



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection**

**Citation for published version:**

Visscher, P, Woolliams, J, Smith, D & Williams, JL 2002, 'Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection', *Journal of Dairy Science*, vol. 85, no. 9, pp. 2368-2375.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Journal of Dairy Science

**Publisher Rights Statement:**

© American Dairy Science Association, 2002

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection

P. M. Visscher,\* J. A. Woolliams,† J. D. Smith,† and J. L. Williams†

\*University of Edinburgh, Institute of Cell, Animal and Population Biology, West Mains Road, Edinburgh EH9 3JT, UK

†Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK

### ABSTRACT

The proportion of cows in the UK dairy herd whose sires were misidentified was estimated using DNA markers. Genetic marker genotypes were determined on 568 cows (from 168 milk samples and 400 hair samples) and 96 putative sires (from semen samples). The estimated pedigree error rate from the hair samples was 8.8%, and from the milk samples, 13.1%, giving an overall estimate of the error rate of 10%. This level of pedigree errors will have a relatively large impact on the efficiency of progeny testing and the accuracy of cow predicted breeding values. We predict a loss of response to selection of approximately 2 to 3% given this error rate.

**(Key words:** pedigree error, exclusion probability, genetic marker, microsatellite marker, dairy cattle)

### INTRODUCTION

Traditionally, pedigree verification in dairy cattle has been carried out using blood groups and protein polymorphisms (Stormont, 1967). This internationally standardized system has worked well, but has some drawbacks. The number of loci used gives relatively low precision and cannot always exclude a putative sire, even if that putative sire is not the real parent. In addition, blood typing cannot be done retrospectively, e.g., after a sire is dead. Molecular techniques using DNA-based microsatellite markers offer several advantages over traditional blood typing. Any sample from an individual (e.g., hair, milk, saliva) can be used, as long as it contains DNA; hence, the sampling technique can be noninvasive and retrospective from stored tissue or semen samples. The accuracy of the DNA test is much greater than for blood group markers, as DNA markers can have many alleles and a virtually unlimited supply of markers are available (e.g., Kappes et al., 1997). DNA-

based markers are becoming the international standard system of identity verification in livestock.

Correct pedigree information is paramount in a successful breeding program, and its importance has increased with the introduction a decade ago of the “animal model” for national genetic evaluation of dairy cattle. In the animal model, a pedigree error, for example, an incorrectly identified sire of a cow, will affect the PTA of that cow *and* all of her relatives. Usually, the correct assignment of the sire and dam to a calf is not questioned. In fact, the proportion of errors in sire identification in pedigree and nonpedigree herds may vary from a few percent to as much as 22% (Christensen et al., 1982; Geldermann et al., 1986; Ron et al., 1996). Errors, particularly in sire identification, will slow down genetic progress by reducing the accuracy of PTA of cows and bulls (Van Vleck, 1970a, 1970b; Israel and Weller, 2000; Banos et al., 2001).

The UK dairy breeding industry relies on progeny testing for genetic improvement and, therefore, correct identification of sires is immensely important. To date, no estimates were available on the reliability of sire identification in UK dairy herds. The objective of this study was to quantify the level of errors in sire identification in the UK dairy herd using easily accessible sources of samples from which DNA could be extracted.

### MATERIALS AND METHODS

A number of pilot studies were conducted to explore the repeatability and efficiency of DNA extraction from milk, saliva, nasal swabs, and hairs. In addition, the efficiency of obtaining DNA from hair samples taken from calves  $\leq 6$  mo of age was compared with that from sampling hairs from adult cows. These initial experiments showed that hair from adult cows and preserved milk samples were the best DNA sources when considering factors such as time, cost, ease of manipulation, starting amount needed, quantity of DNA obtained, and accessibility of the sample. Hair from adult cows was likely to be a better DNA source than hairs from calves because of the larger size of hair root follicles.

Received November 16, 2001.

Accepted February 4, 2002.

Corresponding author: P. M. Visscher; e-mail: peter.visscher@ed.ac.uk.

