TESTING AN APPROACH TO ACCOUNT FOR DAUGHTER MISIDENTIFICATION IN THE ESTIMATION OF BREEDING VALUES

N. Lopez-Villalobos[1], J.R. Bryant[2], B.L Harris[3] and D.J. Garrick[1,4]

[1] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand
[2] DairyNZ Limited, Hamilton, New Zealand
[3] Livestock Improvement Corporation, Hamilton, New Zealand
[4] Department of Animal Science, Iowa State University, Ames, IA, USA

SUMMARY

A stochastic simulation was developed to account for daughter misidentification (DaMi) in the estimation of breeding values (EBV) of bulls. Miss-assignment of parentage was simulated in 20% of daughters of 50% of bulls in 30 of 40 commercial herds. A bias of -40 kg protein was assigned to those cows that were misidentified as heifer calves. The model for genetic evaluation included fixed effects of contemporary group and random effects of additive animal, permanent environment and residual. The approach to account for DaMi was to include a regression coefficient for each sire that reflected DaMi. Compared to progeny test (PT) EBVs, parent average (PA) EBVs were overestimated by 2.3 and 3.4 kg protein for bulls with and without DaMi, respectively. Compared to PT EBVs, reproof (RP) EBVs were underestimated by 3.5 kg and overestimated by 5.1 kg protein for bulls with and without DaMi, respectively. The model that accounted for sires with DaMi removed biases from PA to PT for both groups of bulls, but compared to PT EBVs, RP EBVs were underestimated by 1.0 kg protein for both groups of bulls. This set of simulations indicates that paternity-verified status for a sire-daughter record can remove biases in genetic evaluation caused by DaMi.

INTRODUCTION

Studies in New Zealand dairy cattle (Johnson and Harris, 2010) have documented that estimated breeding values (EBV) of young bulls based on parent average (PA) are typically higher than subsequent EBV which include daughter information from the progeny test (PT) or from reproof (RP) in commercial herds. These biases from PA to PT and from PA to RP breeding values are not unique to New Zealand dairy bulls; there is evidence of similar biases in Denmark (Pedersen *et al*. 1995), the United States of America (Powell *et al*. 2004) and Germany (Rensing *et al*. 2009). Possible factors causing these biases include preferential treatment of bull dams (Pedersen *et al*. 1995) preferential treatment of daughters (Kuhn and Freeman, 1995), misidentification of sire-daughter pairs (Winkelman, 2013), heterogeneous variance across herds and years and effects of heterosis (van der Werf *et al*. 1994).

Using DNA parentage verification the magnitude of sire to daughter misidentification in commercial herds has been estimated at 23% (Bowley et al., 2012). However in PT herds, where sires are initially evaluated for widespread commercial use, the degree of sire misidentification was approximately 5% in Livestock Improvement Corporation PT herds for the 2005 to 2007 seasons (Winkelman, 2013). The impact of daughter misidentification on dairy sire breeding value estimation has been investigated in New Zealand (Winkelman, 2013). Estimates of progeny group yield means for milk, fat and protein of DNA-verified daughters were higher than those of daughters for which paternity had been assigned via mating records. Progeny of genetically superior sires can easily be assigned to inferior sires and their genetic evaluations are biased downward (Bowley et al., 2012). The underestimation of elite sires has been shown via simulation to have a negative impact on rates of genetic gain in the dairy cattle population (Bowley *et al.*, 2012).

In this study a stochastic simulation was developed to evaluate biases caused by daughter-sire misidentification and to test an approach to correct for this bias during successive predictions of breeding values of bulls from birth to 9 years of age representing PA, PT and RP.

