ABSTRACT

Information on diseases of dairy cows in Finland is entered into the national disease register. Before the data from such types of secondary database are used, the quality of the data needs to be validated. In this study, 7,324 veterinary records for culled cows were compared against records in the national disease register in the period 2002 to 2008. Evaluation of the national disease register data was done by calculating completeness and correctness values. Completeness is the proportion of events that were physically recorded in the secondary database, and correctness is the proportion of correctly recorded events of all recorded events. The use of both correctness and completeness is important when describing the accuracy of secondary data. The completeness and correctness values for the Finnish national dairy disease register were 83 and 92%, respectively. We found that 39% of all unmatched diagnostic events were unreported by the artificial insemination technician who transferred the data from cow cards to the register. Logistic regression models showed that diagnostic events for cows born into the herd had twice the odds of being transferred compared with events for purchased cows. Diagnostic events for reproductive diseases had higher odds of being transferred to the register compared with all other disease groups, and the odds for transfer of the diagnostic event decreased as the age increased. We also found that if the diagnostic event was the last diagnostic event on the cow card, then its odds of being transferred to the disease register were significantly lowered. Although the Finnish national dairy disease register has good completeness and excellent correctness values, different disease groups, age groups, origin of the cow, and timing of the diagnostic event affect how well diagnostic events are transferred to the Finnish national dairy register.

Key words: dairy cow, data validation, disease register, secondary database

INTRODUCTION

Background on Dairy Cow Databases

The Finnish national dairy disease register (NDR) has collected information on diseases of dairy cows since 1982. This computer-based dairy cow register is used for clinical research and management. The main reasons for collecting information are to improve the health of animals at both the individual and herd level, to establish nationwide planning on herd health, and to assist in genetic selection of dairy bulls (Gröhn et al., 1986). In addition the register used in Finland, similar types of national dairy registers are used in Denmark, Norway, and Sweden. National dairy registers offer a vast amount of “ready to use” information for epidemiological research, breeding, and economic decision-making (Gröhn et al., 1990; Valde et al., 2004; Zwald et al., 2004).

In Finland, the NDR is a component of the Finnish National Dairy Recording System (FNDRS), which is based on information collected and provided by the Finnish Agricultural Data Processing Centre. Although participation in FNDRS is voluntary for farmers, 79.5% of the country’s 287,000 dairy cows were registered in the system in 2008 (FABA, 2009). Hence, data on a major proportion of the Finnish dairy cow population are included in the recording system. Apart from collecting information, the FNDRS also returns information to the participating farmers on different health and production parameters through health reports. Participation thus helps farmers in herd health planning and in seeing problems at the herd level rather than focusing on individual sick cows.

Why and How the Databases Are Validated

To avoid incorrect results or interpretations of the research done with secondary data, such as register data, quality should be validated by examining the accuracy
of the diagnostic data (Kelton et al., 1997; Hennessy et al., 2003). This is especially true for disease data, which are used for quality improvement and research (Thrusfield, 2007). The accuracy of secondary database data is measured using completeness and correctness values (Hogan and Wagner, 1997). Completeness is the proportion of events recorded in the secondary database, analogous to the expression of diagnostic sensitivity in epidemiology. Correctness is the proportion of correctly recorded disease events of all recorded events in the secondary database, analogous to positive predictive value of a test result. By using both completeness and correctness values, it is possible to provide an adequate description of the data accuracy in registers (Table 1).

Evaluation of register data is common practice in human medicine and many examples of evaluations have been reported (Kristensen et al., 1996; Hlaing et al., 2006; Tolonen et al., 2007). Evaluations have been less commonly carried out in veterinary medicine. They are stimulating interest, however, especially in the Nordic countries, where databases are commonly used. Examples of validations of secondary data in the field of veterinary epidemiology can be found in Sweden, where horse (Penell et al., 2007) and dog (Egenvall et al., 1998) insurance databases have been validated against veterinary practice records and where the dairy production database has been validated against veterinary records (Mörk et al., 2009; Jansson Mörk et al., 2010). Validation has also been done for calf registers in Norway (Gulliksen et al., 2009) and for the dairy cattle database in Denmark (Bennedsgaard, 2003). Outside the Nordic countries, only a few validation studies can be found in the area of veterinary epidemiology (Müller et al., 1994; Pollari et al., 1996).