**Table 1.** Sires from the milk sample trial.

Name sire	Country of first test
Scottish Washington	UK
Lesmay Superstar	UK
Pittendreich Baron	UK
Picston Aramis	UK
Huddlesford Doraman	UK
Douneside Avenger	UK
Brynhyfryd Cascade	UK
Collycroft Manager	UK
Bankstream Advance	UK
Grove Guardsman	UK
Romandale Elevation Supreme	North America
Hanoverhill Lincoln ET	North America
Hanoverhill Starbuck	North America
Russelldale Promise ET	North America
Mapel Wood Tiger Cat ET	North America
Madawaska Aerostar	North America
Startmore Supreme ET	North America
Linde Alfred	Netherlands
Skalsumer Sunny Boy	Netherlands
F16 Rocket C	Netherlands

### Milk Samples

Milk samples from 23 Scottish herds were kindly provided by Livestock Services UK, Ltd. The herds were selected based upon the number of progeny available from a preselected set of 20 sires. All herds participated in an accredited milk recording regimen. The sires were chosen based on two criteria: availability of a large number of progeny in Scottish dairy herds and the original country in which the sire was tested. The latter was to assess whether there is a larger pedigree error rate in progeny of imported sires than UK-tested sires. Thus, 10 Scottish sires, seven North-American sires, and three Dutch sires were selected for the study (Table 1). According to the database information in March 1998, the number of progeny of the sires in the Scottish milk-recorded herds varied from 200 to 4500. All bulls were proven sires at that time.

DNA was extracted from 421 cows from 23 herds using a method adapted from Koumans and Van Haeringen (1997). Milk sample was stored with preservative in a cold room. Then, DNA extraction was performed as follows; 1.5 ml of milk was centrifuged for 10 min at 13,000 rpm in an Eppendorf centrifuge, the supernatant poured off, and the cell pellet washed by resuspending in 300  $\mu$ l of TE (10 mM Tris-HCl, pH 7.5 1 mM EDTA). The cells were repelleted by centrifugation, resuspended in 200  $\mu$ l milk lysis buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.5% Tween, 0.4% Proteinase-K), then, incubated at 56°C for 45 min and subsequently at 94°C for 10 min. These samples were stored at 4°C until used, at which point 1  $\mu$ l supernatant was added to a 10- $\mu$ l PCR.

All samples were tested for the presence and PCR quality of DNA by scoring a single microsatellite marker. Only samples in which the marker was successfully amplified were sent for full genotyping by a commercial laboratory.

### Hair, Nasal, and Saliva Samples

Hair samples from 550 cows were kindly provided by Genus, Ltd. The cows were sampled on 30 herds that participated in Genus' progeny test program. Some samples were rejected (134) because we were unlikely to be able to get DNA (semen) from the putative sires (i.e., from natural service bulls). A total of 416 samples were provided to a commercial genotyping company which extracted DNA using a standard phenol-chloroform extraction method. A total of 76 different (putative) sires were represented in the 416 progeny sampled.

### Semen Samples

Two straws of semen from putative sires were kindly provided by Genus, Ltd. In addition, commercially obtained semen straws from the Roslin Institute and the University of Edinburgh Langhill herd supply were used. DNA was extracted by the commercial genotyping company using a standard phenol-chloroform extraction technique.

### Genotyping

A standard commercial kit, Stockmarks-II (Perkin-Elmer Applied Biosystems), which contains a set of 11

unlinked microsatellite markers was used to determine genotypes from cows and putative parents. After initial genotype results were obtained, one of the markers (TGLA53) appeared to be unreliable, as the number of sire-progeny mismatches considering this marker alone were considerably more than for the other markers and more than would be expected by chance. For example, of 400 progeny-sire pairs tested, this marker indicated exclusion for 50 pairs, whereas the exclusion probability for this particular marker was 0.48, and the overall proportion of exclusions based upon the remaining markers was approximately 0.1. Hence, it was decided to exclude this marker from further analysis, and the results were analyzed for the remaining set of 10 markers.

