Aspects of validation and data quality based on veterinarian diagnoses

C. Egger-Danner[1], W. Obritzhauser[2], C. Fuerst[1], B. Grassauer[3], K. Zottl[4] , K. Fuchs[5] and B. Fuerst-Waltl[6]

[1]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Strasse 89/19, A-1200 Vienna
[2]University of Veterinary Medicine, Vienna, Veterinaerplatz 1, 1210 Vienna, Austria
[3]Chamber of Veterinarians, Hietzinger Kai 83, 1130 Vienna, Austria
[4] Performance Recording for Lower Austria, Pater Werner Deibl Straße 4, Zwettl, Austria
[5]Austrian Agency for Health and Food Safety (AGES), Data, Statistics & Risk Assessment, Beethovenstrasse 8, 8010 Graz, Austria
[6]University of Natural Resources and Applied Life Sciences, Division of Livestock Sciences, Gregor-Mendel-Strasse 33, A-1180 Vienna, Austria

Abstract

In Austria a health monitoring system for cattle was established between 2006 and 2010. Meanwhile registration of veterinarian diagnoses has been implemented into routine genetic evaluation as well as animal health programs. Presently from about 12.500 farms veterinary diagnostic data subject to documentation by law (Law on the Control of Veterinary Medicinal Products [*Tierarzneimittelkontrollgesetz*]) is standardized, validated and recorded in a central database.

Precondition for efficient use is correct health data. Therefore plausibility checks and data validation are of importance. For use and interpretation the data source and logistics of data recording have to be considered. By working with field data anybody has to be aware that the real incidence of disease will probably not be the same as the observed incidence based on the recorded diagnoses especially in subclinical diseases and disorders. Not all diseases will involve treatment or are treated depending on different herd management strategies and therefore lack recording. Additionally the challenge is to distinguish between farms with low frequencies of disease and incomplete documentation and recording. These limitations make data validation a key issue in health monitoring systems. Nevertheless data from broad health monitoring systems are very valuable for genetic evaluations, for surveillance purposes and herd health management.

The present paper covers aspects of data validation based on Austrian experiences with veterinary diagnoses for genetic evaluation, herd health management and the overall monitoring of the health status. Reasons for incomplete data and measures to improve data quality are presented as well.

Introduction

Health traits are becoming more and more important in breeding, herd management and in the context of prevention of diseases. Good data quality is necessary if any benefits are to be gained from health data. Experiences from different countries showed that it is challenging to establish a system of registration of veterinarian

diagnoses. It is important that the system is adjusted to the existing circumstances and requires a minimum of additional work for farmers and veterinarians. The benefit for the stakeholders involved is essential.

Field data have the advantage of a big quantity of data from a wide range of different farms, but do not always provide complete data. Different validation studies from different countries showed different degrees of completeness (Bartlett *et al.*,2001, Mörk *et al.*,2010, Gaddies *et al.*,2012, Espedvedt *et al.*,2012). Field data do involve high emphasis on monitoring of the registration including plausibility checks and validation. Depending on the use of data different restrictions and criteria have to be applied. The present paper describes the aspects of data validation based on the Austrian circumstances and experience.

Material

The study is based on veterinary diagnoses in dairy cattle in Austria. Between 2006 and 2010 an Austrian wide health monitoring system was established. The overall aims were the development of a genetic evaluation for health traits, the provision of information for herd management und preventive work for farmers and veterinarians as well as benchmarks for monitoring the health status for the Austrian Ministry of Health and the Austrian Animal Health Organizations. A detailed description is found in Egger-Danner *et al.*,2012.

Registration of veterinarian diagnoses

The precondition for the recording of diagnoses is the availability of standardised data. By legal obligation, diagnoses and treatments have to be documented in Austria (Law on the Control of Veterinary Medicinal Products [*Tierarzneimittelkontrollgesetz*]). These documents have to be kept for 5 years by the veterinarians as well as the farmers. However, these data have neither been collected nor stored in a database. A standardized code consisting of 65 diagnoses divided into 10 categories was developed for the project and was published by the Ministry of Health before the start of the project in 2006. This coding system only includes diseases, which can be diagnosed on site by the veterinarians, but currently no laboratory results. A two-digit code for the standardized diagnosis was added to the receipt form for the documentation of medication (Law on the Control of Veterinary Medicinal Products).

