or 1. The numerator of the second estimator is the reduction in phenotypic variance (**RV**) of a single trait model when MGV is added to the model as a covariate (full); the denominator of this estimator is the phenotypic variance of the reduced model that does not have MGV as a covariate. Assuming h^2_{tm} to be one, the approximate standard error of the MT estimator can be computed as twice the absolute value of the phenotypic correlation estimate times the standard error of the phenotypic correlation. The standard error of the RV estimator is not available because it is computed from two different analyses of the same data set. Detailed derivations of these estimators and standard errors were provided by Thallman *et al*. (2009).

Hanford *et al*. (2010) described the simulation of 500 replicates from a model that included both additive and non-additive genetic effects for the observed trait and an MGV. The pedigree consisted of 98-100 unrelated sires with an average of 10 offspring each from unrelated dams. Each of the progeny was randomly assigned to one of 20 contemporary groups. The twelve parameter combinations are presented in Table 1. The simulations assumed h^2_{tm} to be one and therefore the MT estimator in this paper is the squared phenotypic correlation. For the above replicates, the narrow sense heritability of MGV ($\mathbf{h^2_{gm}}$) was simulated as 0.8, with the remaining 20% of variation in MGV due to non-additive genetic effects. For each of the same twelve combinations of narrow sense heritability of the observed trait ($\mathbf{h^2_{gy}}$) and the proportion of additive genetic variance accounted for by the DNA test ($\boldsymbol{R_g}^2$), an additional 500 replicates with only additive genetic effects were

Table 1: Proportion of Phenotypic Variance Explained by MGV

Model: Estimator:			Two Trait Phenotypic Correlation Squared				Full & Reduced ST Variance Reduction	
Simulation Parameters			Mean Estimate ± Standard Error of Mean	Root Mean Sq Error	Mean Std Error of Est	Std Dev of Est	Mean Estimate ± Standard Error of Mean	Root Mean Sq Error
h^2_{gy}	R_g^2	R_p^2						
Data Simulated from Additive & Non-Additive Model								
0.1	0.1	0.013	0.0137±0.0003	0.008	0.007	0.008	0.0125±0.0003	0.007
0.1	0.2	0.025	0.0262±0.0004	0.010	0.010	0.010	0.0252±0.0005	0.010
0.1	0.4	0.050	0.0504±0.0006	0.014	0.013	0.014	0.0495±0.0006	0.014
0.1	0.6	0.075[a]	0.0768±0.0008	0.018	0.016	0.017	0.0758±0.0008	0.017
0.3	0.1	0.038	0.0385±0.0006	0.013	0.013	0.013	0.0375±0.0006	0.013
0.3	0.2	0.075	0.0746±0.0007	0.016	0.017	0.016	0.0734±0.0007	0.017
0.3	0.4	0.150	0.1485±0.0010	0.022	0.022	0.022	0.1479±0.0010	0.022
0.3	0.6	0.225	0.2252±0.0012	0.026	0.024	0.026	0.2250±0.0011	0.025
0.5	0.1	0.063	0.0645±0.0007	0.016	0.016	0.016	0.0634±0.0007	0.016
0.5	0.2	0.125	0.1248±0.0009	0.021	0.022	0.021	0.1238±0.0010	0.021
0.5	0.4	0.250	0.2534±0.0011	0.024	0.026	0.024	0.2529±0.0011	0.025
0.5	0.6	0.375	0.3741±0.0011	0.024	0.026	0.024	0.3734±0.0011	0.025

[a]One of the 500 replicates did not converge

generated (results not shown). For each replicate, the MT and RV estimators were computed. The single trait analyses were conducted using PROC Mixed of SAS[††] with an option that allows negative estimates of variances. Those estimates that fell within the parameter space were REML estimates and those that fell outside the parameter space were not REML. The two-trait analyses were conducted using ASReml[††]. The residual correlations were estimated. The parameters were within the parameter space with the exception of a few replicates that caused numerical problems when the additive variance was on the boundary at zero in a preliminary analysis.

