MAYO CLINIC

Automated Diabetes Case Identification Using Electronic Health Record Data at a Tertiary Care Facility

Sudhi G. Upadhyaya, MS; Dennis H. Murphree Jr, PhD; Che G. Ngufor, PhD; Alison M. Knight, MS; Daniel J. Cronk, MS; Robert R. Cima, MD; Timothy B. Curry, MD, PhD; Jyotishman Pathak, PhD; Rickey E. Carter, PhD; and Daryl J. Kor, MD, MSc

Abstract

Objective: To develop and validate a phenotyping algorithm for the identification of patients with type 1 and type 2 diabetes mellitus (DM) preoperatively using routinely available clinical data from electronic health records.

Patients and Methods: We used first-order logic rules (if-then-else rules) to imply the presence or absence of DM types 1 and 2. The "if" clause of each rule is a conjunction of logical **and, or** predicates that provides evidence toward or against the presence of DM. The rule includes *International Classification of Diseases, Ninth Revision, Clinical Modification* diagnostic codes, outpatient prescription information, laboratory values, and positive annotation of DM in patients' clinical notes. This study was conducted from March 2, 2015, through February 10, 2016. The performance of our rule-based approach and similar approaches proposed by other institutions was evaluated with a reference standard created by an expert reviewer and implemented for routine clinical care at an academic medical center.

Results: A total of 4208 surgical patients (mean age, 52 years; males, 48%) were analyzed to develop the phenotyping algorithm. Expert review identified 685 patients (16.28% of the full cohort) as having DM. Our proposed method identified 684 patients (16.25%) as having DM. The algorithm performed well—99.70% sensitivity, 99.97% specificity—and compared favorably with previous approaches.

Conclusion: Among patients undergoing surgery, determination of DM can be made with high accuracy using simple, computationally efficient rules. Knowledge of patients' DM status before surgery may alter physicians' care plan and reduce postsurgical complications. Nevertheless, future efforts are necessary to determine the effect of first-order logic rules on clinical processes and patient outcomes.

 ■ Mayo Clin Proc Inn Qual Out 2017;1(1):100-110

From the Department of Health Sciences Research (S.G.U., D.H.M., C.G.N, J.P., and R.E.C.), Department of Anesthesiology and Perioperative Medicine (A.M.K., T.B.C., D.J.K.), Department of Physiology and Biomedical Engineering (T.B.C.), Department of Information Technology (D.J.C.), Division of Colon and Rectal Surgery (R.R.C.),

Affiliations continued at the end of this article.

Diabetes mellitus (DM) is a metabolic disease resulting from abnormal insulin secretion or insulin resistance, or both. Whether DM is type 1 (type 1 diabetes mellitus [T1DM]) or 2 (type 2 diabetes mellitus [T2DM]), it is the leading cause of microvascular complications, myocardial infarction, stroke, congestive heart failure, and end-stage renal disease that often results in premature death or disability.[1] Patients with DM also undergo surgical procedures at a higher rate than do patients without DM.[2]

Metabolic stresses experienced in the perioperative encounter can alter glycemic homeostasis of patients with DM, which then may result in higher rates of perioperative hyperglycemia, postoperative sepsis, impaired wound healing, and ischemia.[3] In addition, previous studies have shown that postoperative complications such as stroke, urinary tract infection, postoperative hemorrhage, transfusions, wound infection, and even death are more common among patients with uncontrolled DM.[4,5] Thus, it is imperative to identify patients who have DM before

their care episode and initiate appropriate care protocols to optimize their glycemic levels.[3,6,7]

In light of the importance of identifying patients with DM before they receive health care, our objective was to develop a highly sensitive and specific automated electronic phenotyping algorithm for such identification. The algorithm would use available data in the electronic health record (EHR) to drive coordinated clinical processes of care. For instance, results from this algorithm would be used by physicians to facilitate 1 or more of the following steps: (1) schedule a regular glucose check, (2) initiate a protocol-driven DM care pathway, (3) initiate a consult with the diabetes service, or (4) ensure insulin availability in the preoperative care environment. Thus, in addition to high classification performance and computational efficiency, the interpretability of the classification process by care providers was prioritized during the development of this DM classifier.

METHODS

In 2012, planning was initiated for the development of an automated DM identification algorithm at Mayo Clinic, a tertiary care academic medical center with 110 operating rooms and more than 1200 beds in Rochester, Minnesota. Mayo Clinic's Integrated EHR complies with all the definitions of a comprehensive EHR and at its core provides physicians with the ability to capture patients' demographic characteristics and medical history, which includes documenting patients' medical issues in a free-text format in patients' notes.[8] The institution performs approximately 4000 surgical procedures a month, with an estimated 15% to 17% of surgical patients having DM. The surgical practice of Mayo Clinic began a project to standardize the care of patients with DM across the surgical episode. During this process, the practice determined that it would be more efficient, effective, and reliable if the DM diagnosis was captured at a central location in the EHR before the patient arrived on the day of surgery. Clinicians agreed that the data elements needed to determine whether a patient has DM have frequently existed in the EHR, but the lack of clear identification of a DM diagnosis in the EHR was a barrier to implementation.

Data Sources

Previous studies have suggested that the most accessible and reliable sources for DM cohort identification were (1) *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) diagnoses codes (Supplemental Appendix 1, available online at http://www.mcpiqojournal.org); (2) laboratory test values; and (3) patient medication data, in varying combinations and thresholds.[9,10]

In addition to these 3 primary domains, several studies have indicated that natural language processing (NLP) can considerably increase accuracy and precision during identification of health issues documented in clinical notes.[11,12] Therefore, in the patient notes section of Mayo Clinic's Integrated EHR, a keyword-based search technique (see Supplemental Table, available online at http://www.mcpiqojournal.org) was used to detect all cases where a provider makes a positive annotation of DM in a patient's clinical notes, to further improve the accuracy and precision of our algorithm. A typical patient notes section may contain descriptions such as the following:

- The patient denies any chronic medical conditions; however, review of outside health records reveals that at times in the past 5 years, he was medicated for T1DM
- Hyperparathyroidism, C-section, T2DM
- T2DM, last HbA_{1c} unknown IFG/early DM2, 131, January 17, 2014
- Pretransplant history dictated into past medical/surgical history:
  1. Renal diagnosis: ADPKD
  2. Recurrence risk? No
  3. Previous transplants? No
  4. Dialysis pretransplant? No
- T2DM
- History of diabetes, history of hypertension in the past, otherwise normal
- T2DM, diet-controlled
- hx gest DM

Metformin, a common medication used not only in DM management but also for other indications (eg, polycystic ovarian syndrome), was included only in the presence of an abnormal laboratory value measurement. The abnormal values were fasting plasma glucose of greater than or equal to 126 mg/dL, random