The objectives of the study were to evaluate the accuracy of the Finnish NDR and to identify reasons for information loss in data transfer. This was done by comparing diagnostic events recorded on cow health cards with diagnostic events recorded in the NDR and by calculating completeness and correctness values for the NDR.

### Table 1. Completeness was used to assess how many of all recorded events were transferred to the secondary database and correctness was used to evaluate the correctness of the data transfer

<table>
<thead>
<tr>
<th>True health status from cow card</th>
<th>Diseased</th>
<th>Healthy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record present</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>Record absent</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

Completeness = \( a / (a + c) \)

Correctness = \( a / (a + b) \)

1Cell d would represent the truly healthy animals but for diagnostic events it is not possible to determine how many times each individual has been truly healthy.

### MATERIALS AND METHODS

#### Cow Health Card Data

Every cow in Finland has an individual cow health card (Figure 1) that contains notes on diseases, treatments, prophylaxis, and insemination history of the animal. Each time a veterinarian diagnoses a disease, he or she is expected to record the incidence on the card. The veterinarian records the code for the diagnosis as well as information about the treatment: when it was given, the name and dosage of the preparation used, and the length of the treatment. Two hundred individual diagnostic codes are used in Finland. All medications used for dairy cattle need to be prescribed by a veterinarian after a clinical examination, with the exception of some preparations, most commonly intramammary tubes, that the veterinarian can prescribe over the telephone. The herd owner is expected to record the telephone prescriptions on the cow health card. Nonveterinary hoof treatments were not included in our study because they are not recorded on the cow card but are transferred to the NDR through other routes.

#### National Dairy Disease Register

Records are transferred from cow health cards to the NDR by AI technicians and, since 2005, by farmers using computer-based programs (WinSem for AI technician and WinAmmu for farmer; Figure 2). In 2006, most records (97.5%) were transferred by AI technicians (M. Simpanen, Finnish Agricultural Data Processing Centre, Vantaa, Finland; personal communication). Disease reporting via WinAmmu by the farmers has become more popular since 2008 (2,871 disease records in 2006, 27,143 records in 2009; M. Simpanen, Finnish Agricultural Data Processing Centre, Finland; personal communication). The information transferred to the database includes herd ID, cow ID [unique European Union (EU) identification number], ear tag number (number used by the farmer, not unique to a cow), treatment date, diagnostic code, treatment occasion,
and veterinary ID. Each cow in the herd has both an EU identification number (cow ID) and an ear tag number. After data insertion into WinSem by an AI technician, the information is transferred to the NDR and the personal ID number of the AI technician is added to the register data. The digital transfer program makes multiple checks on the transferred data, as no fields can be left blank. Additional checks on date, duplicate information, and cow identifications are done during later stages of the data transfer. If errors are found, the diagnostic event is reported to the AI technician for correction. For farmers, data insertion is limited to that pertaining to their own herd. The data from WinAmmu are also checked automatically for duplicates before they are added to the NDR. All diagnostic codes are included in a dropdown menu so that it is not possible to enter nonexistent diagnostic codes.

**Sampling and Data Management**

To evaluate the data transfer from cow health cards to the NDR, we collected the cards for culled cows and compared the disease information with the information in the NDR. For this study, herd sampling was done within the context of another study, in which 167 randomly sampled dairy herd owners agreed to record all the diagnostic events for the herd [S. Rintakoski, M. Espedtvet (Norwegian School of Veterinary Science, Oslo, Norway), C. Wolff (Swedish University of Agricultural Sciences, Uppsala, Sweden), A. Lind (University of Copenhagen, Copenhagen, Denmark), J. Kyyrō (Finnish Food Safety Authority Evira, Helsinki, Finland), J. Taponen, O. A. T. Peltoniemi, and A.-M. K. Virtala, unpublished data]. For the present study, farmers were asked to submit all cow health cards for cows culled between 2002 and 2008. Forty-seven herd owners (28.1%) submitted the requested cow health cards, termed “cow cards,” for photocopying.

