Quality and Validation of Insemination Data for National Genetic Evaluations for Dairy Cow Fertility in the United Kingdom

Haja N. Kadarmideen [1,2,†] ***and Michael P. Coffey*** [2]
[1] *Statistical Animal Genetics Group, Institute of Animal Science, Swiss Federal Institute of Technology, ETH Zentrum, Zurich CH 8092, Switzerland,* [2] *Animal Breeding and Genetics Department, Scottish Agricultural College, West Mains Road, Edinburgh EH9 3JG, U.K*
[†] *e-mail: haja.kadarmideen@inw.agrl.ethz.ch*

1. Introduction

Sub-optimal fertility in dairy cows is a major cause of involuntary culling and hence cost in many countries. Antagonistic genetic correlations between fertility traits and milk yield (e.g. Kadarmideen et al., 2000 and 2001) would lead to a decline in cow fertility, if selection is for milk production only. Economic implications of this decline in fertility are also of major concern. Thus routine national sire evaluations for daughter fertility are essential to help farmers and breeding companies to select the best bulls for a combination of production and fertility of their daughters. The major hurdle arises in the form of poor quality of insemination and pregnancy data at the national level in the UK.

Insemination and calving dates are used in national recording schemes to compute a variety of interval and score traits that are used as measures of fertility. Unbiased and accurate fertility breeding values depend on the reliability and accuracy of recording reproductive events at herd-level. Unlike Scandinavian countries, recording insemination dates are voluntary in UK and many other countries, which introduces uncertainty in data quality. In voluntary recording schemes, a complete absence or under-recording of service dates and the absence of pregnancy diagnosis are common. This will inevitably lead to under-estimation in fertility measures such as number of services per conception (NSPC), number of days to first service (DTFS), interval between first and last service (IFLS) and overestimation of success rates in conception to first service (STFS). The objective of this study, therefore, is to characterise and compare insemination and calving data collected from two of the national milk recording schemes that differ in their accuracy of recording. Other aims were to investigate their reliability and quality and to provide validation rules for routine fertility genetic evaluations in the United Kingdom and other countries where voluntary recording of insemination events exists.

2. Materials and Methods

Insemination and calving records on Holstein Friesian cows were obtained from two of the UK national recording schemes, offered by Livestock Services UK (LSUK) Ltd and Dairy Information System (DAISY), operated by the University of Reading. Unlike LSUK, DAISY scheme is a computer-based comprehensive recording scheme for insemination and health events. DAISY is used by a group of co-operating farmers and veterinarians who are specifically interested in monitoring fertility and health in their herds.

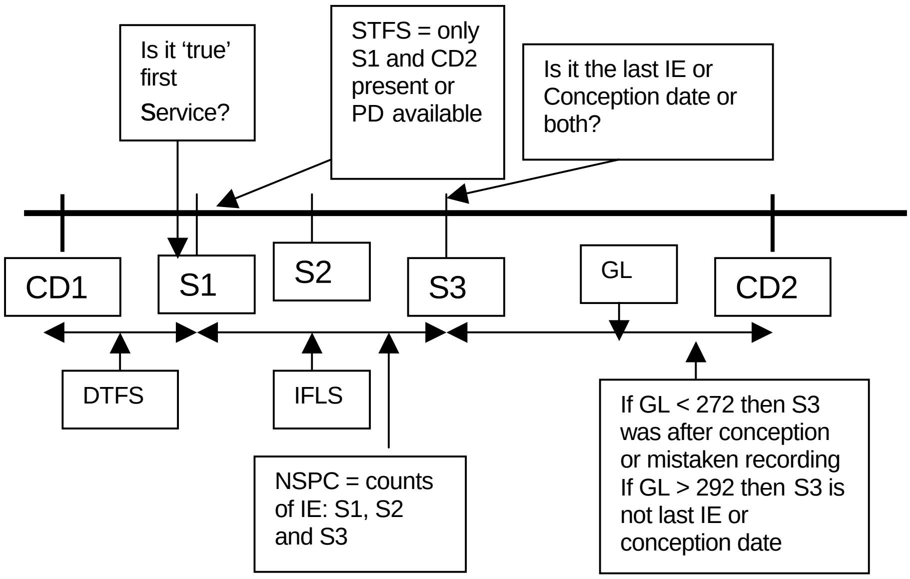
In the LSUK database, there was a total 177,620 lactation records from 932 herds. For each herd, the number of cow records with a complete absence of service or insemination events (IE) compared to the total number of records was calculated (percentage of records with no insemination event; PNIE). The number of herds (N_h) belonging to different ranges of PNIE (10 to 100%) was then calculated. The PNIE was further classified by different herd-size (HS). The PNIE was also computed for 4 years and 4 seasons of calving. Similar calculations were also performed for about 40,000 lactation records from 134 DAISY herds chosen because of known recording quality.

