

## Relationship Between Individual Herd-Heritability Estimates and Sire Misidentification Rate

C. D. Dechow,\*<sup>1</sup> H. D. Norman,† N. R. Zwald,‡ C. M. Cowan,§ and O. M. Meland#

\*Department of Dairy and Animal Science, The Pennsylvania State University, University Park, PA 16802

†Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

‡Alta Genetics Inc., Watertown, WI 53094

§Genetic Visions Inc., Middleton, WI 53562

#Accelerated Genetics, Baraboo, WI 53913

### ABSTRACT

The objectives of this study were to estimate heritabilities within herds participating in Dairy Herd Improvement and determine the relationship of the individual herd heritability with sire misidentification rate. Individual herd heritabilities for milk, fat, and protein yield and somatic cell score (SCS) were calculated with daughter-dam regression and daughter-sire predicted transmitting ability (PTA) regression using 4,712,166 records from 16,336 herds available for August 2000 evaluations and 7,084,953 records from 20,920 herds available for August 2006 evaluations. Herd heritabilities were estimated using regression models that included fixed breed, age within parity, herd-year-season of calving, dam records nested within state, sire PTA within state, and an interaction between sire PTA and herd variance; random regression coefficients were dam records within herd and sire PTA within herd. Average daughter-dam herd heritability estimates ranged from 0.21 (SCS in 2000) to 0.73 (protein percentage in 2006), whereas daughter-sire herd heritability ranged from 0.10 (SCS in 2000) to 0.42 (protein percentage in 2006). Verification of sire identification with DNA marker analysis was provided by Accelerated Genetics and Alta Genetics Inc. Daughter-sire herd heritability was more strongly correlated with sire misidentification rate than daughter-dam herd heritability. The correlation between the first principal component for all measures of herd heritability and sire misidentification rate was  $-0.38$  (176 herds) and  $-0.50$  (230 herds) in 2000 and 2006, respectively. Herd heritability can be estimated with simple regression techniques for several thousand herds simultaneously. The herd heritability estimates were correlated negatively with sire misidentification

rates and could be used to identify herds that provide inaccurate data for progeny testing.

**Key words:** sire misidentification, daughter-dam regression, heritability

### INTRODUCTION

Heritability for yield varies among herds differing in characteristics such as mean milk production, phenotypic variance for yield, herd size, percentage of cows that are registered, average age at first calving, and region of the country (Van Vleck, 1970; Norman et al., 1972; Lofgren et al., 1985; Vinson, 1987; Dimov et al., 1995; Van Tassell et al., 1999; Zwald et al., 2003; Dechow and Norman, 2007). Not all variation in heritability among herds can be explained by general herd characteristics. Dechow and Norman (2007) demonstrated that individual herd heritabilities ( $h^2_H$ ) could be estimated with moderate accuracy using regression techniques that could potentially be applied to large data sets.

Sire misidentification reduces heritability estimates and contributes to variation in  $h^2_H$ . Van Vleck (1970) demonstrated that heritability estimates from paternal sibling correlations vary approximately by the squared percentage of cows with correctly identified sires. A reduction in heritability and associated decline in the accuracy of PTA arising from sire misidentification reduces genetic progress and variance among PTA (Geldermann et al., 1986; Banos et al., 2001). Estimates of sire misidentification rate were 5.2% in the Israeli Holstein population (Ron et al., 1996) and 10% in the United Kingdom (Visscher et al., 2002) based on tests of 173 and 837 cows, respectively. Geldermann et al. (1986) reported an overall sire misidentification rate of 13% for 1,221 daughters of 15 German-Friesian progeny test bulls. The misidentification rate for daughters of individual sires ranged from 4.7 to 24.1%.

The effect of sire misidentification on genetic evaluations is particularly problematic when it occurs in prog-

Received July 20, 2007.

Accepted December 19, 2007.

<sup>1</sup>Corresponding author: cdechow@psu.edu

eny test herds. Predicted transmitting ability based on first crop daughters will largely determine which bulls enter a proven sire lineup after progeny testing and which bulls become sires of sons. Change in PTA between first and second crop genetic evaluations of US sires is persistently higher than predicted based on reliability of first crop daughters (Powell et al., 2004). Inaccurate PTA can cause an inaccurate group of bulls to be marketed, and more importantly, an inaccurate group of bulls to be chosen as sires of sons. Other studies have stated that misidentification generally reduces PTA for high genetic merit bulls while inflating PTA for low genetic merit bulls (Geldermann et al., 1986; Banos et al., 2001). In general, high genetic merit bulls with many misidentified daughters will be disadvantaged relative to sires of equal genetic merit and fewer misidentified daughters because, on average, a misidentified daughter is only an average performer. However, there is an opportunity for misidentified daughters to have an effect on the sire's proof in either direction because not all misidentified daughters are average. Determining which sires were proven in low  $h_{FH}^2$  herds might help identify bulls whose genetic merit are most likely to be inaccurately estimated.