Information from the cow cards was entered as cow card data (CCD) in a digital form (MS Excel, Microsoft Corp., Redmond, WA), including herd ID, cow ID, cow ear tag number, diagnostic code, treatment date, order number of the treatment (i.e., follow-up treatments), veterinary ID, time of culling, and county of herd. The variable “county” included 4 distinct geographic regions; South, West, North, and East. The treatment notes were in the form of free text on the cow cards and were used only if a treatment code was missing on the cow card. Each diagnostic event on a cow card was regarded as an individual observation, with
the exception of duplicate treatments. The duplicate cow card treatment was defined as a second treatment for the same cow with the same diagnostic code on the same date (follow-up treatment). All duplicates were removed from the analyses to make the study comparable with those of other Nordic countries that do not record follow-up treatments for the same day.

The raw CCD originated from 2,327 cow cards, of which 508 did not detail any diagnostic events, providing information only about inseminations. The raw CCD contained 7,966 diagnostic events. After removing duplicate information (n = 98), all nonveterinary treatments of hoof diseases (n = 158), and observations with missing information on cow ID (n = 33) or diagnostic code and treatment date (n = 353), the final CCD contained 7,324 diagnostic events from 1,786 cows.

All disease information for the 2,327 cows sampled was extracted from the NDR according to the unique cow ID value. In total, extracted data included 9,401 diagnostic event records from 1,701 cows, and after removing duplicate information (n = 1,688) and information on nonveterinary treatments of hoof diseases (n = 1,077), 6,636 diagnostic events from 9,401 events were included in the NDR data. The large number of duplicates in the NDR resulted from follow-up treatments.

To check the validity of the data input from the photocopied cow cards into the CCD for transcription errors due to researchers, the CCD was first matched with the NDR data by cow ID, diagnostic code, and treatment date. All diagnostic events that had exactly matching information were considered correctly entered. Thereafter, a random sample of 200 diagnostic events was taken from the 1,442 events that did not match with the NDR. We found a relatively low incidence (0.6%) of transcription errors attributable to researchers for data entered into the CCD.

**Statistical Analyses**

**Completeness and Correctness.** Completeness of the NDR data was calculated by comparing the diagnostic events in the CCD and NDR data sets. The CCD and NDR data were matched using the SQL procedure of SAS (version 9.2, SAS Institute Inc., Cary, NC). Completeness was calculated as the number of diagnostic events in the NDR with a matching observation in the CCD, divided by all events in the CCD (Table 1). Correctness was calculated as the number of diagnostic events in the NDR with a matching observation in the CCD, divided by all events in the NDR (Table 1). Diagnostic event information consisted of herd ID, cow ID, diagnostic code, and treatment date. Completeness and correctness were evaluated at 3 different agreement levels depending on the similarity of the information in the CCD and the NDR data sets (Table 2). Exact binomial confidence intervals for both completeness and correctness were calculated (Casella and Berger, 1990).

**Logistic Regressions.** Multilevel logistic regression was performed to determine the factors affecting the completeness of the NDR. An F-test was used for fixed effects and a likelihood ratio test for random effects. The model was based on maximum likelihood estimation (PROC GLIMMIX of SAS). The dependent variable corresponded with whether the diagnostic event was found in the NDR or not. The predictor variables were geographic region (south, west, north, and east), disease group (all diagnostic codes were divided into 5 groups; reproductive diseases, udder diseases, metabolic diseases, lameness diseases, and other diseases), age of the animal at the time of treatment (<2 yr, 2 yr, 3 yr, 4 yr, and ≥5 yr), and whether the cow was born into the herd or purchased. Herd was included as a random variable to account for intra-class correlation.

