C4.5 Estimation of the Proportion of Genetic Variation Accounted for by DNA Tests.

R.M. Thallman[1], K. J. Hanford[2], R. L. Quaas*[3], S. D. Kachman[2], R. J. Tempelman[4], R. L. Fernando[5], L. A. Kuehn[1], and E. J. Pollak[3].

[1]USDA-ARS U.S. Meat Animal Research Center, Clay Center, NE, [2]University of Nebraska, Lincoln, NE, [3]Cornell University, Ithaca, NY, [4]Michigan State University, East Lansing, MI, and [5]Iowa State University, Ames, IA

An increasingly relevant question in evaluating commercial DNA tests is "What proportion of the additive genetic variation in the target trait is accounted for by the test?" Therefore, several estimators of this quantity were evaluated by simulation of a population of 1000 animals with 100 sires, each with 10 progeny. Three heritabilities (0.1, 0.3, and 0.5) of the target trait and four proportions of genetic variation (0.04, 0.16, 0.36, and 0.64) accounted for by the molecular breeding value (**MBV**) for the DNA test were simulated. In some replicates, non-additive genetic variation accounted for by the MBV was simulated using a heritability of the MBV of 80%. The first estimator evaluated is the reduction in estimated sire variance (**RSV**) when the MBV is added as a fixed covariate to a single-trait model for the target trait divided by the sire variance from the model without the MBV. The second estimator is based on the regression of phenotype on MBV (**RPM**) from a single trait model in which the MBV is a fixed covariate (this is the model that has been standard in independent validations since DNA tests began being reported as MBVs). This estimator is computed as $RPM = \frac{\left(\hat{b}_{P,\,MBV}\right)^2 \hat{\sigma}^2_{MBV}}{\hat{h}^2_{MBV}\ \hat{\sigma}^2_A}$ where $\hat{b}_{P,\,MBV}$ is the regression of the target phenotype on MBV, $\hat{\sigma}^2_{MBV}$ is the phenotypic variance of the MBV, $\hat{h}^2_{MBV}$ is the heritability of the MBV, and $\hat{\sigma}^2_A$ is the additive genetic variance of the target trait. The third estimator is the REML estimate of additive genetic correlation squared ($\mathbf{R_g^2}$) in a two-trait animal model for the target trait and the MBV (as the second trait). In this case, the only fixed effect in the model for MBVs is a mean. The standard error of R_g^2 was computed by multiplying the standard error of the genetic correlation by twice the genetic correlation. The mean estimates of R_g^2 tended to be closer to the simulated values than RSV and RPM, although all three estimators performed reasonably for most replicates. The standard deviations of estimates among replicates of R_g^2 were generally smaller than RSV and RPM and, in some cases, were much smaller. All three estimators can produce erratic results in replicates in which the estimate of the additive (or sire) variance approaches zero. Data sets in which this occurs should be considered inadequate for estimating the proportion of additive variation. The RSV estimator can produce negative estimates and the RSV and RPM estimators can produce estimates > 1 of the proportion of additive variance explained. The R_g^2 estimator has the advantage of producing estimates within the parameter space. The computed standard errors of R_g^2 were similar to the standard deviation of the estimates. This property is another advantage of R_g^2 over RSV and RPM, for which empirical methods for computing standard errors are not obvious. It is recommended that the R_g^2 estimator be used for estimating the proportion of additive variation explained by a DNA test.