Diagnostic data is recorded into the Austrian central cattle database. This has the advantage that validation checks can be done using the information gathered from identification and performance recording. Within the course of each disease the diagnosis is only recorded once (course diagnosis).

The diagnoses are collected by the performance recording organisations or may be sent electronically to the database by the veterinaries. The data are stored within the central cattle database (Rinderdatenverbund, RDV) in Austria.

Participation of farms in health monitoring in Austria

The project was based on voluntariness so that each farmer under performance recording was free to take part. Since 2011 recording of direct health traits is compulsory for the breeding program of most of the breeding organizations in

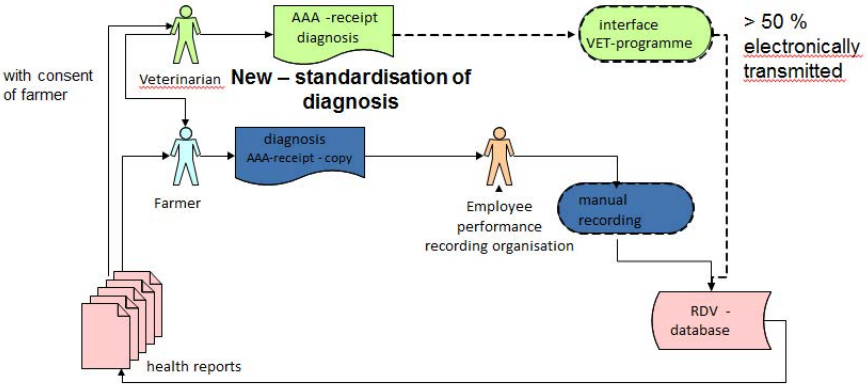

Figure 1. Data logistics for recording of veterinary diagnoses into the central cattle data base in Austria.

Austria. Figure 2 shows a marked increase in participation in 2011 and 2012 due to the decisions of the Cattle Breeding Organisations. The number of farms providing diagnostic data is increasing continuously but is lagging behind the decision of compulsory registration of health data. Big differences exist between federal states. There are some who have achieved an extent of above 90% of veterinary diagnoses, others are substancially lower.

Impact of data recording on frequency of diagnoses

Table 1 shows the incidence rates of the most frequent diagnoses in Austrian cattle depending on the type of data recording. VET means that more than 75% of the diagnoses of a farm are sent to the central cattle data base directly by the veterinarian. The other group includes farms where diagnostic data are recorded either by the employees of the performance recording organisations (PRO) and some farms where there is a mixture of types of recording. Table 1 shows that incidence rates of fertility disorders are higher if calculated with data electronically transmitted. The reason

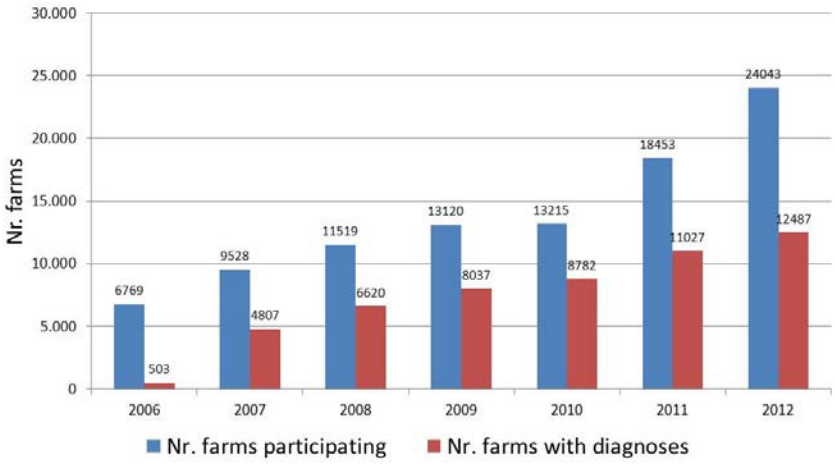

Figure 2. Participation of farms in health monitoring and nr. of farms with diagnoses.