Sire identification can be verified using DNA markers, but costs involved with such testing generally limit the number of herds and cows tested. Even when sire identification is correct, electronic identification errors in the milking parlor and poor record keeping can reduce data accuracy. Thus, identification of herds with low  $h_{FH}^2$  and subsequent correction of the source of data error would improve the accuracy of progeny testing. Estimates of  $h_{FH}^2$  would be particularly valuable if  $h_{FH}^2$  was significantly correlated with sire misidentification. Dechow and Norman (2007) estimated  $h_{FH}^2$  for up to 499 herds at one time, but generating  $h_{FH}^2$  for all herds contributing to genetic evaluations simultaneously has not been attempted before.

The objectives of this study were to determine the feasibility of generating  $h_{FH}^2$  for all herds in a national data set, and determine the relationship between  $h_{FH}^2$  and herd misidentification rate.

## MATERIALS AND METHODS

Data for this study were mature-equivalent milk, fat, and protein yield, and SCS from the national dairy database for first through fifth parity cows. Fat and protein percentages were derived from mature-equivalent yields. Records were included only if cows had records available for all traits. Herd heritability was calculated twice: first for all herds in the data set as of August 2006 and secondly for all herds as of August

2000. Only herds with records in 2006 had  $h_{FH}^2$  included from 2000 to avoid computer memory limitations. Therefore, there were fewer total herds analyzed for 2000. Records were retained for a 2 to 5 yr window depending on herd size. If more than 1,600 lactation records were available in a 2, 3, or 4 yr span, the total years of records used were truncated at <5 yr. Heritability estimated with 1,600 daughter-dam pairs would have an approximate standard error of 0.05 (Falconer and Mackay, 1996). This allowed sufficient observations to estimate  $h_{FH}^2$  with reasonable accuracy, while retaining the most recent records in an attempt to reflect current  $h_{FH}^2$  as accurately as possible. In total, 98.0% of herds in 2000 and 97.4% of herds in 2006 included data from a 5-yr period. The average number of daughter-sire records per herd in 2006 was 339 with a maximum of 11,176, whereas the average and maximum number of daughter-dam observations was 201 and 5,951, respectively.

A lactation record was retained only if the cow's sire had reliability for PTA milk of 0.50 or greater. No restrictions were placed on the breed of the cow. Cows were required to calve between 18 and 120 mo of age. Herd-year-season of calving had 6 bimonthly calving seasons. For herd-year-seasons with <5 cows, seasons were expanded to 4-mo intervals. Herd-year was substituted for herd-year-season if <5 cows were in the herd-year-season after expanding season length; herd-years with <5 cows were excluded.

## Heritability Estimates

Herd heritability for milk, fat, and protein yields, SCS, and fat and protein percentages were estimated with daughter-dam and daughter-sire PTA regression using the methods of Dechow and Norman (2007). Dam records were adjusted for age within parity and herd phenotypic variance ( $\hat{\sigma}_H^2$ ) was estimated with the following model:

$$y_{ij} = b_i A_i + YS_j + e_{ij}, \quad [1]$$

where  $y_{ij}$  = mature-equivalent milk, mature-equivalent fat, mature-equivalent protein yield, SCS, fat percentage, or protein percentage for parity  $i$  of cow in year-season  $j$ ;  $b_i$  = coefficient for fixed regression on age nested within parity ( $A$ );  $YS_j$  = fixed effect of year-season  $j$ ; and  $e_{ij}$  = random residual. The analysis was applied to each herd individually with the MIXED procedure of SAS (SAS Inst. Inc., 2000).

To account for heterogeneity of variance among herds, residual variance from model 1 was used to generate  $\hat{\sigma}_H^2$  that was regressed toward the average residual

variance for a state. Herd residual variance was weighted by the residual degrees of freedom from model 1 and the average residual variance for all herds from a state was assigned a weight of 20, which is equivalent to the weight assigned to the average regional herd-year variance in heterogeneous variance adjustments for US genetic evaluations (Wiggans and VanRaden, 1991).

Herds with no residual degrees of freedom in model 1 were removed from the data set. There were 188 herds with residual degrees of freedom in 2000, but not in 2006; and these were not removed. The final data sets had 4,712,166 records from 16,336 herds for the 2000 sample, and 7,084,953 records from 20,920 herds from the 2006 sample.