### Analysis of Error Rate

For each cow, the presence or absence (0/1) of a detected pedigree error was modeled as a function of the source of DNA (milk or hair), the putative sire (20 for milk samples and 76 for hair samples), the country of first test of the sire (UK, North America, Netherlands) and the herd (23 for milk samples, 30 for hair samples). A simple linear model was used for these analyses.

### Impact on Genetic Progress

Van Vleck (1970a, 1970b) and Christensen et al. (1982) predicted the impact of incorrect parentage on the estimation of genetic parameters and on the effect on genetic gain in a simple halfsib design. If  $p$  is the error rate (nonpaternity rate) in the progeny of bulls, the bias in the estimate of the heritability (the true heritability in the population,  $h^2$ ) from halfsib groups is approximately  $(1 - p)^2$ ,

$$E(h^2) \sim (1 - p)^2 h^2$$

The impact on genetic gain is twofold: 1) The reliability of the bull proofs are reduced because of a lower correlation between the bull proof and the true breeding value, which causes a lower genetic gain than can be achieved when there are no pedigree errors, and; 2) the expected genetic gain is lower than that achieved because of the downward bias in the heritability estimate. Consider an extreme case with very large progeny groups. Then, even with a proportion of daughters misclassified, the bulls will be ranked correctly. However, the expected response will be too low because the estimate of the heritability is too low. Following the derivations of Van Vleck (1970a, 1970b) and Christensen et al. (1982), the total impact of the pedigree error rate can be summarized by calculating the reliability ( $R$ ) of bull proofs:

Theoretical maximum	$R_0 = n/(n + \lambda_0)$	$\lambda_0 = (1 - t)/t$
Achieved	$R_1 = n/(n + \lambda_1)$	$\lambda_1 = (1 - qt)/(q^2t)$
Expected	$R_2 = n/(n + \lambda_2)$	$\lambda_2 = (1 - q^2t)/(q^2t)$

where  $n$  = the number of putative progeny per bull;  $q$  =  $(1 - p)$ , the proportion of correctly classified daughters; and  $t$  = the intraclass correlation,  $(h^2/4)$ . For a given amount of genetic variation and selection pressure, the rate of gain is proportional to the square root of the reliability, and so, the efficiency of response to selection is the square root of the ratio of the achieved and theoretical maximum reliabilities.

$$\begin{aligned} \text{Efficiency} &= [(n/(n + \lambda_1))/(n/(n + \lambda_0))]^{0.5} \\ &= [(n + \lambda_0)/(n + \lambda_1)]^{0.5}, \end{aligned}$$

and the relative loss in response is,  $\text{Loss} = 1 - \text{Efficiency}$ . These equations were used to predict the efficiency and loss in response to selection for a number of different values of  $n$ ,  $h^2$ , and  $p$ .

## RESULTS

### Marker Characteristics

Tables 2 and 3 summarize the characteristics of the markers. All markers were highly polymorphic in the sample, with a range of animals heterozygous for markers varying from 0.57 (SPS115) to 0.84 (TGLA227). The larger the heterozygosity of markers, the better they are for identity validation. This is summarized in Table 2, by calculating the exclusion probability per marker and for all markers combined. The exclusion probability is the probability that a putative sire can be excluded as a parent, given a case of nonpaternity, and is calculated from population frequencies of marker alleles (e.g., Dodds et al., 1996). For individual markers in this study, the exclusion probability ranges from 0.19 (SPS115) to 0.52 (TGLA227). Both the heterozygosity and exclusion probability for TGLA227 are identical to the values estimated by Heyen et al. (1997) from the US Holstein population.

