Available online at www.sciencedirect.com
ScienceDirect
Preventive Veterinary Medicine 82 (2007) 236–251
PREVENTIVE VETERINARY MEDICINE
www.elsevier.com/locate/prevetmed

Validation of computerized Swedish horse insurance data against veterinary clinical records

Johanna C. Penell [a,*], Agneta Egenvall [a], Brenda N. Bonnett [b], John Pringle [a]

[a] *Department of Clinical Sciences, Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, P.O. Box 7054, SE-750 07 Uppsala, Sweden*
[b] *Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1G 2W1*

Received 1 January 2007; received in revised form 19 May 2007; accepted 24 May 2007

Abstract

The aim was to evaluate the agreement between computerized insurance data in a large Swedish horse insurance database and the information in the corresponding clinical records (CR). A random sample of 400 veterinary care and 140 life claims was included. Information on name of the horse, breed and gender, year of birth, specific diagnosis and system diagnosis (e.g. joints, digestive and skeletal) was compared between sources. The concordance for demographic variables was categorized as agreement, disagreement or data missing. For diagnostic information, the categories were agreement, minor disagreement and major disagreement and for system information agreement or disagreement. There were missing values for demographic information in the CR, varying from 2% for name to 16% for breed. The overall agreement for demographic information was >94% (disregarding missing data), 92% for system and 84% for specific diagnosis. For veterinary care and life claims, the observed agreement for diagnosis was 85 and 83%, minor disagreement 6 and 5%, and major disagreement 9 and 12%, respectively.

Using the CR data as gold standard, for the systems evaluated (joints, digestive, skeletal, skin and hooves), sensitivity varied between 62% (skin) and 89% (digestive) whereas the specificity was >96% for all systems. The positive predictive values ranged from 86% (skin) to 97% (digestive). Logistic regression analysis was used to examine factors associated with agreement for diagnosis. Analyses were performed separately for veterinary care and life claims. Factors examined were type of visit (clinic/field), treating veterinarian/clinic (categorized as district veterinarians, private practitioners, small clinics, medium clinics and, for the clinics with ≥20 claims, the specific clinics), computerized or manual CR, processing clerk, whether the CR was included in the paper file, if the claim was rejected or reimbursed, system diagnosis and if a immediate settlement (in analysis for veterinary care claims) or death certificate (in analysis for life

* Corresponding author at: Division of Ruminant Medicine and Veterinary Epidemiology, P.O. Box 7054, SE-750 07 Uppsala, Sweden. Tel.: +46 18 671323; fax: +46 18 673545.
E-mail address: Johanna.Penell@kv.slu.se (J.C. Penell).

0167-5877/$ – see front matter
doi:10.1016/j.prevetmed.2007.05.020

claims) was included in the paper file. For veterinary care claims, in the logistic regression model type of visit was significantly associated with agreement, with clinic visits generating better agreement than field visits.

Keywords: Validation; Computerized; Horse; Insurance data; Clinical records; Diagnosis; Sweden; Correctness; Completeness

1. Introduction

Computerized information in medical databases offers the potential for use in epidemiological research (Lawrenson et al., 1999; Miller et al., 2002). However, the quality of the data should be addressed by, for example, examining the accuracy of diagnostic and demographic data (Hennessy et al., 2003). In human medicine, evaluations of computerized medical information have usually involved validating a database against either a patient survey or a paper record (Roos et al., 1991; Pilpel et al., 1993) but validation of electronic patient records based solely on the contents of the clinical database has also been performed, for example, by comparing morbidity data to recognised diagnostic standards to confirm diagnoses and identify further cases (Hassey et al., 2001). Evaluations of computerized veterinary medical records are less common. One study evaluated the quality of data at a Canadian Veterinary Teaching Hospital by comparing information in the computerized record to the paper medical files (Pollari et al., 1996a).

The use of insurance data for epidemiological studies in human medicine has been examined (Bright et al., 1989; Hennessy et al., 2003). Problems addressed included, for example, the uneven validity and completeness of the diagnoses appearing on claims and the lack of straightforward identification of individuals in the database. Use of insurance databases in equine veterinary medicine has been rather limited (Barkley, 1983; Clausen et al., 1990; Seidensticker, 1999; Leblond et al., 2000; Egenvall et al., 2005; Penell et al., 2005; Egenvall et al., 2006a,b).

Sweden has a tradition of medical insurance for animals with the majority of the nation's horses currently insured. A claims database for horses is maintained at a Swedish Insurance Company (Agria Insurance, P.O. Box 70306, SE-107 23 Stockholm, Sweden, www.agria.se) that insures, in addition to companion and farm animals, a large number of Swedish horses. These insurance data have been used to describe morbidity and mortality rates for insured horses in Sweden during 1997–2000 (Egenvall et al., 2005, 2006a; Penell et al., 2005). However, having demonstrated the feasibility of accessing these data, assessment of data quality is needed to support further research. The overview statistics were guided by an earlier evaluation of the insurance company's dog database where the agreement was considered reasonable when comparing information in clinical records to the computerized information (Egenvall et al., 1998). The objectives of this study were to determine the degree of agreement between the computerized insurance data (CID) and clinical records (CR) for demographic and diagnostic information in a sample of claims for veterinary care and life insurance and to investigate factors related to agreement. Also, for five of the most commonly affected systems the aim was to describe the accuracy (sensitivity, specificity, positive predictive value) and the consistency (κ) when comparing the two data sources.