Table 1. Incidence rates based on different data registration methods (more than 75% of diagnostic data submitted electronically (VET), and recording by employee of a performance recording organization or mixture (RPO)) for dairy cows under observation in 2012

Traits	VET	PRO	Diff
Validated dairy cows	36,756	110,597	
Metabolic disorders	**5.65**	**4.08**	**-1.57**
Milkfever	4.36	3.10	-1.26
Ketosis	1.06	0.78	-0.29
Reproductive disorders	**26.14**	**18.54**	**-7.60**
Metritis	5.34	3.36	-1.97
Anoestrus	9.59	6.50	-3.09
Cystic ovaries	9.90	6.57	-3.33
Prolapse of vagina	0.13	0.08	-0.04
Retained placenta	3.34	3.49	0.15
Puerperal disorders	2.05	1.10	-0.95
Udder disorders	**18.16**	**15.73**	**-2.43**
Acute mastitis	12.72	11.37	-1.35
Chronic mastitis	5.84	4.65	-1.19
Hoof and claw disorders	**4.12**	**3.48**	**-0.64**
Panaritium, DD	2.30	1.78	-0.52
Hoof ulcer	0.92	0.96	0.04

might be that if there is no waiting period for drugs used the documentation might not be complete. Due to the fact that veterinarians using electronic devices usually document all treatments irrespective of drug use because of internal organisational reasons. Another reason for the higher incidence rates might be that veterinarians who are using electronic devices are more active in prevention and routine checks.

Methods

The challenge of validation is to differentiate farms with low frequency versus farms with incomplete health data recording. One important precondition before storage of the data in the cattle data base are plausibility checks.

Plausibility checks

The plausibility checks applied before storage in the cattle data base include checks for correct ID of the animal, checks concerning plausibility of date (the animal has to be at the respective farm at the day of the diagnosis) and code of diagnoses. As only first diagnoses are recorded there is also a check concerning the respective time period for first diagnoses per code. More detailed information is found under Austrian Ministry of Health (2010).

Another important plausibility check is by farmer and veterinarian directly by provision of health reports und use within animal health programs. The inclusion of the registered veterinary diagnoses in the health reports facilitates a check for the correctness of the data by farmers and veterinarians. An incorrect documentation and recording of diagnostic data can be recognized and corrected.

Aspects of validation

Only validated data are used for genetic evaluation and the calculation of different benchmarks. Criteria for validation are continuous data recording and the definition of a valid observation period, the incidence rate per farm and the coding of the diagnoses.

Valid observation period

A valid observation period is defined per farm and per animal. In the first step a valid observation period with compulsory recording of diagnoses is calculated taking the date of joining the health monitoring program, the data of the first diagnoses per farm and information concerning start of data transmission into account. As farms do have very small herdsizes it is possible that for a certain period the farm is already providing reliable data, but no case of diagnoses occurred. Continuity of data recording and an eventual ending of health data registration is checked as well. Based on that information for all dairy cows the valid observation period is defined.

Incidence rates of farms

A minimum requirement of 0.1 first diagnoses per cow and year are applied. Only time periods fulfilling these requirements are considered. The incidence rate of first diagnoses validated for genetic evaluation was on average 0.5 diagnoses per cow and year. For electronically transmitted data the incidence rate is 0.7 first diagnoses per cow.

