Data Quality Assessment and Preliminary Investigations on Fertility in the Italian Holstein Friesian

S. Biffani, A.B.Samoré, F.Canavesi and M. Marusi
A.N.A.F.I. Italian Holstein Breeders Association
Via Bergamo, 292 – Cremona
ITALY

Abstract

Data from insemination and calving events of Italian Holstein cows and heifers collected from January 1996 until December 2002 were used to assess the quality of the Italian fertility data set and to estimate the effects of some environmental factors on fertility measures of Italian Friesian. A validation procedure based on gestation length (GL) and information on service sire is presented. According to these procedure, in Italy 75.9 % of the calvings are associated to service or inseminations events, considering a GL within 240 and 300 days and with the sire equal to the SS. The traits considered in the subsequent analysis were: days to first service (DTFS), interval between two successive calving dates (CI), number of services per conception (NSC) and non-return rate at 56 days (NRR56). The effect of month and year of calving (or insemination), of sampling bull code and of the age of the cow were statistically significant as $p<0.0001$ for all the considered traits. Current findings are going to be used in a subsequent variance components and breeding value estimation.

Keywords: dairy cattle, fertility, data validation..

1. Introduction

In the last years selection for functional cows has become an important breeding goal all over the world. Functionality refers to the ability of the cow to be milked in an efficient way, i.e. the same production with lower costs. Fertility, one of the functional characters (of an animal), can inflate the production cost per cow, being one of the main reason for the involuntary culling .

In Italy there are no national statistics available about the incidence of cow's replacement due to fertility. Nevertheless data collected by some Provincial Breeders Associations (APA) confirm the existence of this problem. Previous studies have shown that fertility efficiency can be enhanced above all by means of a better management but that part of the variation in fertility traits has a genetic origin. This means that selection can be practiced for such traits.

The genetic and phenotypic improvement of fertility efficiency relies on the availability and the accuracy of field data collection (i.e., insemination date, calving date, etc). In Italy, most of the inseminations are performed by herdsman technicians. The registration to the National Italian Holstein Herd-Book requires that all the inseminations must be recorded but in some cases only the last insemination event, or no event at all, is recorded. This system of data collection will, by consequence, affect the estimation of fertility measures. Nevertheless, through an appropriate editing, it is possible to improve the reliability of the data set. This editing will be useful to better estimate fertility measures and to supply information about the insemination event data-flow among the Provincial Breeders Associations (APA) and the National Breeders Association (ANAFI).

Bagnato and Oltenacu (1993,1994) and Miglior et al. (1997) estimated both phenotypic and genetic parameters for fertility traits in Italian Holstein cows. Their results however have not been applied for any national genetic evaluation.

Since more data are now available and fertility problems concern the Italian breeders, it was decided to start a new project, aiming at developing management indicators and a national genetic evaluation for fertility.

The objectives of this study were: 1) to analyze the national fertility data set, 2) to propose an editing procedure, 3) and to estimate the effects of some environmental factors on fertility measures in Italian Friesian population.

2. Material and Methods

Data

Data available for this study included all insemination and calving events, both for artificial insemination (AI) and natural service (NS) of Italian Holstein cows and heifers collected from January 1996 until December 2002. Insemination data set consisted of 13,600,000 records with the following information: animal ID, date of service, service sire (SS), herd, service code (AI, NS, DN = donor, RE = recipient) and sampling bull code (NP = national proven bull, FP = foreign proven bull, YP = sampling bull). The calving data set consisted of 5,100,000 records including: animal ID, date of calving, date of birth, sire and herd. The two data sets were collated and sorted by animal and date events (insemination or calving) to compute and check fertility criteria.