The reliability and accuracy of IE, when available, were assessed by using gestation length (GL), as GL is biologically determined and has low phenotypic variation. GL was computed as the difference between the last IE and subsequent calving date (CD). Assuming "true" gestation length (mean of 282 days) to be constant, one oestrous cycle will result in the maximum of 21 days around this mean. Hence, a GL outside 272 and 292 (282 ±10) days were taken as an 'unacceptable' range for any record. The standard deviation (SD) of true gestation length is about 5 days (Holstein UK and Ireland 2000). If we consider records outwith mean ± 2 SD as outliers, then the restriction of 282 ± 10 days could be considered as a statistically appropriate editing procedure.

Figure 1 illustrates how insemination events (S1, S2, S3,...) could be recorded within lactation and uncertainties associated with the reliability of these events when recorded voluntarily. As depicted in Figure 1, a GL of more than 292 days was taken to indicate that the exact IE that resulted in conception was unknown or there were one or more IE after the last insemination date (S3) that were not recorded. A GL of less than 272 days was assumed to be the result of an inappropriate insemination or mistaken recording of an insemination when the cow was already pregnant (Figure 1). However, the editing based on GL can only validate whether the last IE was correct. In the case of multiple IE under optional recording, it will be difficult to ascertain whether the first recorded IE was genuinely the first or the true first IE was not recorded. In the latter case, the second or subsequent IE would be mistakenly treated as first IE. This adds complexity to validation of IE.

When information on pregnancy diagnosis (PD) is unavailable, as for this data, GL could be used to *partially* validate fertility measures such as NSPC, STFS, DTFS, IFLS as shown in Figure 1. Records that fell within a range of 272 and 292 days were considered as 'useable insemination records' (UIR) for genetic analysis. The distribution of N_h across different percentage of UIR (PUIR) classified by HS was also computed as for PNIE (shown in Figure 3). For analysis of CI, the number of records lost due to the absence of calving interval (CI) or subsequent calving date (CD2) were computed. The proportion of cows without IE that did and did not have an associated CI record was computed for both schemes.

Figure 1. Illustration of calving and insemination events and associated uncertainties in data quality in voluntary recording schemes



3. Results and Discussion

This analysis was aimed at investigating data quality and validation of insemination data from LSUK recording scheme and making comparisons with a specialised fertility recording scheme (DAISY) in the United Kingdom. LSUK operates a free and voluntary recording scheme for recording health and service information. There are 180 out of about 1000 LSUK herds (18%) that are known to actively participate in health and fertility recording, hence the data quality from these herds is expected to be higher than the overall data quality from all 1000 farms. The assessment of overall data quality from all 1000 farms, however, showed that about 50% of lactation records did not have any insemination event recorded. These data would have to be rejected if genetic analysis is based solely on IE. When non-contributing records (those without IE) are rejected for genetic analysis, selection bias may be introduced if rejected data was missing 'non-randomly'. However, results showed that IE are 'randomly' missing across herds although large herds had a larger proportion of missing IE records. These herds made a substantial contribution to the overall rejection rate of 50% of available data. This may be due to some larger herds not recording any IE or using an alternative method of recording fertility data for management purposes such as an on-farm PC. The insemination data quality issues discussed here are also similar to recording disease events by LSUK and DAISY (e.g. clinical mastitis) where herds would be selected for genetic analysis only if at least one disease event is recorded in a given herd-year of calving (Kadarmideen and Pryce 2001; Kadarmideen et al., 2001).