To identify the reasons why some of the diagnostic events were not transferred despite being present on the cow cards, a second regression model was created using the same procedure as previously (PROC GLIMMIX of SAS). The dependent variable was the AI technician’s “transferred” entry on the cow card. Predictor variables were geographic region, age of the animal, whether the cow was born into the herd or purchased, and timing of the diagnostic event on the cow card. Timing of the...
diagnostic event on the cow card indicates whether the diagnostic event was close to culling or not. The variable was binomially classified to event closest to culling (or the only event on the card) and event other than closest to culling. Cow cards did not have the culling date, and thus we defined the variable only by its physical location on the cow card. Herd was also included as a random variable to account for intra-class clustering.

Explanatory variables in both models were tested for collinearity using the CORR procedure and chi-squared test. Backward stepwise regression was used for both regressions. Two-way interactions were tested between all variables, but none were statistically significant and therefore they were not included in the final model in either of the analyses. For model validations, a goodness of fit test was used where values for deviance were divided by the degrees of freedom. Estimates were 0.97 and 0.85, indicating a good model fit (close to 1) for both models. A causal diagram was used to test for confounding in the models. Because age of the cow was associated with both the outcome and timing of the diagnostic event on the cow card, it was considered to be a confounder and therefore necessary to keep in the model. No other confounding effects were detected. For any pairwise testing, P-values were adjusted with the MULTITEST procedure of SAS for false discovery rate (Benjamini and Hochberg, 1995) to control type I errors. Intra-class correlation ($\rho$) for random effect of the herds was calculated according to Dohoo et al. (2003), using the equation $\rho = \sigma^2_{\text{herd}} / \left( \sigma^2_{\text{herd}} + \pi^2 / 3 \right)$, where $\sigma^2_{\text{herd}}$ is the estimated variance of the herd.

RESULTS

Completeness and Correctness

Completeness and correctness values at each agreement level are presented in Table 3. Out of 496 diagnostic events not marked as transferred by the AI technician, 480 events were completely missing from the NDR. This was the single largest reason for data loss and contributed 27.5, 39.1, and 48.4% of all not matched CCD diagnostic events at exact, minor, and major agreement levels, respectively (Table 3).

At the minor disagreement level, 3 main reasons for transfer errors were found. First, the treatment date was associated with a discrepancy of ±7 d in 102 diagnostic events, caused by poor handwriting and transcription errors made by the AI technician. Second, for 138 diagnostic events, a small discrepancy in treatment code was found that was caused by veterinarians’ poor handwriting or, in some cases, because the AI technician had changed the diagnostic code. Finally, in 281 diagnostic events, the treatment code was missing in the CCD but available in the NDR. At the major disagreement level, 2 main reasons causing transfer error were identified. First, in 183 diagnostic events, the treatment date had a discrepancy of >7 but ≤365 d. Second, 55 additional diagnostic events were found when the treatment date was left out and diagnostic events were matched only by cow ID and diagnostic code. Errors at the major disagreement level were the result of poor handwriting or transcription errors. When matches found at the minor and major disagreement levels were added to the exact agreement level, completeness and correctness values both increased (Table 3).

Logistic Regressions Explaining Completeness of the NDR

Completeness of diagnostic events in the NDR data (Table 4) depended on whether the cow was purchased or born into the herd ($P < 0.001$), disease group ($P < 0.001$), and age of the cow ($P = 0.002$). Geographic region ($P = 0.136$) did not have a significant effect on the completeness of the NDR. Herd as a random variable had a significant effect on the model and was kept in the model to account for the effect of intra-class correlation. The proportion of model variance at the herd level was 12%.