Editing and data validation

The number of calving with a complete (2Y), partial (2A, 2N, 2C, 1A, 1Y, 1C, 1N) or missing (3M) presence of service (NS) or insemination (IE) events and the number of herds belonging to these different categories were calculated (Kadarmideen and Coffey, 2001). The herds belonging to category 3M was further classified by different herd size. Categories were assigned by using gestation length (GL) and SS information. If a calving record was not preceded by an insemination record (on the same cow), the code 3M was assigned otherwise a code 1 or 2 was assigned. Assignment of code 1 or 2 relied on the GL and on the SS. When insemination recording is not complete the reliability of the insemination is assessed by using the GL computed from the data and a "true" gestation length of 282 days (± 10) (Kadarmideen and Coffey, 2001; Kadarmideen, 2001). When the computed gestation length is out of the related interval, the NS or the IE are excluded because it is difficult to ascertain whether they are really the first recorded NS or IE. Nevertheless biologically GL ranged from 240 to 272 days or from 292 to 300 days sometimes occurs. In order to recover such records, information on SS might be used. In Italy, when a cow is inseminated more than once, only less then 8 % of successive inseminations are carried out with the same service sire (Marusi, personal communications). This means that if the computed GL is, for instance, 245 days but the sire is equal to the SS, the probability that the IE used to compute GL is not correct, is less then 8 %. According to this theory the code 2Y was assigned to all the calving records with a GL within a range of 240 and 300 days and with the sire equal to the SS. If the cow had more than one calving and the calving was associated to only one IE, the interval between the preceding calving and the unique IE had to be less than 150 days. The code 2N, 2C and 2A were assigned to a calving event with a GL within the range of 272 and 292 days and with the SS not equal to the sire or equal to the sire but belonging to a beef breed or missing, respectively. The remaining cases were assigned to 1Y (sire = SS, GL < 240 days or > 300 days), 1N (sire ≠ SS, GL < 272 or > 292), 1C (sire = SS beef breed, GL < 272 or GL > 292) and 1A (SS missing, GL < 272 or > 292).

Statistical analyses

Using records with the 9 described categories, it was created a data set for investigation of environmental factors. Only the records regarding AI and cows were retained. Records from cows involved in embryo-transfer programs (donor and recipient) were deleted. The analysis considered the following traits: days to first service (DTFS), interval between two successive calving dates (CI), number of services per conception (NSC) and non-return rate at 56 days (NRR56). The choice of traits DTFS and NRR56 was based on recommendation by the GIFT-Workshop on fertility (Groen *et al.*, 1997). Days to first service were restricted to between 21 and 250 days, resulting in about 3,600,000 records. Calving interval and NSC were retained only if the following calving belonged to the category 2Y, 2N, 2C and 2A. This resulted in about 2,200,000 records. Information on NRR56 were deleted if SS was missing. For the estimate of environmental effects, the four traits were analyzed using PROC GLM of SAS®. The model for DTFS included the following fixed effects: the interaction of year by month of calving, the herd-year-season interaction, the age of the cow in classes (> 18 mo and < 260 mo), and the sampling bull code. The model for CI and NSC was the same as previously described with the exclusion of the sampling bull effect. For NRR56 the interaction of year by month of calving was substituted by the interaction of year by month of insemination.

3. Results and Discussion

Data quality and editing

The fertility project developed by ANAFI has three different but correlated objectives: 1) improving the quality of the data collection and of the data-flow among the Provincial Breeders Associations (APA) and the National Breeders Association (ANAFI), 2) supplying the breeders with some management indicators, 3) developing a national genetic evaluation. An accurate analysis and editing of the available insemination and calving events were necessary in order to estimate with the maximum reliabilit some parameters on fertility. Information about the lacks of the data set would also result from that editing allowing corrections and improvements in data collection.

Figure 1 shows the proportion of calvings with a complete (2Y), partial (2A,2N,2C,1A,1Y,1C,1N) or missing (3M) presence of service (NS) or insemination (IE) events. Considering a GL within 240 and 300 days and with the sire equal to the SS, the 75.9 % of the calvings resulted to be associated to service or inseminations events in Italy. It was therefore possible to recover more then 300,000 calvings (10 % of the total) using these criteria. These calvings would otherwise have been lost when using a GL within 272 and 292 days. The analysis of the gestation length distribution (figure 2) does not indicate any unusual peak, supporting the idea that considering GL within 240 and 300 days does not introduce any bias. Nevertheless it should be noted that the proportion of the 9 different categories is not randomly distributed all over the country. This means that some areas have a proportion of 3M that attain 50 %.

The proportion of calvings with missing dates is 10% and the distribution of percentage of herds for different proportion of missing information can be observed in figure 3. Only 4 % of the herds have more then 80 % of missing service dates. Table 1 shows that these herds are also herds with less then 30 animals while large herds had a smaller proportion of missing insemination records. These results were fully expected because small herds, on average, have a less efficient management system due to the geographic localization or to the scarce qualification of the farmer or of the herdsman technician.