The model for estimating  $h_{\text{H}}^2$  in ASReml (Gilmour et al., 2006) was

$$y_{ijklmno} = BD_k + b_l A_l + HYS_m + b_d D_n + b_s S_n \quad [2] \\ + b_{sd}(S \times SD_o) + b_{do} F_o + b_{so} G_o + e_{ijklmno},$$

where  $y_{ijklmno}$  = mature-equivalent milk yield, mature-equivalent fat yield, mature-equivalent protein yield, SCS, fat percentage or protein percentage for the  $i$ th record of cow  $j$  of breed  $k$ , parity  $l$ , calving in herd-year-season  $m$ , in state  $n$  and herd  $o$ ;  $BD_k$  = the fixed effect of breed  $k$ ;  $b_l$  = coefficient for fixed regression on age (A) nested within parity  $l$ ;  $HYS_m$  = fixed effect of herd-year-season  $m$ ;  $b_d$  = coefficient for fixed regression on dam record nested within state (D);  $b_s$  = coefficient for fixed regression on sire PTA (S) nested within state;  $b_{sd}$  = coefficient for fixed regression on the interaction between sire PTA and herd standard deviation (SD);  $b_{do}$  = coefficient for random regression on dam record nested within herd (F);  $b_{so}$  = coefficient for random regression on sire PTA nested within herd (G); and  $e_{ijklmno}$  = effect of random residual. Dechow and Norman (2007) used only records from Holsteins, and the effect of state was not included in their analysis. The purpose of including both fixed and random regression coefficients is to generate  $h_{\text{H}}^2$  that are regressed toward average heritability for all herds within a state. The dam records were the residual values ( $e_{ij}$ ) from model 1, which were averaged for dams with multiple records. Sire PTA were from May 2000 or May 2006 national genetic evaluations. Including sire PTA and dam records simultaneously adjusts for the merit of mates. The random regression coefficients ( $b_{do}$  and  $b_{so}$ ) were assumed to be correlated with the following variance structure:

$$\text{var} = \begin{bmatrix} b_{do} \\ b_{so} \end{bmatrix} = \begin{bmatrix} I_h \sigma_{b_{do}}^2 & \sigma_{b_{do}, b_{so}} \\ \sigma_{b_{do}, b_{so}} & I_h \sigma_{b_{so}}^2 \end{bmatrix},$$

where  $h$  = the number of herds.

Regression coefficients from model 2 were then used to generate  $h_{\text{H}}^2$ . The  $h_{\text{H}}^2$  from daughter-dam regression was  $2(b_d + b_{do})$ . The additive genetic standard deviation estimate from daughter-sire PTA regression for the herd was

$$\sigma_{ds} = [b_s + b_{sd}(SD_o) + b_{so}]SD_{US},$$

where  $SD_{US}$  = genetic standard deviation assumed for USDA-DHIA Holstein genetic evaluations (655 kg for mature-equivalent milk). Herd genetic variance ( $\sigma_{ds}^2$ ) was calculated as  $(\sigma_{ds})^2/R$ , where  $R$  = average sire PTA reliability for all cows in a herd. The  $\sigma_{ds}^2$  estimate for herds that use lower reliability young sires (progeny test herds) would be deflated more severely than other herds with no adjustment for reliability because young sires have less accurate PTA. Herd heritability from daughter-sire PTA regression can then be estimated as  $\sigma_{ds}^2/\hat{\sigma}_H^2$ . Multiple breed genetic evaluations in the United States were implemented in May 2007 (VanRaden et al., 2007). However, PTA originated from single breed evaluations in this study. Predicted transmitting abilities from other breeds were standardized to a Holstein base to accommodate the use of all breeds simultaneously.

### Principal Components

A total of 12  $h_{\text{H}}^2$  were estimated for each herd: daughter-dam  $h_{\text{H}}^2$  for mature-equivalent milk yield, mature-equivalent fat yield, mature-equivalent protein yield, SCS, fat percentage, and protein percentage and daughter-sire  $h_{\text{H}}^2$  for the same traits. Principal components were generated to provide measures of  $h_{\text{H}}^2$  that reflect the general tendency of a herd to have higher or lower  $h_{\text{H}}^2$  and provide a reduction in the number of  $h_{\text{H}}^2$  estimates per herd. Principal components were generated in the PRINCOMP procedure of SAS (SAS Inst. Inc., 2000). The first principal component for the 6 daughter-sire  $h_{\text{H}}^2$  (**PRIN<sub>DS</sub>**) was obtained. The first principal component for daughter-dam  $h_{\text{H}}^2$  (**PRIN<sub>DD</sub>**), and the first principal component for all 12  $h_{\text{H}}^2$  (**PRIN<sub>ALL</sub>**) was also derived.

### Herd Misidentification

Paternity verification results from DNA marker analysis were provided by Alta Genetics Inc. (Westby, WI) for herds that were candidates for the Alta Advantage progeny testing program and by Genetic Visions Inc. (Watertown, WI) for herds that were candidates for the



Accelerated Genetics PACE young sire program. The paternity verification results for Alta Genetics Inc. were from a 15-microsatellite marker analysis with an 11 microsatellite secondary panel in case of inconclusive results. The verification procedures varied slightly over the study period for Genetic Visions Inc., but included from 6 to 9 microsatellite markers and 32 SNP markers. The herds were larger than average and were using progeny test semen or interested in doing so. Not all of the herds were involved with DHI testing or remained on test in 2006, so not all herds with misidentification data had  $h_{\text{H}}^2$ . The DNA results from Alta Genetics Inc. were from a subset of herds sampled during 2 screening periods. An initial paternity verification screening of cows from 301 dairy herds was performed from 1999 to 2001. Follow-up screenings were performed for prospect herds with a low initial misidentification rate and that met other qualifications for the Alta Advantage progeny test program. Paternity verification results were provided for 95 Accelerated Genetics herds that were sampled between July 2001 and April 2007. Both studs tested a random sample ( $n = 3$  to 274) of cows. Herd misidentification rates were merged with herd heritability for 176 combined herds (2000) and 230 combined herds (2006).

