J. Dairy Sci. 96:2271–2282
http://dx.doi.org/10.3168/jds.2012-6143

Data quality in the Norwegian dairy herd recording system: Agreement between the national database and disease recording on farm

M. N. Espetvedt,*[1] O. Reksen,* S. Rintakoski,† and O. Østerås*‡
*Norwegian School of Veterinary Science, Department of Production Animal Clinical Sciences, PO Box 8146 Dep., NO-0033 Oslo, Norway
†University of Helsinki, Department of Veterinary Biosciences, PO Box 66, FI-00014 Helsinki, Finland
‡TINE Extension Services, TINE SA, PO Box 58, 1431 Ås, Norway

ABSTRACT

The majority of herds in Norway participate in the national dairy herd recording system. For disease events, this involves transferring information registered on farm, using individual cow health cards (CHC), to the central cattle database (CCD). Before using data from such a database, validation with an aim of describing data quality should be performed, but is rarely done. In this study, diagnostic events from CHC and CCD from 74 dairy herds were compared. Events in 2008 from female cattle with minimum age of 1 yr were included (n = 1,738). Discrepancies between the 2 data sources and assessment of data quality were evaluated using agreement between events on CHC and in CCD, calculating completeness and correctness for the CCD, and using a multivariable regression model for agreement (1/0). The agreement evaluation described the concordance between the 2 data sources, whereas the calculations of completeness and correctness depended on a reference data source assumed to be more reliable. Completeness of the CCD was defined as the proportion of diagnostic events on the CHC that was recorded therein. Correctness was defined as the proportion of the CCD events that was also recorded on the CHC, and with the same date and diagnostic code. The agreement was up to 87.5%, the majority of disagreement being caused by unreported events on the CHC (between 10 and 12% of all events). Completeness of the CCD was regarded as high, between 0.87 and 0.88, and correctness excellent, between 0.97 and 0.98. The multivariable regression model found 4 factors that increased the odds for diagnostic events being in agreement between CHC and CCD. These were the events occurring during the 305-d lactation period; the herd size being 75 cows or less; the event occurring during the spring, summer, or winter rather than autumn; and lastly, the diagnostic code for the disease event being preprinted on the CHC, involving a simple check mark as opposed to writing a 3-digit code. The model found a high degree of clustering within herd. In conclusion, disease data in the Norwegian national database for dairy cows are valid to use for epidemiologic research, having in particular an excellent correctness, but it is of concern that at least 10% of data are missing. The proportion of unreported data should be taken into consideration whenever data from this database are used. Reasons for discrepancies found are important to be aware of in any work aiming to improve data transfer from farm to central databases.
Key words: database validation, completeness and correctness, data transfer, dairy cow

INTRODUCTION

Several countries have established production animal recording systems (International Committee for Animal Recording, 2012). Dairy cattle recording systems have become an essential part of intensive dairy farming today, recording varying amounts of production and health data, on both the individual and herd level. Electronic recording may occur purely on farm in own herd information management systems; for example, InterHerd (http://www.interagri.org/E/IH/interherd.htm) in several European countries (NMR, 2010); DairyCOMP 305 (http://www.vas.com/dairycomp.jsp), PCDART (http://www.dairyone.com/FarmServices/PCDart/default.htm), and DHI-Plus (http://www.dhiprovo.com/solutions/dhiplus.asp) in the United States (Wenz and Giebel, 2012); and DairyWIN (http://www.dairywin.co.nz/) developed in New Zealand (EpiCentre, 2005). Alternatively, recording is in central databases on regional or even national levels as in the Nordic countries (Olsson et al., 2001), The Royal Dutch Cattle Syndicate (NRS, 1993), the Australian Dairy Herd Improvement Scheme, and the National DHIA in the United States (ADHIS, 2012; National DHIA, 2012). The recording of production, reproduction, and health data can be of great benefit to the farmer, guiding daily decisions and aids in the detec-

Received September 10, 2012.
Accepted January 2, 2013.
[1]Corresponding author: mari.espetvedt@nvh.no

tion of change in health and productivity over time (for example, the emergence of a subclinical mastitis problem). The added advantage with centralized systems is that they can provide comparison among herds and are accessible, useful data sources for epidemiologic research (Bartlett et al., 1986) and breeding programs (Heringstad et al., 2004).