Figure 2 shows the distribution of proportion of herds for different proportion of missing IE records. Table 1 shows the same distribution as Figure 1 but cross-classified by different herd-sizes. Results show a range of missing information across herds. There is a case where about 15% of herds have less than 20% cow records with missing IE (or more than 80% records had at least one IE recorded). At the extreme, about 15% of herds have 90% to 100% of cow records with a complete absence of IE (or only less than 10 % of cow records have at least one IE recorded). There is a further loss of data among records that had at least one IE depending whether it is useable or not, as determined by gestation length in this study (PUIR). Hence, availability of at least one IE for a cow record does not guarantee that it will be used for analysis. This leads to an even greater loss of insemination records (see Figure 3).

The poor reliability of service data in national recording schemes could be expected, because fertility recording is optional. Within this voluntary recording framework, several combinations of management preferences in recording insemination dates contribute to the problem of data quality: some farms record all insemination dates, some only a few, some record the last insemination date that resulted in pregnancy and some record none at all. This inconsistency is complicated by changes in the recording patterns within the same farm over time (e.g. when the herdsman goes on holiday or leaves). As mentioned earlier, this poses uncertainty over correctness and accuracy of what is available on an individual cow basis and hence in subsequent genetic evaluations.

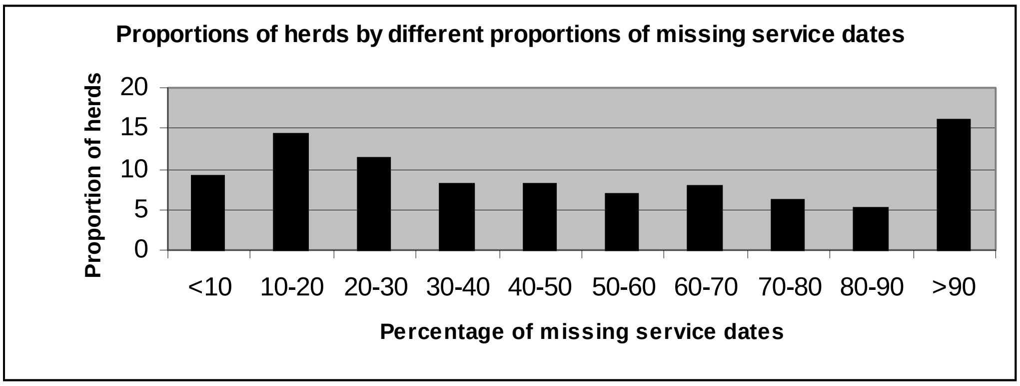


Figure 2. Distribution of LSUK herds with different proportion of missing service dates

