

Data Quality Issues: From Insemination to Genetic Analysis

Gottfried Averdunk¹ and Jörg Dodenhoff,

Bayer. Landesanstalt f. Tierzucht, D-85586 Grub

¹ *Present address: Blombergweg 1, D-85551 Kirchheim, Germany*

E-mail:Gottfried.Averdunk@t-online.de.

1. Introduction

During the last 60 years genetic analysis has gone through a drastic period of evolution: from daughter-dam comparison via contemporary comparison to test-day model. While the first ignored the possibility of environmental effects, the latter tries also to eliminate the effect of the micro-environment at a given test-day.

The analysis of applied methods revealed a large variety in national evaluation systems (Jorjani, 1999; 2000a), which partially are explained by the evolution of systems. The discussions in Zurich and Bled gave us the task to document the differences and stimulate the discussion for more uniformity in national evaluations. Hosein Jorjani did a good job in the documentation of applied methods (2000b), but I am not quite sure, if all the questions were answered in the correct way. Therefore we all should go once more through the resp. country answers to check, if the sheets in IB-bulletin No. 24 contains the right answers.

2. Differences in correcting for environmental effects

2.1 Data quality issues

National evaluation centers rely with regard to insemination, calving and other information upon data from farmers, inseminators and milk recording personal. It is therefore essential, that all people follow given rules, documented by ICAR and other agencies or in an ISO-agreement. It is also essential, that identification systems are clearly documented and straightforward. International bull ID is an example, where it would be helpful, if the ID from the AI-straw is unique in every country and every system. The present practice to renumber

and rename imported bulls might be useful for internal purposes, but cause a lot of problems internationally.

Increasing herd size and reduced care for the individual cow may also cause problems, for instance with the identification of calves in the maternity barn of large units. In most cases blood-typing to clarify parentage is too expensive for commercial herds. In such cases it is better to have an unknown sire rather than a false one. It should be noted, that a higher than normal percentage of wrong sires is reducing the variance between sires and can also be a reason for lower correlations between countries. The publication of spot check results for parentage control should be requested together with the background of the material.

2.2 Data selection

As in most cases, where there is no concrete question, the answers are rather vague regarding limits for inclusion of data. Looking through the country sheets, I found for 23 countries no answer. 8 countries applied age limits for the start of first and/or later lactations, 4 countries have limits on absolute yield or percentages to avoid outliers and 3 countries have limits for age and yield. From my experience, most countries have somewhere in the evaluation process checks for outliers, which are not mentioned or which are conducted somewhere else in the data processing, that the national evaluation unit is not aware of. This fact should definitely be checked within each country and the edits of milk recording organizations should be included.

Preselection of data should be documented.

2.3 Incomplete lactations

In lactation models there are two reasons for using incomplete lactations: to account for selection within first and later lactations and to have an earlier genetic evaluation with running lactations. Table 1 demonstrates large differences in the minimum number of days in milk, required for the inclusion of data. From the mean days, one can extrapolate, that on average three tests are required with a tendency to use less information for culled cows. The column “Not enough information” requires an update, but ***quite a lot of countries do not account for culling during first lactation, which could lead to a severe bias in sire evaluation.***

There are several problems with incomplete lactations:

- a) The modeling of the lactation curve applied to the data (cf. discussion with the test day model), but if it is only applied to culled cows, the factors used have to be adapted regularly. In my opinion these should differ for herd level and time trend. Multiplicative factors may solve some problems, but not all over a time span of 20+ years. Some countries use the last known test day result for projection of whole lactation (e.g. Belgium, Denmark, Italy-HF).
- b) If regularly short lactations are extended, the problem of time trend in factors applied remains with old data.
- c) Quite a lot of countries extend naturally terminated lactations. Here the problem of double counting (short lactation and correction for days open/calving interval) may arise, but I can’t judge, if this happens.
- d) Extension of later lactations seems to be a less important question, since many countries consider this as irrelevant. The effect of selection is less pronounced, since there is no large freedom for voluntary culling in later lactations.

As with other systematic effects, the inclusion of incomplete lactations in the model would be recommended, but there is no simple solution. A mixture of incomplete and culled records could lead to unreal factors and could cause a larger bias. The phenomenon “RIP-drip” is circumscribing the problem, but these are to be tackled both with culled and running records in progress. In Bavaria we thought, dividing the lactation in three 100-day-parts would cause less problems with extension, but now the test-day model with a minimum number of test days (one, two or three?) may be the solution. The question remains:

How to correct best for incomplete records?

2.4 Calving age and comparable expressions

At present most countries apply a pre-correction for calving age with multiplicative factors. In later lactations comparable measures are taken, partially split within lactations. There are some limits for data exclusion to exclude too young or too old animals, which e.g. may have already a second calf. There are not many answers, when the age correction factors were last updated.

Calving age should be included in the model, either fixed or as covariate w.n. lactation. Usually classes are defined w.n. lactations and the distribution in later lactations should be checked, if for instance the calving interval between first and second calf has increased from 390 to 430 days in a time span of 20 years. Despite of out-of-date correction factors, there may be no large bias due to model changes, except in upgrading populations.

2.5 Calving interval and days open

The calving interval (CI) has a pronounced effect in the last third of the gestation, which should be accounted for. This effect can be corrected, if the recording system has information on mating dates, which is commonly the case in a cattle data bank. It can be done either via days open or via the calculated probable calving interval, so both expressions mean the same. The best way to have complete data is, that insemination data are used, but the increased use of “do it yourself”-inseminations and other modes of semen distribution may affect the value and availability of this information. Since the effect of a short and unknown CI is rather pronounced, this gives the possibility to voluntarily inflate cow’s breeding values. At least a correction of pre-calculated CI by calving dates should be applied. The correction of culled records for CI or days open remains an open question.