Multiple-regression equations were used to generate three prediction formulas for herd misidentification rate based on 2006  $h_{\text{H}}^2$  estimates. Misidentification rates from Alta Genetics Inc. were used to derive a prediction equation for misidentification. The resulting formula was then applied to  $h_{\text{H}}^2$  estimates for herds with misidentification observations from Accelerated Genetics and the correlation between predicted and observed misidentification rates was estimated. Likewise, misidentification rate observations from Accelerated Genetics were used to predict misidentification for the Alta Genetics Inc. sample. A pooled sample from both studs was also used to predict misidentification rate for all herds with 2006  $h_{\text{H}}^2$ . All 12  $h_{\text{H}}^2$  estimates and the square of all  $h_{\text{H}}^2$  estimates were eligible for inclusion in the prediction equation. The stepwise regression option of the REG procedure of SAS (SAS Inst. Inc., 2000) was used to identify significant variables. In two instances (daughter-dam heritability for milk and SCS) the squared term was selected by the REG procedure, but was replaced for the final model with the linear term to create a hierarchical model with minimal loss in the model effectiveness. To prevent over-extrapolation,  $h_{\text{H}}^2$  for all traits were constrained to the minimum and maximum observed values for the 230 herds that were used to generate the prediction equation.

**Table 1.** Mean and standard deviation of daughter-dam and daughter-sire estimates of herd heritability for milk, fat, and protein yield, SCS, fat percentage, and protein percentage in 2000 and 2006 and correlations ( $r$ ) between herd heritabilities estimated in 2000 and 2006<sup>1</sup>

Item	2000		2006		$r$
	Mean	SD	Mean	SD	
	Daughter-dam				
Milk	0.34	0.05	0.36	0.06	0.25
Fat	0.39	0.07	0.41	0.07	0.30
Protein	0.30	0.06	0.32	0.06	0.25
SCS	0.21	0.06	0.26	0.06	0.16
Fat percentage	0.68	0.10	0.72	0.12	0.42
Protein percentage	0.67	0.12	0.73	0.12	0.41
	Daughter-sire				
Milk	0.21	0.10	0.18	0.08	0.39
Fat	0.22	0.09	0.18	0.08	0.34
Protein	0.22	0.12	0.17	0.09	0.42
SCS	0.10	0.04	0.11	0.04	0.23
Fat percentage	0.29	0.10	0.39	0.13	0.34
Protein percentage	0.29	0.10	0.42	0.13	0.37

<sup>1</sup>Heritability estimates are averages of 16,336 herds in 2000 and 20,920 herds in 2006. Correlations were for 16,148 herds.

## RESULTS AND DISCUSSION

The mean and SD of  $h_{\text{H}}^2$  for 16,336 herds in 2000 and 20,920 herds in 2006 are reported in Table 1. In 2000, mean daughter-dam  $h_{\text{H}}^2$  ranged from 0.21 for SCS to 0.68 for fat percentage. Daughter-dam  $h_{\text{H}}^2$  was highest for protein percentage (0.73) and lowest for SCS (0.26) in 2006. Herd heritabilities were lower when estimated with daughter-sire PTA regression, ranging from 0.10 to 0.29 in 2000, and from 0.11 to 0.42 in 2006. Herd heritability estimated with daughter-sire PTA regression was lowest for SCS and highest for fat or protein percentage in both 2000 and 2006.

Daughter-dam  $h_{\text{H}}^2$  were reported to be higher than daughter-sire  $h_{\text{H}}^2$  by Dechow and Norman (2007). Daughter-dam heritability estimates for milk yield are also higher than heritability estimates from paternal half-sibling correlations (Van Vleck and Bradford, 1965, 1966). Maternal and common environmental effects among daughters and dams likely inflate daughter-dam heritability (Van Vleck and Bradford, 1965). Lower heritabilities estimated with techniques that rely on sire-daughter relationships may also indicate that sire misidentification is more common than dam misidentification. Whereas mean daughter-sire  $h_{\text{H}}^2$  was lower than daughter-dam  $h_{\text{H}}^2$ , the standard deviation for daughter sire  $h_{\text{H}}^2$  (range = 0.04 to 0.13) was often higher than daughter-dam  $h_{\text{H}}^2$  (range = 0.05 to 0.12).