Table 2 shows that the distribution of number of herds falling within different proportions of useable insemination records classified by herd-size. Results show that about 20% and 8% of herds have more than 90% and less than 60% UIR, respectively. Table 2 also shows different proportions of herds for all possible combinations of PUIR by HS. Distribution of UIR across 4 years of calving ranged from 47% to 49% with an average of 43% and across 4 seasons of calving ranged from 32% to 38% with an average of 36%. Out of 50% available IE records, there was 83.8% UIR (or 16.2% records unusable). The mean and SD of GL in the 83.8% of UIR was 282.53 and 4.59 days, respectively. The mean and SD of GL in the 16.2% of unusable insemination records was 293.5 and 32.5 days, respectively. The abnormal mean and SD of GL is due to missing IE and will be closer to normal limits when missing insemination dates can be *imputed*. For herds with a mean GL > 288 days (upper 3rd of the range), the SD of GL was 18.2 days. These characteristics of unusable insemination records suggest that these are records from herds that are inconsistent in reporting IE. The summary for each herd in terms of mean and SD of GL indicated that herds with less than 60% UIR (Table 2) have a relatively high mean and SD of GL (results not shown) which again emphasises the inconsistent recording. Distribution of UIR across 4 years and seasons of calving ranged from 83% to 87% with an average of 84%. The overall percentage of records with no CI record was 15.7%. Among 50% of cow records with no IE, 23.21 % of cows did not have a CI record.

Table 1. Distribution of percentage of herds for different PNIE (row-wise) classified within different HS (column-wise). Column total is percentage of herds with a given PNIE and row-total is percentage of herds with a given HS

	Per cent records with no insemination event (PNIE)					
Herd Size	<20%	20-40%	40-60%	60-80%	80-100%	Total
<= 50	17.91	26.87	20.40	23.88	10.95	**21.57**
51-100	31.38	17.55	17.02	14.89	19.15	**20.17**
101-150	27.81	27.15	11.26	11.26	22.52	**16.20**
151-200	25.17	23.78	25.87	12.59	12.59	**15.34**
201-250	25.97	22.08	12.99	9.09	29.87	**8.26**
> 250	16.86	10.47	13.95	10.47	48.26	**18.45**
Total	**23.82**	**21.14**	**17.27**	**14.59**	**23.18**	**100.00**

Comparison of fertility data quality from LSUK and DAISY schemes

All DAISY herds had at least one IE recorded since these were herds selected as being interested in fertility recording. At least one IE was present for 97.4% records in DAISY whereas it was only 50% of records in the LSUK scheme Although recording insemination events is also *voluntary* for DAISY farmers, these results show that recording is relatively more accurate and complete in DAISY than LSUK. Furthermore, computerised data validation that checks the logic of fertility recording ensures improved data quality. Hence, the DAISY scheme allows a greater degree of confidence in the results (Kadarmideen et al., 2001). It is not surprising because the DAISY recording scheme is specifically designed for active recording of insemination, calving and health events. Even within the DAISY scheme, when available IE was validated using gestation length, it was found that one or more IE were missing for 13.7% of records. This however, may correspond to natural services that were not recorded in some farms. It may also be because of poor estimation of pregnancy length by vets during pregnancy diagnosis, missed heat periods or it could genuinely indicate the problem of voluntary recording of insemination events even within specialised recording schemes. Although quality of insemination data is high in the DAISY scheme, it currently suffers from the lack of sufficient data to produce national genetic evaluations for fertility, for which large volumes of data would be needed to achieve acceptable reliability in the bull proofs. The volume of data from LSUK, even after loss of data due to stringent editing, was sufficient to achieve acceptable reliability in the bull proofs.

Table 2. Distribution of percentage of herds for different PUIR (row-wise) classified within different HS (column-wise). Column total is percentage of herds with a given PUIR and row-total is percentage of herds with a given HS

	Percentage of useable insemination records (PUIR)					
Herd Size	>90%	80-90%	70-80%	60-70%	<60%	Total
< 50	18.11	36.49	17.27	12.81	15.32	**44.76**
51-100	18.00	47.50	20.00	9.50	5.00	**24.94**
101-150	26.02	51.22	15.45	6.50	0.81	**15.34**
151-200	20.97	64.52	9.68	4.84	0.00	**7.73**
201-250	18.75	53.13	28.13	0.00	0.00	**3.99**
>250	30.77	53.85	15.38	0.00	0.00	**3.24**
Total	**19.95**	**44.89**	**17.46**	**9.48**	**8.23**	**100.00**