MATERIALS AND METHODS

Simulation scheme. The cow population comprised 40 commercial herds and another 20 herds for the sire proving scheme (SPS). Average herd size was 518 lactating cows plus replacements. Protein yield (Py) for each lactation of a cow was calculated as:

$$Py = \mu + TBV + year + herd + p + e$$

where μ is the general mean of the population (μ = 160 kg protein); TBV is the true breeding value of an animal modelled as TBV = 0.5×(TBVsire + TBVdam) + (z x 0.7 x genetic SD) where z is a random number from a normal distribution with mean=0 and SD=1; year is a year effect simulated from a normal distribution with mean 0 and variance=14.0 kg^2; herd is a herd effect simulated from a normal distribution with mean 0 and variance=28.0 kg^2; p is a permanent environmental effect of the cow through her productive life and repeated for each lactation; e is a residual effect for each lactation. Estimates of genetic, permanent and residual variances were 82.1 kg^2, 85.5 kg^2 and 111.7 kg^2, respectively.

Genetic evaluation. Genetic evaluation for protein yield was performed each year using an animal model. The model equation for genetic evaluation was the following:

$$y_{ikm} = \mu + hyl_i + a_k + p_m + e_{ikm}$$

where y_{ikm} is the protein yield record for cow m in contemporary group i defined as cows of the same lactation number calving in the same herd and year (hyl), a_k is the random additive genetic effect of animal k, p_m is the random permanent environmental effect of cow m, and e_{ikm} is the random residual effect unique to y_{ikm}. Lactation yields of protein were mature equivalent.

The best 68 cows and best 6 progeny-tested bulls were used to produce 34 young bulls to be progeny tested in SPS herds, the best 20% of these were selected when the bulls were 5 year old based on progeny test results. Selected bulls were used in commercial herds for two years.

Simulation of daughter-sire misidentification and bias. A bias of -40 kg protein was added to the yields of 20% of cows in 30 of 40 commercial herds. The introduction of this bias was at the level of contemporary group (herd-year-lactation number) and the bias was repeated for the same cows across all repeated lactations. This introduction of bias was to represent miss assignment of parentage whereby genetically elite sires get assigned progeny of inferior sires. Those cows were progeny of 50% of the bulls selected at random. All progeny of the other 50% of the bulls were correctly assigned to their sire. The model equation for genetic evaluation to attempt to remove bias caused by daughter misidentification was the following:

$$y_{ijkm} = \mu + hyl_i + \beta_j + a_k + p_m + e_{ijkm}$$

where y_{ijkm} is the protein yield record for cow m in contemporary group i, daughter of sire j and β_j is the fixed regression coefficient for sire j that reflects misidentification (0 if rightly assigned daughter and 1 if wrongly-assigned daughter).

RESULTS AND DISCUSSION

Table 1 shows distributions of true and estimated breeding values for protein yield and their reliabilities for 238 bulls evaluated at different ages; 129 bulls had 20% of daughter misidentification in each of 30 commercial herds and 109 bulls were without daughter misidentification. Compared to PT EBVs, PA EBVs were overestimated by 2.3 kg protein for bulls with daughter misidentification and 3.4 kg protein for bulls without daughter misidentification. When selected bulls entered the bull team and a significant number of daughters in the commercial population contributed to the reproof of these bulls, a divergent bias occurred for the two groups of bulls. Compared to PT EBVs, RP EBVs of bulls with daughter misidentification were underestimated by 3.5 kg protein whereas RP EBVs of bulls without daughter misidentification were overestimated by 5.1 kg protein.

The biases in protein EBVs from PA to PT in this simulation agree with Johnson and Harris (2010) for New Zealand dairy cattle. They reported overestimation of 3 to 4 kg protein of PA EBV compared to PT EBV, and suggested that such bias could reflect drift in genetic evaluations (Powell *et al.* 2004) and may include parentage error associated with sires of sons as they accumulate daughter numbers.

Protein EBVs from progeny test results were similar for both bull groups, but RP EBVs of bulls without daughter misidentification increased whereas RP EBVs of bulls with daughter misidentification decreased. This trend was expected in the simulation because misidentification of sire-daughter was always penalised with a negative bias. This agrees with Winkelman (2013) who reported that EBVs for production traits were, on average, biased downwards when all progeny was not DNA-verified.