Daughter-dam and daughter-sire  $h_{\text{H}}^2$  were positively correlated. Model 2 assumed correlated daughter-dam and daughter-sire PTA regression coefficients within

herd ( $b_{do}$  and  $b_{so}$ , respectively). Correlations between  $b_{do}$  and  $b_{so}$  (not shown) were generally lower for yield traits (range = 0.12 to 0.19) than for SCS, fat percentage, and protein percentage (range = 0.21 to 0.42). Correlations between  $h_{H}^2$  in 2000 and 2006 are reported in Table 1 for 16,148 herds with  $h_{H}^2$  in both years. The correlations ranged from 0.16 (daughter-dam  $h_{H}^2$  for SCS) to 0.42 (daughter-dam  $h_{H}^2$  for fat percentage and daughter-sire  $h_{H}^2$  for protein yield). Because a maximum of 5 yr of data was used to estimate  $h_{H}^2$  for any herd, there was no overlap of records in 2000 and 2006. The correlations between 2000 and 2006 estimates of  $PRIN_{DD}$ ,  $PRIN_{DS}$ , and  $PRIN_{ALL}$  were 0.38, 0.47, and 0.48 (not shown), respectively. The moderate correlations between years indicate that  $h_{H}^2$  are not constant over time. Moreover, heritability estimates are relatively imprecise, and only  $h_{H}^2$  estimated with more than 1,600 daughter-parent combinations would have a standard error of 0.05 or less (Falconer and Mackay, 1996). Correlations between 2000 and 2006 herd heritability estimates were impacted by herd size. For herds with 100 or more cows in 2006, the correlation between  $PRIN_{DS}$  in 2000 and  $PRIN_{DS}$  in 2006 was 0.61, whereas the correlation of  $PRIN_{ALL}$  in 2000 and 2006 was 0.58. The corresponding correlations for herds with fewer than 100 cows were 0.45 for  $PRIN_{DS}$  and 0.44 for  $PRIN_{ALL}$ . The correlation between  $PRIN_{DD}$  in 2000 and 2006 was low regardless of herd size; the correlation was 0.36 for herds with fewer than 100 cows vs. 0.38 for herds with 100 or more cows.

Average daughter-dam and daughter-sire mature-equivalent milk  $h_{H}^2$  estimates for various breeds are displayed in Table 2. A herd was included in the breed average if at least 75% of cows in that herd represented the respective breed. A herd with less than 75% of all breeds or with more than 75% crossbreds (7 herds) was pooled into a mixed breed category. Average daughter-dam  $h_{H}^2$  ranged only from 0.35 to 0.37, whereas daughter-sire  $h_{H}^2$  had a larger range from 0.17 to 0.25. Among breeds with more than 100 herds represented (all except Milking Shorthorn and Red and White) the range of daughter-sire  $h_{H}^2$  was 0.17 (mixed breed) to 0.21. The Brown Swiss and Jersey breeds have higher assumed heritabilities for milk yield (0.35) in US genetic evaluations (AIPL, 2007) than Holstein (0.30). Herd heritability estimates for nonHolstein herds may have been underestimated marginally by including all breeds in the analysis after standardizing PTA to the Holstein scale. However, the ability to analyze all herds and all cows within a multiple breed herd simultaneously compensates for minimal bias across breeds. Herd heritability was estimated prior to implementation of mixed breed

analysis, which should reduce potential bias further in the future.

### Misidentification

The average initial herd misidentification rate for the 396 herds screened by Alta Genetics Inc. and Accelerated Genetics was 26%. The average misidentification rate for 230 herds that also had  $h_{H}^2$  available was 18.2%. The misidentification rate for individual herds ranged from 0 to 93.7% and among herds with 50 or more cows tested, the range was 1.5 to 50%. Not all herds from the initial screening were, or are currently, part of the Alta Advantage progeny test program or the Accelerated Genetics PACE program because the accuracy of identification did not meet the standards of the respective young sire programs. Therefore, the misidentification rates presented are substantially higher than the misidentification rate of Alta Advantage or PACE progeny test herds.

It is clear that misidentification rates are high in many large dairy herds, which severely compromises the accuracy of genetic evaluations when those herds are involved in progeny test programs. However, there are also many large herds with low misidentification rates and that can be valuable contributors to progeny test programs. Aggressive DNA sampling of progeny test daughters can also allow misidentified daughters to have their parentage corrected, as is currently the practice in Alta Advantage herds.