For illustration purposes, we describe one locus (BM1824) in more detail. Six alleles were found, with frequencies of 0.256, 0.232, 0.255, 0.002, 0.212, and 0.043 (these frequencies sum to 1.00). Heterozygosity ( $H$ ) is the probability that a random sample of two alleles from the population are different, i.e., that two alleles are the same:

$$\begin{aligned} H(\text{BM1824}) &= 1 - 0.256^2 - 0.232^2 - 0.255^2 - 0.002^2 \\ &\quad - 0.212^2 - 0.043^2 = 0.77. \end{aligned}$$

The exclusion probability is the probability that two random (unrelated) individuals do not share any al-



**Table 2.** Marker characteristics using the hair sample data.

	Marker									
	1	2	3	4	5	6	7	8	9	10
Name	TGLA227	BM2113	ETH10	SPS115	TGLA126	TGLA122	INRA23	ETH3	ETH225	BM1824
No. alleles	12	7	8	6	5	16	10	9	7	6
Heterozygosity	0.84	0.80	0.66	0.57	0.69	0.80	0.77	0.74	0.77	0.77
Exclusion probability	0.52	0.43	0.26	0.19	0.27	0.45	0.39	0.35	0.38	0.36
Cumulative										0.990

les, so that parentage can be excluded. For example, if one individual has alleles 1 and 2 (denoted 1/2) and another, alleles 3/6, parentage can be excluded. The exclusion probability is calculated by adding the probabilities for all genotype-pair combinations which results in exclusion (see Dodds et al., 1996, for an excellent review). For marker BM1824, the exclusion probability is 0.36.

Table 3 summarizes the distribution of loci which show a sire-progeny mismatch, given a case of nonpaternity. The probability was calculated from the individual exclusion probabilities by considering all possible combinations of outcomes if we sample from 10 independent Bernoulli distributions with different probabilities. For example, the probability for 1 out of the 10 loci showing a mismatch, given a case of nonpaternity, is calculated by summing up the probabilities:

Prob(marker 1 shows exclusion and all other markers show no exclusion) + Prob(marker 2 shows exclusion and all other markers show no exclusion) + ... + Prob(marker 10 shows exclusion and all other markers show no exclusion):

$$\begin{aligned}
 &= (0.52 * (1 - 0.43) * (1 - 0.26) * \dots * (1 - 0.36) \\
 &+ ((1 - 0.52) * 0.43 * (1 - 0.26) * \dots * (1 - 0.36) \\
 &+ \dots + ((1 - 0.52) * (1 - 0.43) * \dots * 0.36) \\
 &= 0.062.
 \end{aligned}$$

Hence, even if there is a misidentification, the probability of seeing only 1 out of the 10 loci showing exclusion is more than 6%. This is then extended by summing the probabilities of observing 2 out of 10 showing exclusion, 3 out of 10, etc. The probability of seeing 2 out of 10 loci showing exclusion is 0.163 (Table 3). Note that by chance we can observe that none of the 10 loci show exclusion, because two individuals from the population can by chance share an allele at each of the 10 loci ( $P = 0.01$ , see Table 3). One minus this probability is equal

to the exclusion probability using all 10 markers and corresponds to the value of 0.99 in Table 2.

### Milk Samples

DNA was extracted from 421 milk samples. The cows were putative progeny from 20 sires (see Table 1). Unfortunately, only 168 (40%) of DNA samples yielded either full or partial genotyping information, which is a very low success rate. However, after exhaustive testing by ourselves and the genotyping company that carried out the work, the source of the problem was not revealed. The likely explanation was a combination of 1) variation between milk samples in the number of somatic cells present in the milk (these cells provide the DNA), 2) variation between milk samples in unidentified (bio)chemical compounds which may have interfered with DNA extraction, or remain in the DNA preparation and interfere with the genotyping reaction, 3) degradation of DNA during storage (at +4°). From these results, we cannot recommend the use of milk for routine genotyping at present without a full phenol extraction being carried soon after the samples are collected.

Of the 168 samples with partial or complete genotypes, 39 (23%) showed exclusion of the putative parent at one or more loci. If we take only the samples that showed exclusion at two or more loci, the error rate reduced to 13%. These rates of misidentification are large, particularly given that they are an underestimate of the population-wide error rate because of partial genotypic information. The error rates are summarized in Table 4.