---

Table 3. Completeness and correctness figures for the Finnish national dairy disease register (NDR) data compared with cow card information received from 47 herd owners concerning all culled cows between 2002 and 2008 (n = 7,324)\(^1\)

<table>
<thead>
<tr>
<th>Agreement level</th>
<th>Matched diagnostic events</th>
<th>Not matched CCD diagnostic events</th>
<th>Completeness (% (95% CI))</th>
<th>Not matched NDR diagnostic events</th>
<th>Correctness (% (95% CI))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact agreement</td>
<td>5,574</td>
<td>1,750</td>
<td>76 (75–77)</td>
<td>1,062</td>
<td>84 (83–85)</td>
</tr>
<tr>
<td>Minor disagreement</td>
<td>6,095</td>
<td>1,229</td>
<td>83 (82–84)</td>
<td>541</td>
<td>92 (91–93)</td>
</tr>
<tr>
<td>Major disagreement</td>
<td>6,333</td>
<td>991</td>
<td>86 (85–87)</td>
<td>303</td>
<td>96 (95–97)</td>
</tr>
</tbody>
</table>

\(^1\)Number of diagnostic events found in both the NDR (n = 6,636) data and on the cow card data (CCD) at different agreement levels and number of observations that were only found from either of the databases.
Nontransferred Diagnostic Events

Timing of the diagnostic event on the cow card (P < 0.001), age of the cow (P < 0.001), disease group (P < 0.001), and geographic region (P = 0.024) had significant effects on whether the AI technician marked the diagnostic event as transferred to the NDR (Table 5). Whether the cow was purchased or born into the herd was of no statistical significance (P = 0.238). Herd as a random effect was significant, and proportion of model variance at herd level was 27%.

DISCUSSION

Completeness and Correctness

Our results on data completeness were very similar to those reported for Swedish and Danish studies. In Sweden, Jansson Mörk et al. (2010) reported completeness of 75% compared with our completeness of 76% at exact agreement level. In Denmark, completeness was 80 to 85% (Bennedsgaard, 2003), very close to our completeness at the minor disagreement level (82%). In Sweden and Denmark, researchers used copies of veterinary receipts and notes to validate disease registers, whereas we used copies of cow health cards. The advantage of the Finnish cow cards was their easy availability. Jansson Mörk et al. (2010) spent a lot of time and effort in collecting the data from individual receipts. Cow cards, in contrast, were stored in binders and all treatments of the individual cow were arranged in chronological order. Another difference was the role of the AI technician in data transfer in Finland. In Sweden and Denmark, veterinarians transferred the diagnostic events to the NDR, a process requiring one less intermediary step and possibly less data loss caused by poor handwriting or transcription errors in those countries.

Important Reasons Affecting Data Loss

The greatest reason affecting completeness was the 480 diagnostic events that were not transferred to the NDR by the AI technician. Regression analyses for these nontransferred diagnostic events showed that the primary reason for this data loss was that the timing of a diagnostic event occurred closest to removal of the cow due to death or culling. After a cow is removed from the herd, the cow card is archived and the AI technician is not able to transfer the diagnostic event. These nontransferred events, however, affect only completeness values because the data are entirely missing from the NDR. The AI technicians in eastern Finland had 3 times the odds of transferring a diagnostic event to the NDR compared with AI technicians in southern Finland. The instructions for the AI technicians...
to transfer disease records have not been uniform in all regions and data collection has been emphasized in eastern Finland (H. Rautala, University of Helsinki, Finland; personal communication), explaining these regional differences.

Other important reasons affecting completeness were whether the cow was born into the herd or was purchased, disease of the animal, and age of the cow. Diagnostic events for cows born into the herd had twice the odds of being transferred to the NDR compared with events for purchased cows. This was a result of the use of 2 types of cow ID: ear tag and EU identification number. The AI technicians cannot use the EU identification number because of restrictions in the computer program, and therefore have to use the ear tag number. For cows born into the herd, the ear tag number remains the same, but when a cow is purchased, its ear tag may be re-numbered. The old ear tag number, however, may stay on the card, which can cause the AI technicians to confuse the cow card, leading to an incorrect NDR record.