Coding of diagnoses

In Austria diagnoses are standardized by codes for 65 diagnoses. In Egger-Danner *et al.* (2012) the distribution of the different codes of diagnoses per lactation is shown. On average about 40% of the diagnoses are due to fertility disorders and around 35% are due to udder health problems. Veterinarians working with practice management software often use a more detailed list of diagnoses for their own documentation. To link this to the standard codes of diagnoses published by the Austrian Ministry of Health in 2006 a list of synonyms is provided. For genetic evaluation complex traits definitions are used e.g. acute and chronic mastitis are combined to clinical mastitis. Therefore variations in coding of specific diseases between the veterinarians do not cause major problems for genetic evaluation presently, for calculation of incidence rates and monitoring of specific diseases it matters much more. Figures 3 to 7 show the variation in codes of diagnoses by veterinarians restricted to veterinarians with a minimum of 500 diagnoses. The reasons might be that some vets are more specialized in certain diseases or that some veterinarians are working in prevention and e.g. ultrasound examinations

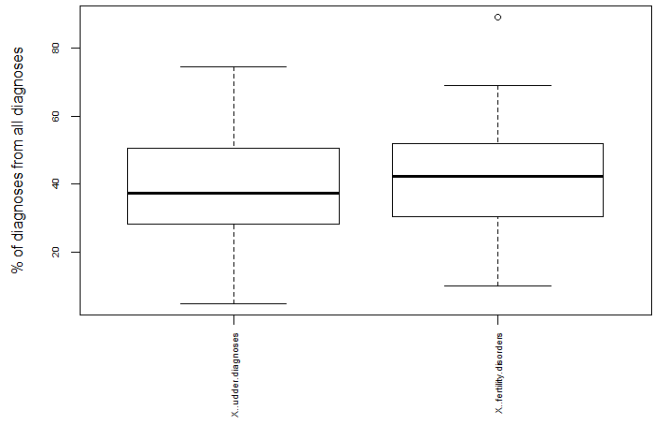

Figure 3. Variation in the percentage of fertility disorders and udder diseases per veterinarian of farms validated for genetic evaluation in 2012.

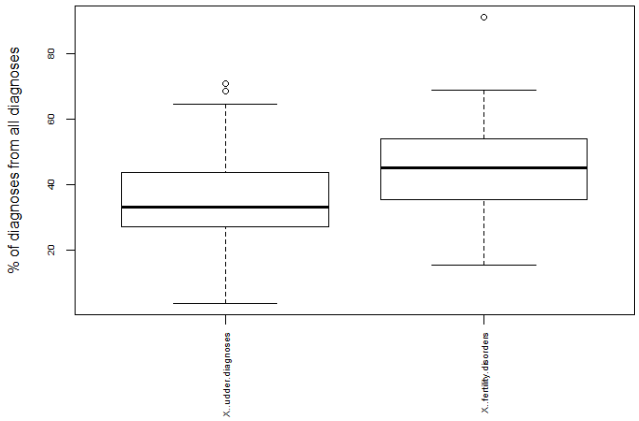

Figure 4. Variation in the percentage of fertility disorders and udder diseases per farm validated for calculation of benchmarks in 2012 (elect. data transmission).

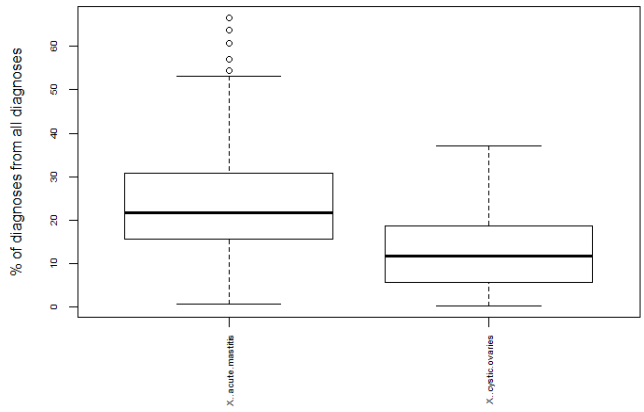

Figure 5. Variation in the percentage of acute mastitis and cystic ovaries per veterinarian of diagnoses from farms validated for genetic evaluation in 2012.