### Hair Samples

After genotyping the cows using hair samples and checking for the availability of genotypes of putative sires, 400 samples were left. Of these 400 samples, the

**Table 3.** Probability of the number of loci showing nonpaternity, given a random case of nonpaternity.

0	1	2	3	4	5	6	7	8	9	10
0.010	0.062	0.163	0.250	0.248	0.165	0.075	0.023	0.005	0.001	0.000

**Table 4.** Summary of sire misidentification rates (proportion  $\pm$  SE).<sup>1</sup>

Sample	No.	No. markers showing exclusion	
		1 or more	2 or more
Milk	168	0.232 (0.03)	0.131 (0.03)
Hair	400	0.125 (0.02)	0.088 (0.01)

<sup>1</sup>SE =  $\sqrt{\hat{p}(1 - \hat{p})/n}$ , with  $\hat{p}$  the proportion and  $n$  the sample size.

proportion of missing genotypes, i.e., the number of unknown marker genotypes as a proportion of the total number of possible marker genotypes ( $= 400 \times 10$ ), was low, 3%, indicating a high success rate of DNA extraction and subsequent genotyping. In none of the samples were all 10 of the marker genotypes missing. The sire was misidentified for 50 samples, i.e., an error rate of 0.125. When misidentification was based upon two or more exclusions, the error rate dropped to 0.088.

### Model Fitting

A simple linear model was fitted to the data. Data were the presence or absence of an identification error on  $400 + 168 = 568$  observations. The narrow definition of identification error was used, i.e., two or more markers had to show exclusion. There was no significant effect of sampling group (hair/milk), sire, or country of origin. Sampling group approached a significance level ( $P < 0.12$ ). Herd of sampling was tested for hair samples ( $N = 400$ ), and no significant effect was found ( $P < 0.17$ ).

### Loss in Response to Selection

A number of examples of the reliabilities and loss in genetic gain as a function of  $n$ ,  $h^2$ , and  $p$ , are given in Table 5. For example, for an error rate of 10%, a heritability of 25%, and 50 progeny per sire, the reliability is reduced by 5% (relative to 0% errors), and genetic progress is reduced by 3%. The variance in PTA is proportional to the reliability, so that a reduction of 5% in reliability corresponds to a reduction of about 2.2% in the SD of bull proofs. These values may seem small, but are similar to the predicted increase in genetic gain of including additional traits in a national selection index or similar to the likely increases in reliability by changing from our current evaluation system to a test-day model based genetic evaluation system.

## DISCUSSION

### Sampling Methods

Our study suggests that the most robust and reliable way to obtain noninvasive DNA samples is by taking hair samples. It is not clear why the milk samples

showed such variable and disappointing results, given the fact that our preliminary work (pilot study) was promising. As highlighted before, the likely causes are variation between milk samples in the number of somatic cells and unidentified (bio)chemical compounds, degradation of DNA during storage, and perhaps non-transportability of extraction methods across laboratories. As a result of our findings, we cannot recommend the use of milk for routine genotyping, particularly on large numbers of samples of limited size.

Herds from which milk samples were taken were in an accredited milk-recording regimen and were believed to be representative for the Scottish population of milk-recorded herds. Herds from which the hair samples were taken participated both in milk recording and in a progeny-testing regimen, and are geographically separated from those with milk samples (in England). Both types of herds should reflect the UK cow population in which bulls are progeny tested and to which genetic progress is disseminated.

### Pedigree Error Rate

Our results show a relatively large pedigree error rate of  $\sim 10\%$ . There are many reasons for such errors (Christensen et al., 1982) including: 1) mistakes by AI company or companies in labeling semen; 2) AI technicians incorrectly identifying semen straws by mistake; 3) the insemination of cows holding to a previous insemination; 4) errors when the bull's herdbook number or name is entered into the insemination record; 5) the use of natural-service bulls leading to pregnancies of previously inseminated cows which were assumed to be pregnant from the AI bull; 6) mistakes in sire identification when a cow enters the milking herd in schemes where pedigree information on milk recorded cows is obtained through the milk recording program; and 7) interchange of calves at birth.