In both regressions, the younger the cow the better the information transfer to the NDR. Reproductive diseases had the highest odds of being transferred to the NDR, followed by udder diseases and other diseases. Metabolic diseases and lameness had the lowest odds of being transferred to the NDR. Differences in age and disease groups are explained by cow removal and long data transfer intervals in Finland. The AI technicians collect disease information only when they are on the farm to inseminate a cow. Although AI technicians should, in theory, transfer all disease events from the cow card, the work is poorly monitored and the time gap between previous and subsequent data transfer can be long. If a cow is treated but removed from the herd (due to culling or death) before the next insemination, the disease information cannot be transferred by the AI technician because the cow card is archived. Older cows are more likely to be culled, explaining the lower odds as the age increases. Reproductive diseases occur shortly before or after insemination, meaning that an AI technician is very likely to visit the cow soon after the reproductive diagnostic event. Treating cows for metabolic diseases or lameness is more likely to lead to removal of the cow, and diagnostic events are therefore less likely to be transferred to the NDR. Our results suggest a systematic error in the NDR: diseases leading to culling or death are underrepresented in the NDR.
Reasons for Data Loss on Different Agreement Levels

At the minor disagreement level, the discrepancy errors in treatment date were caused by poor handwriting on the cow cards and transcription errors of the AI technician (e.g., the previous diagnostic code was chosen instead of the correct code). For those cases in which the diagnostic code had only a small discrepancy, we found 2 explanations. The first was a transcription error due to poor handwriting on the cow card, but the error did not change the actual diagnosis. For example, code 300 on a cow card was marked as 303 in the NDR; both code for mastitis. Second, we found that some of the AI technicians had used their own judgment in disease coding. On those occasions, the AI technicians changed the diagnostic codes, although the information should simply have been copied from the cow card. Such adjustments to the diagnostic codes by the AI technicians caused discrepancies that affected completeness and correctness calculations at the exact agreement level. One reason for these changes was that the AI technician interpreted the code differently if the veterinary ID code was missing from the card. In those events, some AI technicians changed the code to match the “Farmer’s treatment by phone prescription medicine” code. However, this did not explain all the changes. We also found events in which the diagnostic code information was missing from the cow cards but was present in the NDR data. In most of these situations, the AI technicians were able to code the diagnosis correctly, presumably by asking the farmer for the treatment of the cow or by using the veterinarians’ handwritten information on the medication. The majority of these events were mastitis diagnoses (87%), for which the medicines used were presumably widely known among the AI technicians and therefore easy to recognize. At the major disagreement level, the discrepancy was caused by transcription errors for treatment date.

Errors causing mismatches between the CCD and NDR data can be classified into 2 groups: systematic errors and random errors (Sørensen et al., 1996). According to Arts et al. (2002), causes of systematic errors include programming errors, unclear definitions for data items, or violation of the data transfer protocol. Random errors can be caused by imprecise data transcription. We found both types of errors affecting the completeness and correctness evaluation for the NDR in our study. Misclassifications of diagnostic events in the NDR will also give rise to conservative estimates of disease incidence with missing data. On the other hand, incidence estimates will be too liberal for diseases where a diagnostic event is wrongly recorded. Misclassifications may therefore result in incorrect deductions when data from the NDR are linked to production data from the national recording systems (Jansson Mörk et al., 2010). The differences for date and diagnostic codes found at the minor disagreement level were so small that we believe they are unlikely to induce incorrect results for incidence calculations or production parameter calculations. Hence, we suggest that the matches at a minor disagreement level should be regarded as correct and exact when evaluating the total functionality of the NDR.

This study measured data transfer from established records. Correctness is therefore viewed as the proportion of all recorded diagnostic events that have been transferred correctly from the cow cards, not as the proportion of all diagnostic events that concern the disease in question. No database is 100% perfect. Some information is inevitably lost, as not all diseases are treated by veterinarians. This induces discrepancy between the true clinical status of the cow and the information recorded on the cow cards. Lost information or underreporting of diagnostic events would have a substantial influence on annual statistics if a large proportion of treatments were not reported. Underreporting, however, is very difficult to resolve. We do not have any evidence that underreporting or discrepancy between true clinical status and recorded information in general would affect the results of our study. Underreporting may be a problem, however, for phone-prescribed treatments. In a pilot study, Østerås et al. (2003) showed that the incidence of mastitis in Finland was very low compared with that in other Nordic countries. We know that Finland is the only country among the Nordic countries that uses phone prescriptions and it is likely that, although not studied, many of these phone prescriptions are not recorded on the cow card.

