

Comparison of Diagnostic Codes in a Clinical-Research Database and an Administrative Database

J.P. Kerby, K. Keshavjee, A.M. Holbrook, Centre for Evaluation of Medicines,
St Joseph's Hospital and McMaster University, Hamilton, Canada

Purpose

To design a research database using records from a structured Electronic Medical Record (EMR) system in 18 Family Practice clinics in Hamilton, Ontario, Canada.

Background

The COMPETE Study (Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness) is an electronic medical record project based in a primary care setting. Physicians were recruited over a period of 12 months from March 1998 to March 1999.

The Administrative database consists of patient demographic information, OHIP billing diagnosis and procedure code. All data are stripped of patient identifying information before being downloaded to the research centre. The Clinical database consists of data extracted from the EMR as described below in the Methods section. A research database consisting of the Administrative and Clinical databases was created to support the research team. Interpretation of the data has proven challenging due to factors such as variability in physician recording, ambiguities posed by data entered as text and discrepancies between administrative and EMR data.

Methods

18 family medicine clinics, comprising 32 Family Physicians, were recruited, fitted for, installed and trained in the use of an EMR (Purkinje Inc.'s DCI). Physicians chart a mean of 76% of their patients into the EMR on a particular day (min-max 0-100%). Data are collected using Remote Access by modem dial-in. These data include all patients seen by the physician since they joined the study. The two data sets were compared for matching diagnoses for visits on the same day and for all diagnoses recorded for a patient. Diagnostic

categories compared include diabetes, musculoskeletal conditions and cardiovascular conditions.

Results

The Administrative Data set (RDB Version 1.1, data collected from Mar 25, 1998 to Jan 7, 2000) has 201,871 diagnoses on 201,871 invoices. This represents 45,617 patients. The Clinical Research Database has 122,449 diagnoses on 114,984 clinical encounter notes. This represents 26,511 patients or 58% of the patients seen in the clinics participating in the project. The percentage of males and females in the two data sets are comparable at 42% and 58%, respectively. The clinical database has 21% of patients older than 65 years of age versus 15% in the administrative database. For a diagnosis with a single diagnostic code, e.g., diabetes, the match was much higher at 51% (95% CI = 10.4) of notes having a matching code in the billing database. Matches for MSK diagnoses were 32% (95% CI = 6.2) and 40% (95% CI = 7.3) for cardiovascular diagnoses. The matches improve substantially if all the diagnoses recorded for a patient in the administrative database are compared to all the diagnoses recorded for the same patient in the clinical database. Matches for patients with diabetes, MSK and CVS diagnoses are 62%, 35% and 57%, respectively. The percentage agreement between the administrative database and the clinical research database for all diagnoses is quite low, with an overall match of 20% (95% CI = 3.1).

Discussion

Although administrative data may capture larger numbers of patient encounters, we suggest that administrative data may be less satisfactory for clinical research than previously realized. EMR clinical data provide more diagnostic detail and may be more suitable for clinical research.