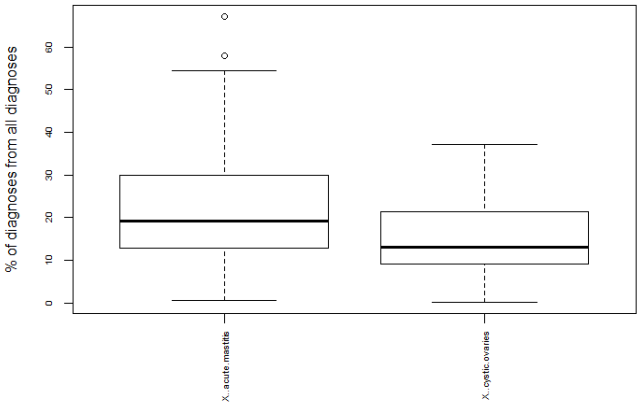

Figure 6. Variation in the percentage of acute mastitis and cystic ovaries per veterinarian of diagnoses from farms validated for calculation of benchmarks in 2012 (elect. data transmission).

are standard. Incorrect coding or mistakes in linkage of codes of diagnoses according the list of synonyms might also be reasons as well as a higher incidence of specific diseases in certain farms or regions by time. Future studies are needed to assess differences in coding in detail.

Other aspects of validation

The completeness of recording of diagnoses could be validated by the usage of additional information. Since 2012 in Austria additional health relevant observations are recorded together with recording of calving ease. This information will be considered in the future to achieve a higher degree of completeness of information.

Use of data

The requirements for validation are depending on the use of data. For herd management only plausibility checks are applied. For genetic evaluation less stringent validation is applied compared to the calculation of benchmarks. The statistical models used in genetic evaluation account for different environmental effects (effect of farm, year *season and type of recording) (Fuerst *et al.*,2010). As a huge amount of data is needed, there is the question about more stringent validation and higher heritabilities or a higher amount of data and lower heritability (Egger-Danner *et al.*,2009). One aspect is that due to low frequencies of certain diagnoses and the requirement of stable breeding values very often a combination of different diagnoses is used for estimation. Koeck *et al*. (2010) shows that the correlation between different fertility disorders are very high and that e.g. early fertility disorders are more stable than single traits like retained placenta, puerperal disorders and metritis. Therefore she recommends the aggregation of different single fertility disorders to early and late fertility disorders. If this is done the conditions concerning coding are not that strict. The conditions for the calculation of benchmarks and the

monitoring of diseases are different. The benchmarks should reflect the real incidence to a high degree. Therefore the requirements on the data quality are higher and more stringent validation criteria are applied. The amount of farms participating is not that important.

Results and discussion

As shown in Egger-Danner *et al.* (2012) about 30% of the farms with diagnoses are excluded for genetic evaluation. The percentage varies between regions. Studies from other countries (e.g. Neuenschwander (2010), Gaddies *et al.* (2012)) also show that high percentages of farms are excluded. If producer-recorded health data are used the fact that a farm is working one year more on e.g. udder health is of higher relevance compared to documentation systems based on the legal documentation requirements connected with the application of drugs. Nevertheless it has to be considered that there are differences in documentation of diseases requiring the application of drugs with waiting period or not.

For the calculation of benchmarks for the Austrian Ministry of Health and the Austrian Animal Health Organizations only data from farms where at least 75% of the diagnoses are transmitted electronically by the veterinarians are used. Further work on validation is in progress.

Three different levels of validation were applied. Restrictions on herdsize were the same for all three levels. Figure 7 shows the distribution of farms according to the percentage of cows with diagnoses in 2012.

No validation/loose (LOOSE). There was no emphasis put on the definition of a valid observation period per farm. The first recorded and last recorded diagnoses per farm were relevant for the calculation of the percentage of animals per farm with certain diagnoses. This group includes 6,111 farms.

Validation for genetic evaluation (EBV). Due to the small herd sizes in Austria the first recorded and last recorded diagnoses per farm might not be the reliable factor for determination of the observation period. Therefore additional information from the veterinarian and the employee of the performance recording organization was used. Incidence rates per farm and year were calculated and only years where the incidence rate was above 0.1 first diagnoses per cow and year were considered. This group consists out of 4,579 farms.