Some of these causes could be prevented by a good recording system and verification systems at AI. For example, if insemination records and milk recording (including date of calving) are linked and a single cow ID assigned throughout her lifetime, 3), 5), and 6) could be avoided. Causes 1) and 2) could be minimized by quality control measures taken at the AI companies.

**Table 5.** Response to selection of sires in the presence of random pedigree errors.

Error rate <sup>1</sup>	$h^2$	$n^2$	Reliability <sup>3</sup>	Response to selection <sup>4</sup>
0.00	0.50	50	1.00	1.00
0.05	0.50	50	0.99	0.99
0.10	0.50	50	0.97	0.98
0.15	0.50	50	0.95	0.98
0.20	0.50	50	0.93	0.96
0.00	0.25	50	1.00	1.00
0.05	0.25	50	0.97	0.99
0.10	0.25	50	0.95	0.97
0.15	0.25	50	0.92	0.96
0.20	0.25	50	0.88	0.94
0.00	0.50	100	1.00	1.00
0.05	0.50	100	0.99	1.00
0.10	0.50	100	0.98	0.99
0.15	0.50	100	0.97	0.99
0.20	0.50	100	0.96	0.98
0.00	0.25	100	1.00	1.00
0.05	0.25	100	0.99	0.99
0.10	0.25	100	0.97	0.98
0.15	0.25	100	0.95	0.98
0.20	0.25	100	0.93	0.96

<sup>1</sup>Random pedigree error rate.<sup>2</sup>Progeny group size.<sup>3</sup>Reliability relative to the case of no pedigree errors.<sup>4</sup>Response to selection relative to the case of no pedigree errors.

Interestingly, only 7) would give rise to equal identification error of sires and dams. Hence, a study in which trios of heifers/cows and both their putative parents are sampled should pinpoint whether the errors are because of poor recording or whether large herd sizes are involved.

Genotyping all cows and bulls in a population would enable pedigree errors to be identified and corrected as long as all putative sires have been genotyped. In addition, the genotype of a cow could be used for positive identification at any stage from birth, during herd life, and at slaughter, thus superseding the costly and error-prone tracking of cattle by ear tagging. However, the genotypic information on an animal should remain with that animal, for example, in the case of animals moving to a different farm, so that tagging systems, such as implanted electronic devices or ear tags, will remain important.

There are few reports on the pedigree error rates in the literature. Christensen et al. (1982) reported rates between 5 and 15% in Danish dairy cattle, Geldermann et al. (1986) found a misidentification rate of 13% in German dairy cattle, and Ron et al. (1996) found a sire error rate of 5% in Israeli Holstein cows, based upon genotyping 12 microsatellite markers on 173 cows and their putative four sires. If exclusion was based on at least two markers, the error rate was reduced to 2%. Although the samples are much smaller than in the present study, the detected error rate in Israeli Holstein cows appears significantly lower than found in this

study, whereas the other reported proportions of misclassifications are similar to our estimates in the UK dairy herd. The estimated pedigree error rates in the UK detected by blood typing over the past 30 yr were approximately 10% (R. L. Spooner, personal communication).

Geldermann et al. (1986) investigated the effect of sire misidentification on genetic gain and concluded that the loss in response could be similar to the proportion of progeny misidentified. Subsequently, Ron et al. (1996) used these results (i.e., 5% misidentification rate equals 5% reduction in genetic progress) and concluded that paternal verification of test daughters can be economically justified. We disagree with the theory and conclusions from Geldermann et al. (1986). The authors appear to show in their paper that both the accuracy of selection and selection intensity are reduced in the presence of pedigree errors. However, in their calculations they 1) treat sires as fixed rather than random, and 2) omit the Mendelian sampling term in the derivation of the variance of predicted breeding values. In agreement with Van Vleck (1970a, 1970b) and Christensen et al. (1982), we conclude that the loss in response to selection due to random pedigree errors in a simple progeny test scenario is a function of the reduction in heritability, and, therefore, reliability only. This becomes intuitively obvious when we consider very large progeny groups. Even if half of the progeny are randomly misclassified, the ranking of the average per-

formance of the progeny group would still correspond to the ranking of the true breeding values of the sires.

