Original research

The completeness of electronic medical record data for patients with Type 2 Diabetes in primary care and its implications for computer modelling of predicted clinical outcomes

Michael Staff [a,*], Christopher Roberts [b], Lyn March [c]

[a] Public Health Unit, Northern Sydney Local Health District, Sydney, Australia
[b] Northern Clinical School, Hornsby Hospital, University of Sydney, Australia
[c] Rheumatology and Musculoskeletal Medicine, Northern Clinical School, Royal North Shore Hospital, University of Sydney, Australia

ABSTRACT

Aim: To describe the completeness of routinely collected primary care data that could be used by computer models to predict clinical outcomes among patients with Type 2 Diabetes (T2D).

Methods: Data on blood pressure, weight, total cholesterol, HDL-cholesterol and glycated haemoglobin levels for regular patients were electronically extracted from the medical record software of 12 primary care practices in Australia for the period 2000–2012. The data was analysed for temporal trends and for associations between patient characteristics and completeness. General practitioners were surveyed to identify barriers to recording data and strategies to improve its completeness.

Results: Over the study period data completeness improved up to around 80% complete although the recording of weight remained poorer at 55%. T2D patients with Ischaemic Heart Disease were more likely to have their blood pressure recorded (OR 1.6, $p = 0.02$). Practitioners reported not experiencing any major barriers to using their computer medical record system but did agree with some suggested strategies to improve record completeness.

Conclusion: The completeness of routinely collected data suitable for input into computerised predictive models is improving although other dimensions of data quality need to be addressed.

* Corresponding author at: Northern Sydney Local Health District Public Health Unit, c/- Hornsby Hospital, Palmerston Rd, Hornsby 2077, Australia. Tel.: +61 2 9477 9400; fax: +61 2 9482 1650.
E-mail address: Michael.Staff@health.nsw.gov.au (M. Staff).

1. Introduction

The prevalence of and health burden associated with Type 2 Diabetes Mellitus (T2D) is increasing and its management poses great challenges for primary care practitioners [1] with large differences in metabolic control existing between countries [2]. Over recent decades many countries including Australia have attempted to address these by introducing primary care programs [3] aimed at improving the quality and coordination of routine care for people with T2D [4,5]. However, a significant gap still exists between recommended guidelines and primary care practice in countries such as Australia [6,7].

Recently there has been interest in using computer models to predict clinical outcomes for patients with T2D [8,9]. It has been suggested that multi-factorial interventions may result in greater reductions in adverse outcomes [10] and models such as the UKPDS Outcomes Model [11] could be used to provide further evidence to support this approach. This model's ability to predict life expectancies using individual level data has also been proposed as a novel approach to evaluating primary care initiatives aimed at improving the management of T2D [12].

The data, which models require, should be routinely collected within primary care [13] and include patient history, clinical observations, and laboratory results. Data extraction tools that access computerised medical record systems are increasingly being used for quality assurance activities within Australian primary care with recent attention being given to the validity of the information obtained [14,15]. One study has concluded that current tools may be unreliable [14] and recent experience has also raised concerns about the completeness of these data outside of formal research trials [12,16,17].

One of the major challenges potentially restricting the usefulness of extracted data is the extent to which primary care practitioners enter data through free text into a general notes field rather than through structured coded fields. Previous research has identified that barriers to physicians entering clinical data as coded entries include time constraints during consultations [18], issues with software interfaces and codes [18,19], and the under appreciation of the usefulness of coded data as a quality indicator [20]. Although attempts have been made to extract clinical observation data from free text fields [21,22] such approaches are likely to have inherent limitations related to variations in users' text recording practices. A recent literature review looking at routinely collected electronic clinical data and chronic disease management identified completeness, accuracy, correctness and timeliness as major dimensions that need to be considered when assessing data quality for both research and patient care purposes [23].

This study examines electronic data extracted from Australian primary care computerised medical records being used routinely outside of a formal research trial environment. The objectives of the study were first to describe the completeness of clinical data relevant for the management of T2D in primary care; second to determine whether patient characteristics were associated with completeness of data and; third to determine general practitioners' self-reported barriers to recording data and strategies that might overcome these barriers.

2. Methods

2.1. Medical record system

Data were electronically extracted from twelve primary care practices during practice visits by one of the researchers. These practices had previously participated in a study looking at evaluating the impact of a primary care program on the management of T2D in the Australian primary care setting [12]. All practices used the medical software Best Practice™ as their computer medical record system. This software has the ability to record current and past visit notes, past medical conditions, clinical observations, and laboratory results.

Practitioners can record a patient's 'reason for visit' or 'past medical condition' by selecting from drop down menus or entering free text. Entries created by using the drop down menus are stored as coded data. There are fields within patients' records designed to be used to record clinical observations such as blood pressure, and if used, facilitate data retrieval and analysis. These readings can also be recorded as free text in the general notes area but data recorded in this manner cannot be readily retrieved. Laboratory data can be automatically imported into the system if provided in a standard format (Health Level Seven (HL7) [24]) and coded to facilitate its storage, analysis and presentation. All laboratory data are required to be reviewed by the general practitioner prior to incorporation into a patient's medical record.

2.2. Identification of T2D patients and their clinical data

Clinical data were extracted from coded text fields and designated clinical observation fields with limited free text searching of the 'past medical condition' field. No free text searching of the general notes section of patient records was undertaken.

Patients with T2D were identified through searching of the 'past medical condition' and 'reason for visit' fields [17] and supplemented by a search for glycated haemoglobin (HbA1c) levels above 6.5% (48 mmol/mol) [25]. Predictive computer models can set patient eligibility criteria based upon age at T2D diagnosis [11]. Consequently only regular patients who were aged between 45 and 64 years at the time of their diagnosis were included in the study analysis. Patients were only included in the study for the period after they were diagnosed with T2D. Being a regular patient was defined as having presented to an individual practice on at least two occasions per year. This was assessed each calendar year allowing the status to alternate from year to year dependent upon the number of visits made to a practice during the relevant year. The clinical data assessed included blood pressure, weight, HbA1c, total cholesterol and HDL-cholesterol. These five data parameters were chosen as they are often required to be included as inputs into models that predict clinically relevant outcomes (e.g. life expectancy) [11]. The presence of at least one reading in a