Results and Discussion

Table 1 presents comparisons of the two alternative estimators of R_p^2 of the replicates simulated with non-additive, as well as additive, effects in the model. There was essentially no difference between the performance of the MT and RV estimators; both performed very well, with little bias and relatively small root mean squared errors (**RMSE**). The approximate standard errors of the MT estimator were very close to the standard deviations of the estimates. The statistics in Table 1 were also computed for the replicates simulated with only an additive genetic component (results not shown). The results were very similar to those in Table 1 and no difference in conclusions would be reached based on them. In general, RMSE were considerably smaller (proportionally) for the estimators of R_p^2 than for the analogous estimators of R_g^2, which were reported by Hanford *et al.* (2010) for the same replicates reported here. This was expected because the former estimators are not highly dependent on partitioning variation into additive genetic and residual components as are the latter. Furthermore, the denominator of each of the former estimators contains the phenotypic variance, and therefore, is essentially assured of being substantially positive, but this is not true of the latter estimators in which the denominator is the additive genetic variance.

For application to MAS, estimation of R_g^2 requires adequate pedigree structure to partition genetic from residual variance. Availability of such populations has been a limiting factor in the independent validation of DNA tests for MAS (Van Eenennaam *et al.* (2007)). However, an alternative approach would be to estimate R_p^2 from populations that have little pedigree structure and to divide by h^2_{gy} to obtain an estimate of R_g^2. This would obviously require an external estimate of h^2_{gy}, but for many traits, such estimates are readily available. This approach would utilize two populations to partition the problem of estimating R_g^2 into two separate problems. A pedigree structure and DNA tests would no longer need to be present in the same populations. However, there is risk that this approach will fail to identify populations in which the estimated h^2_{gy} was unusually low (had it been possible to partition it from the residual). Thus, populations that are not adequate to estimate the genetic covariance would not be identified. There may also be risk in using populations in which h^2_{gy} is

---

[††] Reference herein to any specific commercial products by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government, and shall not be used for advertising or product endorsement purposes.

unusually high. Thallman *et al.* (2009) provided extensive discussion of practical issues in estimating R_g^2 and R_p^2 and applying those estimates to the independent validation of DNA tests.

In order for an estimate of R_p^2 (R_g^2) to be useful, it is critical that the MGV (or molecular breeding value) not be derived from the same (or a very closely related population) as is used to estimate R_p^2 (R_g^2). Otherwise, the estimate of R_p^2 (R_g^2) will be biased up, perhaps very seriously. Finding appropriate populations can be a considerable challenge for traits for which phenotypes are quite limited (e.g., residual feed intake). A practical guideline is that the average relationship between the animals in the discovery data and those in the data from which R_p^2 (R_g^2) will be estimated should be approximately equal to the average relationship between the discovery population and the target population for application of the test.

For MAM, whether breed effects should be fit in the models for phenotypes and MGV depends on the application. One opportunity in MAM is to use the markers' ability to estimate breed composition to enhance the prediction of total genetic merit. Omitting breed effects seems appropriate in analyses for the validation of tests intended to be used for MAM in mixed breed populations in which breed composition is unknown (a common situation in feedlot applications). However, if the intended application of DNA tests is MAM within populations that are uniform in breed composition, it will be more useful to evaluate the tests with models that include breed effects. Evaluating tests with both models that include and do not include breed effects may be useful in determining the extent to which the tests' predictive ability is derived from the estimation of breed composition. However, simple correlations of unadjusted observations would generally not be appropriate for estimating R_p^2.

Conclusion

The proportion of variation accounted for by DNA tests should be a very useful tool for cattle producers to use in determining the value of DNA tests in their production systems and breeding programs. Fortunately, either of the two statistics considered estimate R_p^2 quite well. Therefore, DNA tests for MAM can be evaluated quite effectively in populations in which there is little pedigree structure, although the estimate could be biased up if pedigree structure exists, but is not accounted for in the model.

References

Hanford, K.J., Thallman, R.M., Quaas, R.L. *et al.* (2010). In *Proc. 9th WCGALP* (submitted)

Thallman, R.M., Hanford, K.J., Quaas, R.L. *et al.* (2009). In *Proc. BIF 41st Res. Symp & Annu. Meet.*, Sacramento, CA, pages 184-209.

Van Eenennaam, A.L., Li, J., Thallman, R.M., *et al*. (2007). *J. Anim. Sci.*, 85:891-900.

Van Eenennaam, A.L., Thallman, R.M., Quaas, R.L. *et al*. (2009). In *Proc. 18th Conf. AAABG*, Queensland, Australia, 18:129-132.