Israel and Weller (2000) conducted a stochastic simulation study of a large dairy cattle population undergoing selection to investigate the effect of a 10% incorrect paternity on genetic progress over a period of 20 yr. They estimated that the annual rate of response was reduced by 3 to 4%. Our simple predictions, based on a reduced reliability, suggests a loss in response of 2% for one round of sire selection using the same heritability (0.25) and progeny group size (100) as in Israel and Weller (2000). The difference presumably reflects both the impact of incorrect identification of both sexes (e.g., daughters of dams with incorrect pedigrees occurred in the simulation of Israel and Weller) and the accumulated effect of pedigree errors over multiple generations. In agreement with our prediction, Israel and Weller (2000) found that the estimated genetic gain was lower than the actual genetic gain.

Banos et al. (2001) used actual data from the US dairy population, simulated a random and unidentified 11% paternity error rate, and reevaluated cow and bull PTA, using the same heritability as is used for the routine evaluations ( $h^2 = 0.3$ ). By using 0.3 in his reevaluation, Banos was assuming this value to be the 'true' heritability. This differs from Van Vleck (1970a): When Van Vleck assumed unbiased estimates of variance components, it was also assumed that the fraction of cows that had been misspecified was also known. In this study, we are dealing with an additional situation in which there was no bias in the estimates of the variance components, but it was incorrectly assumed that there were no pedigree errors. For milk yield, the genetic trend for cows and bulls was reduced by 11 and 14%, respectively, and the variance in EBV was reduced by 9%. Using our simple one-generation predictions, which are based upon Van Vleck (1970a, 1970b) and the parameters of Banos et al. (2001), i.e., an average progeny test size of 57 daughters and a heritability of 0.3, we attempted to predict their results. The reduction in the variance of sire EBV relative to the case of no paternity errors is predicted as  $q^2R/R_1 = q^2(n + \lambda_1)/(n + \lambda_0)$ , with  $R$  and  $R_1$  the assumed and realized reliability, respectively. For  $q = 0.89$ ,  $n = 57$ , and  $t = 0.075$  ( $= 1/4 \times h^2$ ), this results in a predicted proportional reduction in sire EBV variance of 0.83, or a proportional reduction in the sire EBV standard deviation of 0.91 (9% reduction), exactly as observed by Banos et al. (2001). Since Banos et al. (2001) used the same heritability (0.3) in their genetic reevaluation, the assumed regression of BV on EBV was 1.0. Hence, the reduction observed by Banos et al. in the apparent genetic trend was mainly a consequence of the reduction in the variance of EBV. The achieved genetic trend (which could not be observed

in Banos et al.) would be larger than the apparent trend, because the achieved regression of BV on EBV would be larger than 1.0. Using the one-generation argument and assumption, the prediction of this regression coefficient is  $R_1/(qR)$ , or 1.07, from the parameters of Banos et al. (2001). Hence, predicted loss in achieved response is only  $1.07 \times 0.91 = 0.97$ , i.e., 3%. It could be argued that the apparent genetic trend is important for a breeder or breeding company, whereas the achieved trend is important for the farmer.

In the predictions of the loss in genetic gain, it was assumed that pedigree errors are random with respect to the sire and that the error rate was the same for all bulls. In practice, it may be that there are differential pedigree error rates for specific groups of bulls and that there is variation in the error rate across bulls. For example, some bulls may have all their progeny correctly identified, while others may have 10 to 20% of progeny misclassified. Natural service bulls, progeny test bulls, and proven bulls may, on average, have different error rates. A differential error rate between bulls could result in a reduction in the predicted loss of genetic progress if those sires that contribute most genes to the population, i.e., the proven elite sires, have a lower degree of misspecified progeny.