At present, pre-correction as well as inclusion in the model is applied; in future the inclusion in the model as class effect should be recommended. Since there are trends towards increased CI in nearly all breeds, the distribution should be checked, when data are evaluated over a long time span. There are still some countries with no correction for CI, which may have an effect on the sire evaluation.

The preceding calving interval also has an effect upon the following lactation, which is somewhat confounded with days dry. Some countries apply both corrections, which is not documented in the data analysis. Nevertheless this effect is minor compared with the effect on the running lactation. Correction for days in milk or days dry may have the same effect.

Herdbook rules require that mating dates are recorded before the calf is registered, which in quite a lot of cases is done in a joint data bank. Also insemination data are the base for a sire evaluation for fertility. I just want to draw the attention to the problem of missing information, if the regular registration system is influenced by other ways of semen distribution.

Correction for calving interval / days open should be done in all national evaluations.

2.6 Herd effect

Today in nearly all national systems the herd is considered as a fixed factor in the model, which means that the genetic quality of the herd-mates is considered. There is no doubt, that herd-year-season (HYS) is the best definition of the systematic environment to be corrected for, but herd-sizes differ tremendously between and within countries. This is the reason, that the definition of herd is varying from a real HYS to herdlevel-region-year-season effects. Since the views on minimum number of herd-mates have changed in recent years from 10-15 to 6-8, there might be a way to change to real herds everywhere in the near future. This could be easier, if more lactations would be included +and selection will be accounted for directly in the model. Combining data for several years could be another solution, where year-seasons could be accounted for separately, but the fallacy of ignoring interactions with years is well known.

One major problem is the fact, that groups of cows in larger herds are managed differently. If such management groups² may occur, these are difficult to record. Such management-groups could be cows milked three-times and receiving different concentrate mixtures. BST-application is in some parts of the world considered as a routine management-tool as is feeding according to production, but animals could also be treated differently. In general there is a tendency towards greater variation in management within herds, which is increasing the within herd variance.

A correction for heterogeneous herd variance is applied in some national evaluations, mostly as pre-correction. It must be stated, that the effect of differing management within the herd can not be corrected by such a method; it can only be a crude approximation. It may be that the test day model can give additional information for a grouping within herds, but cluster analysis will always have the problem, where to separate and results in instabil results.

² The expression “Management group” has a different meaning as used in IB.24 and Jorjani’s summary.

Herd-Year-Season (herd-year-testday) is the best way to describe the environmental influence.

Reasons for heterogeneous variance should be monitored.

2.7 Changing recording methods

There is a tendency world-wide towards alternative recording systems, having longer time intervals (4, 6 and 8 weeks) and/or only one sample per day. Also systems with two milk-recordings and one sample are ICAR-approved. ICAR is trying to follow all these developments, but there is a lack of actions in this field (Wilmink, 1998).

In my opinion INTERBULL should try to put more pressure on changes, especially in the development of collecting data from processing-PC's in the milking parlour.

Direct information should be available for evaluation purposes. This would enable the continuous adaptation of results with changing methods, even if these results may differ from the information given to the farmer directly after recording for management purposes. The same should be true in applying optimal methods to calculate lactation yields or other expressions of production in national evaluations.

One unsolved problem are recording procedures for robot milking; the definition of a 24-hour day has problems as well as taking samples for fat and protein during a given period, where some cows may be missing. There are working groups in this field, but I feel, that questions of use in genetic evaluation are of minor importance.

An open question is, if the different recording systems should be included as effect in a national evaluation system. At present the solution might be to have additive corrections as with milking frequency (two vs. three times), which might be applicable for robots. I am not aware if an analysis for other systems is available (e.g. A4, A6, B4 or B6). Certain requirements for herdbook registration or export may cause systematic differences, especially in the genetic background.

The development of recording methods should be followed.

Consequences for animal evaluation systems should be monitored.

3. Conclusions

The existence of Interbull has greatly influenced national evaluation systems towards more uniformity. Nearly all systems are now applying an Animal Model. There was also pressure to larger evaluation centers, especially in Germany. Regarding the definition of systematic effects still exist some differences, which may influence national results and could be the reason for genetic correlations less than unity. Inclusion of some of the systematic effects in the evaluation model could reduce some biases possible, for instance through calving interval and age. Increasing herd size and the inclusion of herd-year-season could also reduce biases. Workshops like this in Verden might contribute to less variable definition of effects, even through cooperation of evaluation centers and exchange of programs.

Definition of traits also differs between different countries and/or breed groups. There is a trend to include more lactations in national evaluations, but more than three lactations do not contribute to precision and the information gain through increased number of contemporaries is also limited. Different trait definitions are also one reason for lower correlations.

Nevertheless the environmental and sociological conditions in the world are variable and this could be the reason for genotype x environment interactions. National evaluation units have also the responsibility to explore the data for extension purposes.

In general the service for farmers, especially from the recording organizations should be improved. If they do not offer this service for horizontal comparisons, other organizations will do it and this may finally reduce the base of our work. Therefore we need to support the evolutionary process, initiated by the founders of Interbull.

References

- Jorjani, H. 1999. Preliminary Report of the Survey on Genetic Evaluation Procedures for Dairy Production Traits in the Interbull/ICAR Member Countries. *Bulletin No. 22, International Bull Evaluation Service*, p.157, Uppsala, Sweden.
- Jorjani, H. 2000a. Report presented at the IB-open meeting in Bled/Slovenia.
- Jorjani, H. 2000b. Opacity of Information. Discussion material for the Workshop in Verden. Uppsala 20.9.2000.
- Wilmink, H. 1998. Invited Paper, ICAR-Conference Roturoa, New-Zealand.