Validation for the calculation of benchmark (VET75)s. Additional to the criteria for genetic evaluation the dataset was restricted to farms where more than 75% of the diagnostic data ware sent electronically by the veterinarian. 1,522 farms are included.

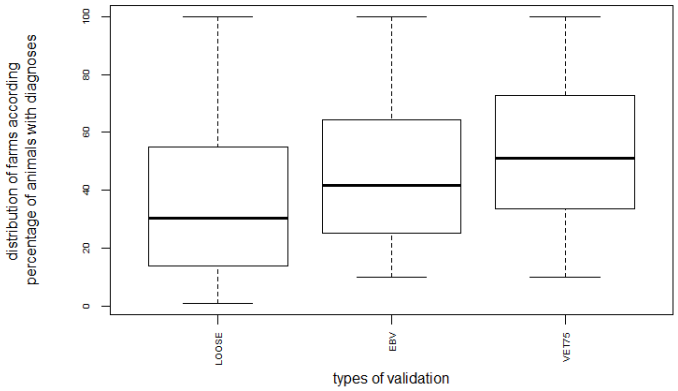

Figure 7. Distribution of farms according to percentage of animals with diagnoses in 2012.

Evaluation of reasons for incomplete data

To improve data quality it is important to know about the reasons of incomplete data. In 2010 a survey based on 600 farms, which were excluded due to validation criteria, was carried out. The survey showed that despite promotion, not all farms or their veterinarians are providing reliable health data. The main reasons for incomplete data mentioned were missing documentation of treatments in general, missing standardization, or that not all available receiptswere provided by the farmer. However, with awareness-building activities and regular information, data quality could be improved.

Measures to improve data quality

Monitoring of recording

Continuous monitoring and evaluation of measures to improve health data quality is done regularly. There is need for further improvement of validation and the provision of regular feedback.

Recording of health data

To improve completeness of data additionally to the registration of veterinarian diagnoses and the possibility for the farmers to record health observations in the data base, recording of health observations around calving by the EPO was started in 2012. Some federal states have already reached a completeness of recording of 95%.

The first priorityare veterinarian diagnoses. Legal documentation requirements are existing. The veterinarian diagnoses are the precondition for the use of synergies and the collaboration of farmer and veterinarian in prevention of diseases.

Continuous information and motivation

Precondition for implementation and maintaining a health monitoring system is the motivation of farmers and veterinarians for documentation and provision of information. The benefit of use of the data is essential. The Austrian experiences show that information and motivation is more challenging than technical aspects.

Further development of benefits of recording

To increase the benefit of health data recording further developments together with partners from the German federal states Bavaria and Baden-Württemberg are in elaboration. These include further work on genetic evaluation and further services for herd management for farmers and veterinarians. An online-platform for veterinarians has been developed. Features for easier documentation of drug use on a voluntary basis are in elaboration..

Real incidence and observed incidence

If diagnostic data based on a nationwide health monitoring system are used it has to be considered that the observed incidence will not necessarily reflect the real incidence of diseases. It depends on the decision of the farmer whether a veterinarian is consulted. Some farmers react earlier others do not consider e.g. a SCC above a certain level a problem Nevertheless field data have the advantage that a huge amount of data is available at low costs. But, high emphasis has to be put on validation.

Conclusions

A good data quality is the precondition for benefit out of health data recording. To motivate farmers and veterinarians to put emphasis on documentation and recording of diagnostic data, additional benefit has to be provided to farmers and veterinarians. Benefits can be breeding values, information for herd management and prevention or easier documentation requirements. Constant monitoring of the registration as well as feedback to the people involved is important to assure a good data quality continuously. Emphasis has to be put on validation especially when a system of health data recording is newly established. The requirements for validation may depend on the use of data. Field data offer the chance that a huge amount of data with limited effort of recording is available. Nevertheless data from broad health monitoring systems are very valuable for genetic evaluations, for surveillance purposes and herd health management.