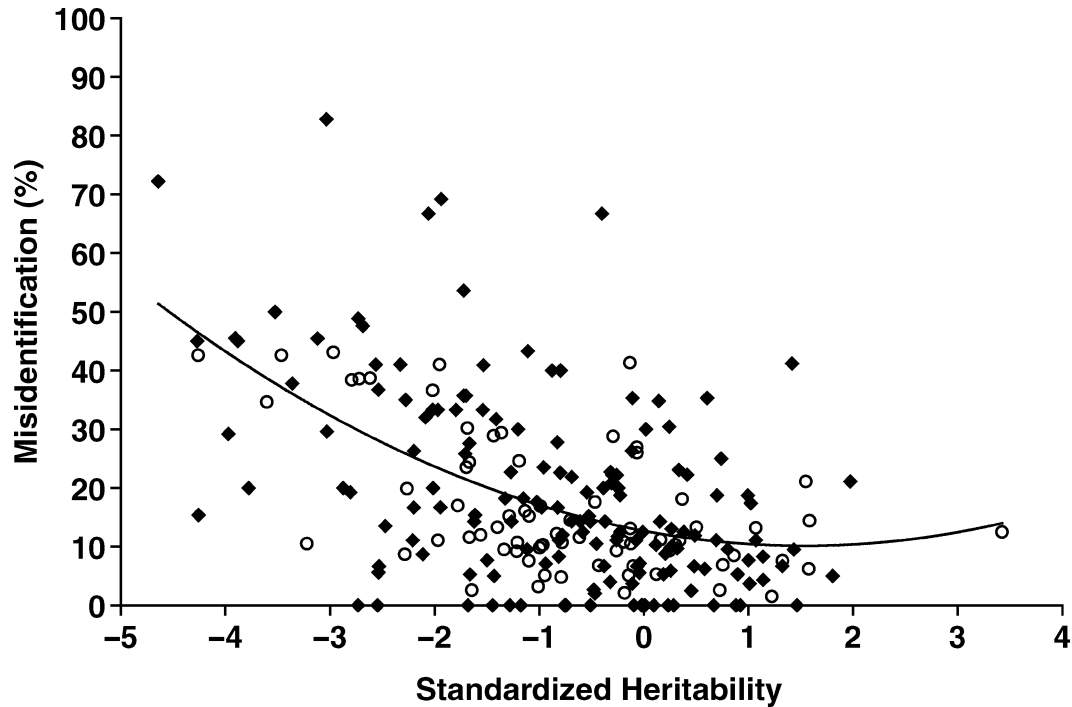
Correlations between  $h_{H}^2$  and herd misidentification rate are reported in Table 3. Higher  $h_{H}^2$  were correlated with lower sire misidentification rates. Correlation between  $h_{H}^2$  in 2000 and misidentification rate are based on 176 herds and range from -0.14 (daughter-dam  $h_{H}^2$  for SCS) to -0.29 (daughter-sire  $h_{H}^2$  for fat and fat percentage). Correlations of misidentification rate with  $h_{H}^2$  in 2006 (230 herds) were stronger (range -0.28 to -0.43) than those for 2000. The majority of DNA sam-

**Table 2.** Number of herds (n) and average daughter-dam and daughter-sire herd heritability for milk yield by breed<sup>1</sup>

Breed	n	Daughter-dam	Daughter-sire
Ayrshire	110	0.37	0.21
Brown Swiss	265	0.36	0.21
Guernsey	174	0.36	0.21
Holstein	18,437	0.36	0.20
Jersey	1,106	0.37	0.18
Milking Shorthorn	34	0.37	0.25
Red and White	20	0.37	0.23
Mixed breed <sup>2</sup>	774	0.35	0.17

<sup>1</sup>Herd must be composed of at least 75% of the respective breed.

<sup>2</sup>All herds with less than 75% of cows representing a single breed.



**Figure 1.** Relationship between herd misidentification rate and a standardized principal component for all herd heritability measures.

ples were taken after 2000, which is likely the reason for stronger correlations of misidentification rate with 2006  $h^2_{\text{H}}$ . In general, daughter-sire  $h^2_{\text{H}}$  was more strongly correlated with sire misidentification rate than daughter-dam  $h^2_{\text{H}}$ . Daughter-dam  $h^2_{\text{H}}$  might be a stronger indicator of dam misidentification rate than daughter-sire  $h^2_{\text{H}}$ , but dam misidentification rates were not available.

Correlations between herd misidentification rate and  $\text{PRIN}_{\text{DD}}$ ,  $\text{PRIN}_{\text{DS}}$ , and  $\text{PRIN}_{\text{ALL}}$  are reported in Table 4. Correlations of misidentification rate with  $\text{PRIN}_{\text{DD}}$ ,  $\text{PRIN}_{\text{DS}}$ , and  $\text{PRIN}_{\text{ALL}}$  generated from 2000  $h^2_{\text{H}}$  were  $-0.29$ ,  $-0.37$ , and  $-0.38$ , respectively. Correlations were

stronger between misidentification rate and  $\text{PRIN}_{\text{DS}}$  ( $-0.48$ ) and  $\text{PRIN}_{\text{ALL}}$  ( $-0.50$ ) than with  $\text{PRIN}_{\text{DD}}$  ( $-0.40$ ) generated from 2006  $h^2_{\text{H}}$ . A plot of misidentification rate vs.  $\text{PRIN}_{\text{ALL}}$  for 2006 is shown in Figure 1. There was significant evidence of a nonlinear relationship (quadratic effect significant at  $P < 0.01$ ) between misidentification and  $h^2_{\text{H}}$ . Misidentification rate was relatively constant for herds with average or higher  $h^2_{\text{H}}$ , but increased significantly as  $h^2_{\text{H}}$  became lower. Misidentification has been shown to reduce heritability estimates previously (Van Vleck, 1970; Geldermann et al., 1986).