The pedigree errors also effect the estimation of genetic parameters in the population and, this, in turn, may alter genetic properties such as inbreeding rates. If the misspecification was to be removed or reduced, the heritabilities in the population would be higher. In the absence of any selection procedure to explicitly manage rates of inbreeding, inbreeding rates are lower when the heritabilities are higher (Grundy et al., 1994), since less reliance is placed on information from collateral relatives. While this will be offset by the pedigree errors themselves, there is the potential for high rates of pedigree errors to encourage greater loss of genetic variation. The balance between these opposing trends has, to our knowledge, not been examined.

When using a commercial service for pedigree testing, the client (breeding company or farmer) should be aware of the rules used to declare whether a parent can be excluded. Using our results, excluding a parent based on two or more individual markers showing exclusion would miss 6% of cases in which there really was a case of nonparentage. However, these calculations ignore the error rate of genotyping. Presumably, for the provider of the parentage verification service, declaring nonpaternity when that is false is worse than not declaring exclusion in a case of nonpaternity. Declaring nonpaternity when a single marker shows exclusion would increase the pedigree error detection rate but would also increase the false-positive rate. These problems could easily be resolved if more markers were



used in the parentage test. If 20 to 30 markers were used, there should be no or negligibly small ambiguity in the outcome, even in the presence of genotyping errors. In conclusion, we have estimated the sire misidentification error from a sample of 400 hair samples and 168 milk samples from the UK national Holstein-Friesian dairy cattle population. The estimate of the pedigree error rate was 10%, which has a relatively large impact on genetic progress in this country.

### ACKNOWLEDGMENTS

We thank the Milk Development Council for financial support (project 96/R2/01); Genus, Ltd. and Livestock Services UK, Ltd. for their generous supply of samples; Roel Veerkamp and Jenny Pryce for help with database queries, and the referees for helpful suggestions.

### REFERENCES

- 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.
- Christensen, L. G., P. Madsen, and J. Petersen. 1982. The influence of incorrect sire identification on the estimates of genetic parameters and breeding values. *Proc. 2nd World Congress on Genetics Applied to Livestock Production*, Madrid, Spain. 7:200–208.
- Dodds, K. G., M. L. Tate, J. C. McEwan, and A. M. Crawford. 1996. Exclusion probabilities for pedigree testing farm animals. *Theor. Appl. Genet.* 92:966–975.
- 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.
- Grundy B., A. Caballero, E. Santiago, and W. G. Hill. 1994. A note on using biased parameter values and nonrandom mating to reduce rates of inbreeding in selection programs. *Anim. Prod.* 59:465–468.
- Heyen, D. W., J. E. Beever, Y. Da, R. E. Evert, C. Green, S. R. E. Bates, J. S. Ziegler, and H. A. Lewin. 1997. Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semi-automated parentage testing. *Anim. Genet.* 28:21–27.
- Israel, C., J. I. Weller. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *J. Dairy Sci.* 83:181–187.
- Kappes, S. M., J. W. Keele, R. T. Stone, T. S. Sonstegard, T. P. L. Smith, R. A. McGraw, N. L. LopezCorrales, and C. W. Beattie. 1997. A second-generation linkage map of the bovine genome. *Genome Res.* 7:235–249.
- Koumans, J. T. M., and H. Van Haeringen. 1997. A simple and rapid method for the isolation of DNA for genotyping purposes from milk samples. *Genet. Mol. D023*:147.
- 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.
- Stormont, C. 1967. Contribution of blood typing to dairy science progress. *J. Dairy Sci.* 50:253–260.
- Van Vleck, L. D. 1970a. Misidentification in estimating the paternal-sib correlation. *J. Dairy Sci.* 53:1469–1474.
- Van Vleck, L. D. 1970b. Misidentification and sire evaluation. *J. Dairy Sci.* 53:1697–1702.