Acknowledgement

The authors gratefully acknowledge the collaborative work of the project partners in Austria (Federation of Austrian Cattle Breeders, ZuchtData EDV-Dienstleistungen GmbH, University of Natural Resources and Life Sciences Vienna, University of Veterinary Medicine, Chamber of Agriculture, Chamber of Veterinaries, animal health organizations, performance reporting organizations, breeding organizations) to establish a “Health monitoring system in cattle”, and also wish to thank all participating dairy farmers and veterinarians. The project was financed by the Federal Ministry of Agriculture, Forestry, Environment and Water Management (BMLFUW) of Austria, the Ministry of Health and the Federation of Austrian Cattle Breeders (ZAR). Validation and provision of benchmarks are especially supported by the Austrian Ministry of Health. The authors thank the partner organizations from the German federal states Bavaria and Baden-Württemberg for the good cooperation in developing a health monitoring system.

List of references

Austrian Ministry of Health. 2010. Kundmachung des TGD-Programms Gesundheitsmonitoring Rind. *www.bmg.gv.at/cms/home/attachments/9/7/3/CH1141/CMS1271936439807/tgdkundm74200_46-ii-b-10-10gesundheitsprogrammrindprogramm.pdf*). Accessed May 20, 2013.

Bartlett, P.C., J. F. Agger, H. Hous & L.G. Lawson. 2001. Incidence of clinical mastitis in Danish dairy cattle and screeing for non-reporting in a passibely collected national surveillance system. Prev. Vet. Med. 48(2001): 73-83.

Egger-Danner, C., A.Koeck, W.Obritzhauser, C. Fuerst, & B. Fuerst-Waltl, B.2009. Impact of health data quality on breeding efficiency in Austrian Fleckvieh cows. In: Book of Abstracts of the 60th Annu. Meeting of the EAAP, Barcelona, Spain. Wageningen Academic Publishers, Wageningen, the Netherlands. 2009;p. 411.

Egger-Danner, C., A. Willam, C. Fuerst, H. Schwarzenbacher und B. Fuerst-Waltl. 2012. Impact of breeding strategies using genomic information on fitness and health. J. Dairy Sci. 95:4600–4609.

Espetvedt, M.N., C. Wolff, S. Rintakovski, A.Lind & O. Osteras, 2012. Completeness of metabolic diseases recordings in Nordic national databases for dairy cows. Prev. Vet. Med. 105 (2012): 25-37.

Egger-Danner, C., A. Willam, C. Fuerst, H. Schwarzenbacher und B. Fuerst-Waltl. 2012b. Impact of breeding strategies using genomic information on fitness and health. J. Dairy Sci. 95:4600–4609.

Fuerst, C., A. Koeck, C. Egger-Danner, and B. Fuerst-Waltl. 2011. Routine genetic evaluation for direct health traits in Austria and Germany. Interbull Meeting, August, 26-28, 2011, Stavanger, Norway.

Gaddies, K.L. P. , J.B. Cole, J.S. Clay & C. Maltecca, 2012. Incidence validation and relationship analysis of producer-recorded health event dta from on-farm computer systems in the United States. J. Dairy Sci. 95: 5422-5435.

Koeck, A., C. Egger-Danner, C. Fuerst, W. Obritzhauser & B. Fuerst-Waltl. 2010. Genetic analysis of reproductive disorders and their relationship to fertility and milk yield in Austrian Fleckvieh dual purpose cows. J. Dairy Sci. 93: 2185-2194.

Mörk, M. J., C. Wolff, C., A. Lindberg, I. Vagsholm & I, A. Egenvall, 2010. Validation of a national disease recording system for dairy cattle against veterinary practice records. Prev. Vet. Med. 93(2010): 183-192.

Neuenschwander, T.F.O. 2010. Studies on disease resistance based on producer-recorded data in Canadian Holsteins. PhD thesis. University of Guelph, Guelph, Canada.