About 15 % of the calvings have partial service or insemination event information which therefore can not be used for estimating neither phenotypic nor genetic indexes.

A nested classification like the one used in the present study could be useful not only to validate data for successive genetic analyses but it could act "retroactively", indicating what kind of information are missing and in which areas. As a matter of fact some statistics can be calculated at farm, at regional or at provincial level, allowing a faster identification of the problem. These information can be used by the national Herd-Book to contact the technician of the local Provincial Breeders Association in order to recover information lost. This would improve the existing data-flow among the herd, the Provincial Association (APA) and the National Association (ANAFI), increasing the proportion of calvings associated to insemination records. As a feedback the breeders will have fertility parameters more precise and will also receive a more realistic analysis of the efficiency of their reproductive management.

Environmental effects

Results from the GLM analysis are summarized in table 2. All the fixed effects considered were significant for the four traits as in the previous findings by Bagnato and Oltenacu (1993,1994) and Miglior et al. (1997). Average DTFS was 84.6 ± 33.2 days, similar to what related by Bagnato and Oltenacu (87.05 ± 36.21 days). In Great Britain, Kadarmideen et al. (2000) found that on average it took 84 days after calving for the cows to be ready to be inseminated for first time.

The model for DTFS also considered the effect of the sampling bull code because sometimes when a cow delays her return to heat (*e.g,* milk fever, mastitis, ketosis, …) the breeder prefers to use semen from young bulls, due to its lower cost. The longer interval observed when the insemination is carried out with semen from a young sampling bull, seems to confirm this hypothesis.

The phenotypic trend for DTFS (figure 4) shows an increase of the interval from calving to first service, confirming the breeders' concern about fertility efficiency.

Average CI was 392,6 ± 51.6. The effect of month*year of calving showed that cows calving in April, May and June had higher CI. This was in agreement with results obtained by Bagnato and Oltenacu (1994). This effect reflects both the stress due to seasonal high temperature during the breeding period of the cows and the willingness of of breeders, who delay the insemination to avoid breeding the cow during the warm months**.** Average SPC was 1.96 ± 1.03, a higher value than the one related by Bagnato and Oltenacu (1994), who found a mean SPC of 1.63 ± 1.03. This difference can be due to the actual worsening of the reproductive efficiency and by the different editing on data sets, supporting the idea that the estimate obtained by the previous authors was underestimated.

The quality and editing of the data set have probably also affected the results for NRR56. The mean value for NRR56 was 62.6 % ± 47.1, slightly below the value estimated by Miglior *et al* (1997) as 65.4%. A further difference between the two analysis concerns the number of inseminations considered because Miglior et al. (1997) used only first insemination between 1990 to 1995. Over the last five years the quality of the events recorded have been improved suggesting that probably the estimate by Miglior was underestimated because of the biased data set and of the different editing procedure. As a matter of the fact the R-square of the present model was 16 %, 50 % higher than the one obtained by Miglior *et al* (1997). In conclusion it is believed that the mean value estimated in the present study is more realistic and in agreement with the current situation of Italian herds. The phenotypic trend reported in figure 5 clearly shows that over the last 5 years the proportion of cows subsequently re-bred after 56 days is increasing steadily.

Inseminations carried out during the summer August) or during the successive period (September-October) resulted in a higher proportion of cows being re-bred after 56 days. For all the traits the effect of the age of the cow was included in classes (6) as fixed effect. Only cows older than 18 months were considered, because previous studies (Jansen, 1985) showed that fertility in heifers and cows should be considered and analyzed separately. Anyway a worsening in fertility efficiency was observed in older cows.

4. Conclusion

The first objective of the fertility project developed by ANAFI was to assess the quality of the available information regarding calvings and insemination events and to develop a system to validate the data. Information on service sire and gestation length can be used to salvage or eliminate calving records. Different categories can be used to classify the eliminated records and to calculate some statistics at farm, at provincial or at regional level. These information will be useful to identify where the insemination data-recording is complete or not, allowing by consequence future corrections and improvements. Variance component estimations and breeding value estimation for fertility traits in Italian Holstein Friesian will be based on results of this research.