The Norwegian Dairy Herd Recording System (**NDHRS**) is a centralized nationwide system that has included the option of disease recording since 1975. In 2011, 98.3% of Norwegian dairy herds, a total of 8,935 herds, participated with disease recording [TINE SA (The Norwegian Dairy Association), Ås, Norway; TINE Rådgiving, 2012]. The NDHRS records production parameters such as milk yield and composition, carcass quality at slaughter, and calving and reproductive information, together with disease and treatment data. Recording of disease data on farm is mandatory, but participation in the NDHRS with transfer of disease data to the central database is voluntary. Historically, the high participation in the NDHRS is related to the long standing cooperative organization of dairy farming, the national dairy cooperative starting in the late 1800s. Today, usefulness of access to several types of data from the system is well recognized by herd managers and veterinarians. Since 1996, health reports have been sent periodically to participating dairy herds. In 2005, Internet-based information systems for health data, including disease events, reproductive events, and milk sample results became available for the farmer and selected herd health advisors (Østerås et al., 2007). Descriptions of variables in the central database are available at http://www.kkvet.no (NDHRS, 2012). Numerous examples of published research fully or partly based on data available in this cattle health database can be found (Hardeng and Edge, 2001; Sogstad et al., 2006; Whist and Østerås, 2007; Garmo et al., 2008; Andersen et al., 2012). Despite extensive use of data in the NDHRS for several years, only recently has validation work describing data quality of recorded disease events in cows been carried out. This was in a comparative validation study of the Danish, Finnish, Norwegian, and Swedish recoding systems (Espetvedt et al., 2012; Lind et al., 2012; Wolff et al., 2012). Before this, only calf health data were validated in Norway (Gulliksen et al., 2009). The study with the 4 Nordic countries was concerned with finding the proportion of farmer-observed clinical disease events that was recorded in national databases for specific diseases, but did not separate out data loss only occurring during the transfer from written records on farm to the central database. Examples of other animal disease database validation studies outside of Norway are limited, and the majority can be found in other Nordic countries: some for cattle (Bartlett et al., 2001; Mörk et al., 2009, 2010; Rintakoski et al., 2012), horses (Penell et al., 2007, 2009), and dogs and cats (Egenvall et al., 1998; Nødtvedt et al., 2006) Whereas Pollari et al. (1996) studied data quality for a small animal hospital database at the Ontario Veterinary College, Mulder et al. (1994) assessed the quality and usefulness for research of practice-generated computerized medical records for dairy cows in Canada, and Salman et al. (1988) validated disease diagnoses reported to the National Animal Health Monitoring System for a Colorado beef feedlot.

Investigating and describing data quality in animal disease databases have been carried out in different ways and using varying terminology, including evaluating agreement between 2 data sources for a dog and cat insurance database (Egenvall et al., 1998), describing completeness and correctness for the Swedish and Finnish cattle databases (Mörk et al., 2010; Rintakoski et al., 2012), and calculating sensitivity and positive predictive value (comparable to completeness and correctness, respectively) for an equine insurance database (Penell et al., 2007). Furthermore, in Sweden, the investigation of the spatial relationship between recorded mastitis incidence and SCC (Wolff et al., 2011), and investigation of risk factors for data loss in the Finnish system using logistic regression (Rintakoski et al., 2012) have been done. Specific guidelines do not exist for such work, but certain frameworks have been suggested within the context of human databases, where more validation work has been carried out (Sørensen et al., 1996; Hogan and Wagner, 1997; Arts et al., 2002). However, the most appropriate way may vary according to the design of the registration system and the aim of the validation work (Abate et al., 1998). In a systematic review of papers assessing electronic human patient records, it was concluded that there were a lack of standardized methods for assessment of quality of data and that this made it difficult to compare results between studies (Thiru et al., 2003). When carrying out database validation in the human medical field, systematic review papers have suggested that the terms completeness and correctness are the most appropriate for describing data quality (Hogan and Wagner, 1997; Jordan et al., 2004). These terms are also appropriate for validating animal disease databases. Completeness is the proportion of observed or actual disease events that is present in the database. Correctness is the proportion of registered events that is a correct representation of the actual disease event (i.e., conforming to the definition for the diagnostic event registered).

The majority of diagnostic events that are registered in the NDHRS are veterinary attended, although it is possible for the farmer to independently register events.