Whereas there is a clear relationship between misidentification rate and measures of  $h^2_{\text{H}}$ , there are notable exceptions. There were 14 herds with 0% misidentification rates, but that had  $\text{PRIN}_{\text{ALL}}$  lower than average. The lack of misidentification in those herds may have

**Table 3.** Correlation of herd misidentification rate in 2000 and 2006 with daughter-dam and daughter-sire herd heritability for milk, fat, and protein yield, SCS, and fat and protein percentages<sup>1</sup>

Item	2000		2006	
	Daughter-dam	Daughter-sire	Daughter-dam	Daughter-sire
Milk	-0.20*	-0.27***	-0.32***	-0.43***
Fat	-0.24**	-0.29***	-0.33***	-0.33***
Protein	-0.23**	-0.27***	-0.29***	-0.40***
SCS	-0.14	-0.20**	-0.29***	-0.28***
Fat %	-0.24**	-0.29**	-0.32***	-0.38***
Protein %	-0.21**	-0.23**	-0.29***	-0.28***

<sup>1</sup>Correlations based on 176 herds in 2000 and 230 herds in 2006. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

**Table 4.** Correlations of herd misidentification rate with first principal component (PC) for daughter-dam herd heritability estimates, daughter-sire herd heritability estimates, and all heritability estimates for 176 herds in 2000 and 230 herds in 2006

PC based on:	2000	2006
Daughter-dam	-0.29***	-0.40***
Daughter-sire	-0.37***	-0.48***
All estimates	-0.38***	-0.50***

\*\*\* $P < 0.001$ .

**Table 5.** Measures of daughter-dam and daughter-sire herd heritability and the corresponding regression coefficients selected to predict misidentification rates when based on Alta Advantage, PACE, or pooled misidentification data sets

Trait	Alta Advantage	PACE	Pooled
R <sup>2</sup>	0.36	0.32	0.35
Intercept	63.7	66.2	65.3
	Daughter-sire		
Milk	-165.9		-127.3
Milk <sup>2</sup>	257.9		199.1
Fat %	-74.7	-137.5	-87.7
Fat % <sup>2</sup>	76.6	130.9	86.5
Protein		-34.2	
SCS		-65.5	
	Daughter-dam		
Milk	-27.4		-26.6
SCS	-18.1		-23.5

reflected limited sampling. The average number of cows tested in those herds was 14.5, with a range of 3 to 37. It is also likely that factors other than misidentification rate influenced the heritability estimates. Zwald et al. (2003) reported that several factors other than misidentification rate were associated with varying heritability levels, including yearly rainfall, maximum monthly temperature, percentage of North American Holstein, and herd size.

The measures of heritability and corresponding regression coefficients used to predict herd misidentification rate are reported in Table 5. The bull stud specific equations predicted misidentification rates for samples originating from the opposite bull stud with moderate accuracy. The correlation of predicted and observed misidentification rate for Accelerated Genetics was 0.54 and the mean absolute difference between observed and expected misidentification was 10.51, whereas the correlation and mean absolute differences for Alta Genetics Inc. were 0.42 and 11.18, respectively. This indicates that  $h_{\text{H}}^2$  could be a useful tool to help identify herds with poor sire misidentification, but that the predicted misidentification rate for an individual herd is not precise.

In general, the prediction equations were not more strongly associated with misidentification than PRI-

$N_{\text{ALL}}$ . However, the prediction equations were valuable for the purpose of estimating industry misidentification rates. The average predicted misidentification rate for all 20,920 herds with 2006  $h_{\text{H}}^2$  was 15.4%, and ranged from 0 to 56.5%. Average, minimum and maximum predicted misidentification rates stratified by the average number of first parity cows on DHI test annually are reported in Table 6. Average predicted misidentification rates ranged from 14.5 to 20.9 and increased with herd size for those herds milking more than 50 first lactation daughters. The standard deviation of predicted misidentification also increased as herd size increased because the methods of generating  $h_{\text{H}}^2$  regress estimates to the mean more severely for small herds than for larger herds (Dechow and Norman, 2007), but in all herd size groups the maximum was 47% or more.

## CONCLUSIONS

Current levels of sire misidentification in the US dairy industry could have a severe impact on the accuracy of genetic evaluations. Heritability estimates for 20,920 individual herds were successfully estimated using daughter-dam and daughter-sire PTA regression and higher  $h_{\text{H}}^2$  was correlated with lower herd misidentification rates. Progeny test programs should avoid herds with lower than average  $h_{\text{H}}^2$  for multiple heritability estimates, or verify parent identification in those herds with DNA marker analysis. Small herds had lower average predicted misidentification rates, but larger herds had  $h_{\text{H}}^2$  that were more stable over time and herds of all sizes were predicted to have misidentification rates as high as 47%. Selection of herds with low sire misidentification rates and accurate herd records should be a higher priority than herd size in progeny test programs.