Acknowledgements

The authors thank Giorgio Civati, herdbook co-ordinator, for his useful comments and suggestions on data editing.

References

Bagnato, A. & Oltenacu, P. 1993. Genetic study of fertility traits and production in different parities in Italian Friesian cattle. *J. Anim Breed. Genet. 110,* 126-134.

Bagnato, A. & Oltenacu, P. 1994. Phenotypic evaluation of fertility traits and their association with production of italian friesian cattle. *J. Dairy Sci. 77,* 874-882.

Groen, A. et al. 1997. EU concerned action "Genetic improvement of functional traits in cattle (GIFT) – Annual Report 1997 Workshop on Genetic Improvement of Functional Traits in Cattle - Fertility and Reproduction - November 23-25, 1997, Grub, Germany.

Jansen, J. 1985 Genetic aspects of fertility in dairy cattle based on analysis of A.I. of data – A review with emphasis on areas for further research. *Livest. Prod. Sci. 12,* 1-12.
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. & Coffey, M. 2001. Quality and validation of insemination data for national genetic evaluations for dairy cow fertility in the United Kingdom. *Interbull Bulletin 27.* Proceedings of the 2001 Interbull Meeting in Budapest, Hungary, August 30-31 2001.
Kadarmideen, H.N., Coffey, M.P., Kossaibati, M.A. & Esslemont, R.J. 2001. Genetic evaluations and parameter estimates for dairy cow fertility in the United Kingdom. Proceedings of the BSAS annual meeting, April 2001, York, UK, page 36.
Miglior, F., Pizzi, F. & Guaita, N. 1997. Effect of environmental factors on Non Return Rate in Italian Holstein-Friesians. *Interbull Bulletin 27.* Proceedings of the 1997 Interbull Meeting in Grub, Germany, November 23-25, 1997.

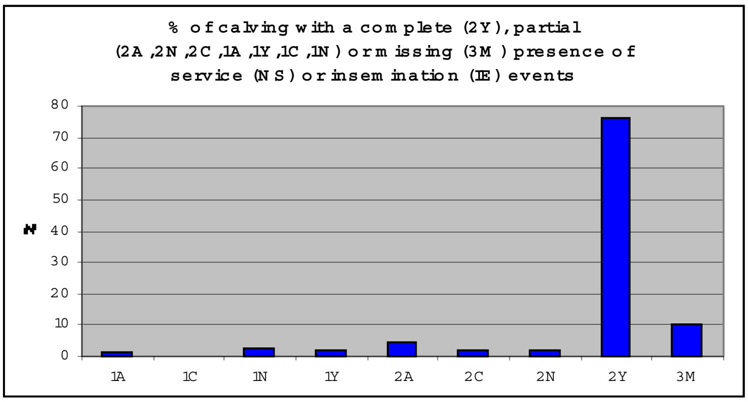

Figure 1. Proportion of calvings with complete, partial or missing information on inseminations events.

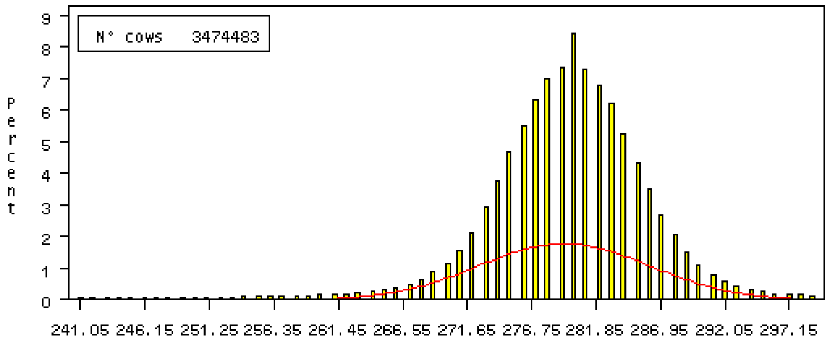

Figure 2. Gestation length distribution.