Study Validity

We received only a 28% response to requests for participation in this study, which could cause bias toward a liberal completeness estimate if farmers were to have a large effect on disease recording and data transfer. In Finland, veterinarians make most of the cow card records and AI technicians do most of the data transfer. Therefore, we do not suspect this type of bias in our study. We have no reason to expect any major loss of data or
other selection bias because we studied diagnostic events from all removed cows from sampled herds with full treatment history from birth to death.

The Finnish dairy disease recording system is currently under major revision and is moving into the digital era. In the future, veterinarians will transfer all disease records on the day of diagnosis. The new system will minimize most of the causes for data loss found in this study by removing various data transfer steps that exist in the current system. The new digital system will not be error free, however, and will generate new types of error sources more characteristic of digital transfer. Therefore, the new system will also need to be evaluated in terms of completeness and correctness. Possible differences in incidence rates may be obtained with different recording systems, caused by differences between the recording systems. Knowledge of the completeness and correctness values of the systems will enable useful comparisons to be made.

CONCLUSIONS

The data in the Finnish national dairy disease register, which have been used since 1982, capture more than 80% of all events reported on the cow cards. The quality of the data can be considered excellent because >90% of the events are correct diagnostic events. However, the reasons for data loss found in this study (e.g., a large percentage of nontransferred diagnostic events, differences in transfer accuracy between disease groups and between cows born into the herd and purchased cows, increased data loss with a cow’s increasing age, and misclassification of diagnostic events or treatment date) all suggest that improvements are still needed. Improvements are being made in Finland through design and implementation of a new digital recording system.

ACKNOWLEDGMENTS

Access to data used in this study was provided by the Finnish Agricultural Data Processing Centre Ltd. (Vantaa, Finland). The study was financially supported by grants from the Nordic Joint Committee for Agricultural research (NKJ, Oslo, Norway) via the Ministry of Agriculture and Forestry in Helsinki, Finland. The Orion-Farmos Research Foundation (Helsinki, Finland) and the Finnish Veterinary Foundation (Helsinki, Finland) also provided funding. Participating farmers played a crucial role in this research and the authors are grateful for their contribution. The authors thank Marjo Simpanen (Finnish Agricultural Data Processing Centre Ltd., Vantaa, Finland), Maarit Vehmas (University of Helsinki, Helsinki, Finland), Frøydis Hardeng (Norwegian School of Veterinary Science, Oslo, Norway) and Jonna Kyyrö (Finnish Food Safety Authority Evira, Helsinki, Finland), as well as the members of the Finnish steering committee of the project, for their input. Thanks also to all the other researchers in the DAHREVA (dairy herd disease register evaluation) research group; Mari Espetvedt (Norwegian School of Veterinary Science, Oslo, Norway), Cecilia Wolff (Swedish University of Agricultural Sciences, Uppsala, Sweden), Ann-Kristina Lind (University of Copenhagen, Copenhagen, Denmark), Agnete Egenvall (Swedish University of Agricultural Sciences, Uppsala, Sweden), Hans Høye (University of Copenhagen, Copenhagen, Denmark), Ulf Emanuelsson (Swedish University of Agricultural Sciences, Uppsala, Sweden), Peter Thomsen (Aarhus University, Tjele, Denmark), Olav Reksen (Norwegian School of Veterinary Science, Oslo, Norway), Ann Lindberg (National Veterinary Institute, Uppsala, Sweden), and Olav Østerås (TINE Norwegian Dairies, Ås, Norway) for valuable comments.

REFERENCES