Recommendations on Future Recording and Genetic Evaluations for Fertility

Accuracy and reliability of insemination dates and hence fertility genetic evaluations could be improved via two approaches. One is by aiming for consistent and uniform recording of insemination dates (e.g. recording all insemination dates at all times) by supplying guidelines to farmers, vets and MROs or by providing incentives to do so. The second is to apply editing and validating rules, as shown here, to salvage those records that can be salvaged and to eliminate records that have missing insemination dates and have the potential to introduce bias in genetic evaluations. Better data quality could be achieved at the recording level by designing appropriate software that checks records at the point of entry into the database.

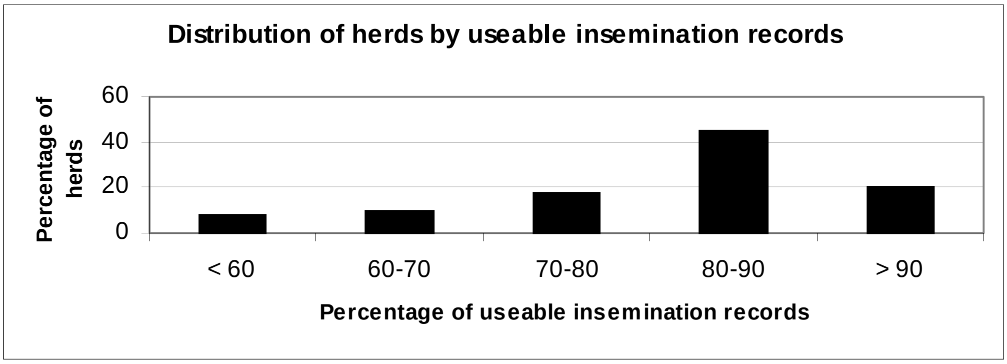


Figure 3. Distribution of LSUK herds with different proportion of useable insemination records

Future recording for dairy cow fertility should address two essential components: first the cyclicity- how early the cow comes to heat after calving; and the second, the ability to conceive at the first insemination. Because heat observations are difficult to record, number of days between calving and first service can be used, instead. Avoiding either of these criteria (cyclicity or pregnancy rate) may not reflect true fertility. Therefore these components should both be addressed. Future national genetic indexes therefore should eventually include at least two fertility measures for genetic evaluations of bulls and cows when accurate data becomes available. Based on the above data validation checks, current national data sets may be used for calculating bull fertility breeding values. The availability of fertility indexes may also subsequently stimulate better recording of insemination records by farmers.

4. Conclusions

In this study, we characterised UK national insemination data, provided data validation rules and investigated the reliability and accuracy of currently available data from two major recording schemes. Recommendations for future recording were made. An optimum between data quality and data rejection needs to be determined in future. Current findings would form a basis for using insemination data for bull evaluation for daughter fertility based on direct measures.

Acknowledgement

The authors thank LSUK and DAISY for providing data and the UK Milk Development Council for funding this study.

References

Kadarmideen, H.N., Thompson, R. & Simm, G. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. *Anim. Sci. 71*, 411-419.

Kadarmideen, H.N. & Pryce, J.E. 2001. Genetic and economic relationships between somatic cell count and clinical mastitis and their use in selection for mastitis resistance in dairy cattle. *Anim. Sci. 73,* 19-28.

Kadarmideen, H.N., Rekaya, R. & Gianola, D. 2001. Genetic parameters for clinical mastitis in Holstein-Friesians: a Bayesian analysis. *Anim. Sci. 73,* 000-000

Kadarmideen, H.N., Coffey, M.P., Kossaibati, M.A., Thompson, R. & Esslemont, R.J. 2001. Multiple trait genetic parameter estimates and evaluations for dairy cow fertility with and without milk yield. *Livest. Prod. Sci.* (submitted). Also as paper for BSAS Annual meeting, University of York, York, UK, 9-11 April 2001.