The model that accounted for sires with misidentified daughters, on average, removed biases from PA to PT for both groups of bulls, but compared to PT EBVs, RP EBVs were underestimated by 1.0 and 1.2 kg protein for bulls with and without daughter misidentification, respectively

Livestock Improvement Corporation has offered SNP-based DNA sire verification services to customers since mid-1990s. The test is based on approximately 100 SNPs, where recorded sire was deemed correct if concordance with his daughter was at least 99% (Winkleman 2013). Likewise, customers of CRV Ambreed can obtain DNA sire verification services via Genomz. Test results from either service provider are recorded in the national database and nine classes of verified paternity of cows can be derived, including DNA-verified and paternity assigned using mating and calving records. Further research is being undertaken to include parentage verification status in the genetic evaluation for each sire for all traits as an attempt to correct biases caused by daughter misidentification.

Table 1. Distributions of true (TBV) and estimated (EBV) breeding values for protein yield and their reliabilities for 129 bulls that had 20% of misidentified daughters and 109 bulls without misidentified daughters, evaluated in different years following selection on protein EBV obtained with *a model without adjustment* for daughter misidentification

		TBV		EBV		Reliability	
Age of bull	N	Mean	SD	Mean	SD	Mean	SD
		Bulls with 20% of misidentified daughters					
0	129	16.4	3.6	18.8	7.5	36	1.8
4	129	16.4	3.6	16.5	6.0	83	1.6
9	129	16.4	3.6	13.0	7.4	89	3.7
		Bulls without misidentified daughters					
0	109	16.7	3.9	20.0	7.0	36	1.6
4	109	16.7	3.9	16.6	6.4	83	1.7
9	109	16.7	3.9	21.7	12.0	89	4.0

Table 2. Distributions of true (TBV) and estimated (EBV) breeding values for protein yield and their reliabilities for 120 bulls that had 20% of misidentified daughters and 118 bulls without misidentified daughters, evaluated in different years following selection on protein EBV obtained with *a model that included adjustment* for daughter misidentification

		TBV		EBV		Reliability	
Age of bull	N	Mean	SD	Mean	SD	Mean	SD
		Bulls with 20% of misidentified daughters					
0	120	18.0	4.6	20.2	3.9	36	1.7
4	120	18.0	4.6	20.2	5.5	83	1.5
9	120	18.0	4.6	19.2	5.0	89	3.1
		Bulls without misidentified daughters					
0	118	18.1	4.8	20.6	3.7	36	1.9
4	118	18.1	4.8	20.6	5.5	83	1.5
9	118	18.1	4.8	19.4	5.3	90	4.0

CONCLUSION

This simulation study shows that misidentification of sire-daughter pairs can be a source of bias in the genetic evaluation of dairy sires, a model that includes parentage verification status in the genetic evaluation for each sire can potentially correct for this bias.

REFERENCES

Bowley F.E., Amer P.R. and Meier S. (2012) *Proc. 33rd Conf. Int. Soc. Anim. Genet*. P1004.

Johnson D.L. and Harris B.L. (2010) *Proc. 9th World Cong. Genet. App. Livest. Prod.* Leipzig, Germany. Communication No. 0227.

Kuhn M.T and Freeman A.E. (1995) *J. Dairy Sci.* **78**: 2067.

Pedersen G.A., Christensen L.G. and Petersen P.H. (1995) *Acta Agric. Scand. Sect. A, Anim. Sci.* **45**: 26.

Powell R.L., Sanders A.H. and Norman H.D. (2004) *J. Dairy Sci.* **87**: 2614.

Rensing S., Pasman E. and Reinhardt F. (2009) *Interbull Bulletin* No 40 pp. 123.

Van der Werf J.H.J., Meuwissen T.H.E. and de Jong G. (1994) *J. Dairy Sci.* **77**: 3174.

Winkelman A.M. (2013) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **20**: 25.