## ACKNOWLEDGMENTS

The authors appreciate funding from the Agricultural Research Service, USDA. The DNA parent verification data provided by Alta Genetics Inc. and Genetic Visions Inc. is greatly appreciated. The cooperation of AgriTech

**Table 6.** Number of herds (n) and mean, standard deviation, and minimum and maximum predicted misidentification rate among herds stratified by the average annual number of first-parity cows

Cows	n	Mean	SD	Minimum	Maximum
≤25	15,039	15.3	6.0	0.0	49.7
26 to 50	3,401	14.5	7.3	0.0	47.0
51 to 100	1,431	15.6	8.1	0.0	56.3
101 to 200	603	17.4	9.0	0.4	52.0
201 to 300	194	18.7	10.2	2.6	50.4
≥301	252	20.9	9.8	4.4	56.5



Analytics (Visalia, CA), AgSource Cooperative Services (Verona, WI), Dairy Records Management Systems (Raleigh, NC), DHI Computing Services (Provo, UT), and Texas DHIA (College Station, TX) in supplying yield data through the National Genetic Improvement Program was invaluable.

## REFERENCES

- AIPL (Animal Improvement Programs Laboratory). 2007. Description of national genetic evaluation systems. <http://aipl.arsusda.gov/reference/xm-yield.htm> Accessed Dec. 19, 2007.
- Banos, G., G. R. Wiggans, and R. L. Powell. 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *J. Dairy Sci.* 84:2523–2529.
- Dechow, C. D., and H. D. Norman. 2007. Within-herd heritability estimated with daughter-parent regression for yield and somatic cell score. *J. Dairy Sci.* 90:482–492.
- Dimov, G., L. G. Albuquerque, J. F. Keown, L. D. Van Vleck, and H. D. Norman. 1995. Variance of interaction effects of sire and herd for yield traits of Holsteins in California, New York, and Pennsylvania with an animal model. *J. Dairy Sci.* 78:939–946.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman Group, Essex, UK.
- Geldermann, H., U. Pieper, and W. E. Weber. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. *J. Anim. Sci.* 63:1759–1768.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. *ASReml User Guide: Release 2.0*. VSN International Ltd., Hemel Hempstead, UK.
- Lofgren, D. L., W. E. Vinson, and R. E. Pearson. 1985. Heritability of milk yield at different herd means and variances for production. *J. Dairy Sci.* 68:2737–2739.
- Norman, H. D., B. T. McDaniel, and F. N. Dickinson. 1972. Conflicts between heritability estimates of mature equivalent and herd-mate-deviation milk and fat. *J. Dairy Sci.* 55:507–517.
- Powell, R. L., A. H. Sanders, and H. D. Norman. 2004. Stability of genetic evaluations for active artificial insemination bulls. *J. Dairy Sci.* 87:2614–2620.
- Ron, M., Y. Blanc, M. Band, E. Ezra, and J. I. Weller. 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *J. Dairy Sci.* 79:676–681.
- SAS Inst. Inc. 2000. *SAS/STAT® User's Guide, Version 8*. SAS Institute, Cary, NC.
- VanRaden, P. M., M. E. Tooker, J. B. Cole, G. R. Wiggans, and J. H. Megonigal Jr. 2007. Genetic evaluations for mixed-breed populations. *J. Dairy Sci.* 90:2434–2441.
- Van Tassell, C. P., G. R. Wiggans, and H. D. Norman. 1999. Method R estimates of heritability for milk, fat, and protein yields of United States dairy cattle. *J. Dairy Sci.* 82:2231–2237.
- Van Vleck, L. D. 1970. Misidentification in estimating the paternal sib correlation. *J. Dairy Sci.* 53:1469–1474.
- Van Vleck, L. D., and G. E. Bradford. 1965. Comparison of heritability estimates from daughter-dam regression and paternal half-sib correlation. *J. Dairy Sci.* 48:1372–1375.
- Van Vleck, L. D., and G. E. Bradford. 1966. Genetic and maternal influence on the first three lactations of Holstein cows. *J. Dairy Sci.* 49:45–52.
- Vinson, W. E. 1987. Potential bias in genetic evaluations from differences in variation within herds. *J. Dairy Sci.* 70:2450–2455.
- Visscher, P. M., J. A. Woolliams, D. Smith, and J. L. Williams. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J. Dairy Sci.* 85:2368–2375.
- Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustment for heterogeneous variance. *J. Dairy Sci.* 74:4350–4357.
- Zwald, N. R., K. A. Weigel, W. F. Fikse, and R. Rekaya. 2003. Identification of factors that cause genotype by environment interaction between herds of Holstein cattle in seventeen countries. *J. Dairy Sci.* 86:1009–1018.