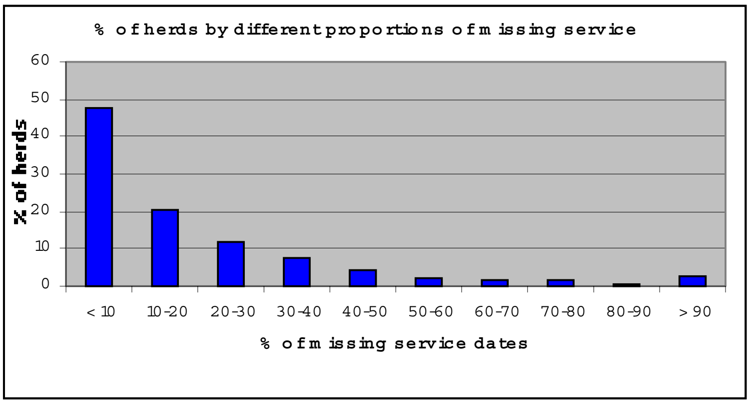

Figure 3. Distribution of herds with different proportion of missing services dates.

Table 1. Relationship between herd size and % of missing service dates (column-wise).

herd size	% of missing service dates (3M)		
	0-30	**30-60**	**60-100**
<10	12.7	32.55	45.46
10-30	22.59	27.68	29.11
30-60	24.57	20.14	15.82
60-90	14.61	8.3	5.59
90-120	9.06	4.66	1.9
120-150	5.62	2.56	1.16
150-180	3.55	1.13	0.32
180-210	2.52	0.72	0.42
>210	4.78	2.26	0.21

Table 2. Results from GLM analyses.

Effect	**trait**			
	DTFS (days)	**CI (days)**	**NSC (days)**	**NRR56 (%)**
herd-year-season	* ($p<0.0001$)	* ($p<0.0001$)	* ($p<0.0001$)	* ($p<0.0001$)
month*year of calving	* ($p<0.0001$)	* ($p<0.0001$)	* ($p<0.0001$)	-
month*year of insemination	-	-	-	* ($p<0.0001$)
age at calving (6 classes)	* ($p<0.0001$)	* ($p<0.0001$)	* ($p<0.0001$)	* ($p<0.0001$)
Sampling code	* ($p<0.0001$)	-	-	* ($p<0.0001$)
Mean	84.6 ± 33.2	392.6 ± 51.6	1.96 ± 1.03	62.6 ± 47.6
R^2	0.25	0.21	0.24	0.16

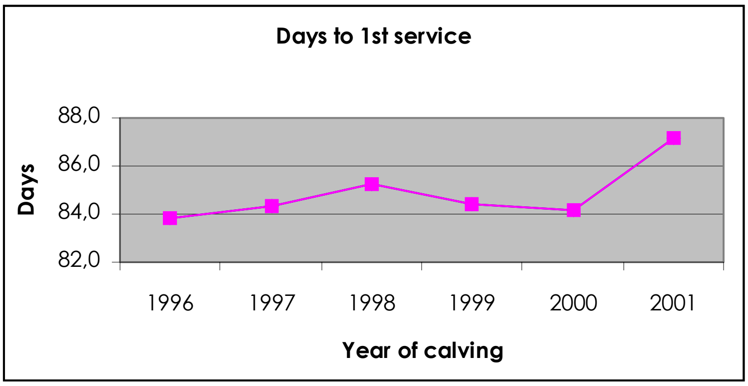

Figure 4. Phenotypic trend for days to first service.

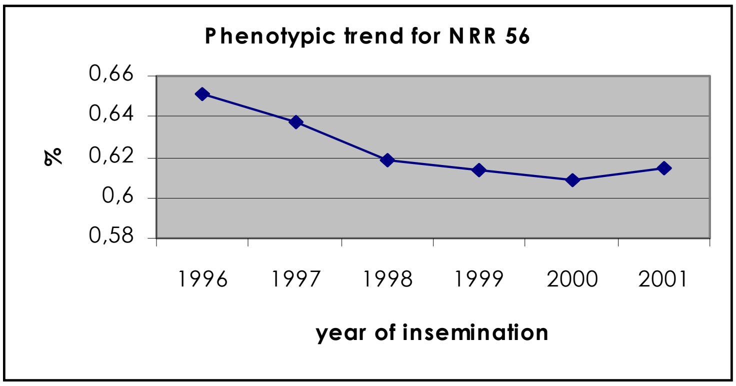

Figure 5. Phenotypic trend for non-return rate at 56 days.