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.

© 2017 Mayo Foundation for Medical Education and Research. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
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 HbA1c 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
plasma glucose of greater than or equal to 200 mg/dL, or the presence of a glycosylated hemoglobin test result (hemoglobin A1c [HbA1c]).

As patients treated with metformin for indications of DM may have HbA1c values that are within normal limits, a decision was made by our endocrinology colleagues to classify patients with any value of HbA1c and a prescription for metformin in their EHR as having DM.

Algorithm Design
We used first-order logic (if-then-else) rules to imply the presence or absence of T1DM and T2DM. This rule-based classification model had a series of logical statements using logical AND operators and logical OR operators. Beyond modeling actual human decision making with logical statements, the approach based on the if-then-else rule strikes a balance between accuracy and interpretability for general classification problems. An overview of the classification model is provided in Figure 1. With this method, our rule classified a patient as having DM when the patient’s EHR contained the following:

1. One or more outpatient diabetes-related ICD-9-CM diagnosis codes
   or
2. At least 1 hypoglycemic medication reported during outpatient medication reconciliation
   or
3. A combination of metformin use and a laboratory value exceeding the maximum threshold value
   or
4. Any positive annotations of DM in the patient’s clinical notes

Existing Phenotype Algorithms
Similar rule-based phenotyping algorithms for the identification of patients with DM have been developed previously. Examples include Electronic Medical Records and Genomics Network (eMERGE), Surveillance Prevention and Management of Diabetes Mellitus (SUPREME-DM), Chronic Condition Data Warehouse (CCW), Durham Diabetes Coalition (DDC), Hemoglobin A1c of New York City (HbA1c of NYC), and Harvard Medical School. Although many aspects of these various algorithms appear similar, considerable differences exist in how these case-definition concepts are operationalized and in the distinctive underlying population characteristics, for which these case-definition strategies were developed. For example, the phenotyping algorithm developed by eMERGE focused on identifying only patients with T2DM for large-scale genomic research. SUPREME-DM is a rule-based method developed by a consortium of 11 integrated health systems to identify patients with T1DM and T2DM for research purposes, but this project considered only insured patients. The CCW algorithm was designed to identify T1DM and T2DM among older patients on the basis of Medicare beneficiary data. The DDC designed its DM phenotyping algorithm to identify T2DM cases specifically among racial/ethnic minority groups and low-income patients in Durham County, North Carolina, for public health reasons. HbA1c of NYC is a case-definition method designed for mandatory patient HbA1c reporting to a public health authority for disease surveillance and tracking. The DM phenotyping algorithm designed by members of Harvard Medical School was targeted toward identifying patients with T1DM and T2DM and was incorporated into an EHR-based public health surveillance and reporting system. Hence, these various DM case-definition methods differ widely on the basis of their stated objectives, and uncertainty remains about their generalizability across various study populations, particularly the cohort of interest in the present investigation—heterogeneous surgical patients.

In light of such limitations, as well as our noted desire to use the results from the developed algorithm for DM determination before the surgical encounter, we were motivated to develop a rule set that could identify accurately the presence of DM in the preoperative environment. This work was performed in collaboration with our endocrinology colleagues to identify a patient who has either T1DM or T2DM at 1 or 2 days before the scheduled surgery date.

Algorithm Validation
Several authors have adopted various approaches to validate results from phenotyping algorithms. For example, Spratt et al
adopted a stratified sampling approach based on Begg and Greenes method, whereas Newton et al adopted an iterative approach where information obtained at each step is used to fine-tune and improve the final phenotype algorithm method. In our study, we followed the iterative approach suggested by Newton et al and evaluated the performance of the proposed algorithm for patients scheduled for surgery during August 2014. Data related to these patients were extracted from the existing legacy Mayo Clinic Unified Data Platform that stores structured, unstructured, and other patient care-related data elements from various sources that support research and quality improvements.

The ascertainment of true DM incidence as the reference standard was created by comparing results from the proposed method with those of the existing process, which is a manual review of the EHR for a DM diagnosis by bedside nurses. When the bedside nurse identifies DM, its documentation is noted in the preoperative evaluation patient flow sheet. This nurse preoperative evaluation patient flow sheet document serves as a surgery intake tool, nursing communication tool, and assessment tool. The nurse documents whether the patient has DM or a history of DM in the flow sheet. This is based on patient response, surgical listing information, the patient’s current medication regimen, and a clinical notes review. The nurse synthesizes this information and documents in the EHR if the nurse confirms that the patient has DM. To develop this reference standard, we first considered concordant cases with agreement between the results from the proposed method and the current manual process (eg, both suggest the presence of DM or neither suggests DM). Among such concordant cases, a randomized sample of 100 patients each (ie, a total of 200 concordant cases) was screened by an expert reviewer, who has 35 years of clinical nursing experience, including patient classification, and 16 years of experience in chart abstraction. This review was to confirm whether the patients indeed did or did not have DM (N=200). Second, we considered all discordant cases (ie, a total of 231 discordant cases) where the proposed method disagreed with the findings from the current

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mayo Clinic proposed</th>
<th>CCW</th>
<th>DDC</th>
<th>SUPREME-DM</th>
<th>eMERGE</th>
<th>HbA1c of NYC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.99 (0.90-0.99)</td>
<td>0.67 (0.54-0.79)</td>
<td>0.72 (0.61-0.79)</td>
<td>0.93 (0.90-0.96)</td>
<td>0.94 (0.93-0.96)</td>
<td>0.93 (0.91-0.95)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.99 (0.99-1.00)</td>
<td>1.00 (0.99-1.00)</td>
<td>1.00 (0.99-1.00)</td>
<td>1.00 (0.99-1.00)</td>
<td>1.00 (0.99-1.00)</td>
<td>1.00 (0.99-1.00)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.99 (0.99-0.99)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.99 (0.99-1.00)</td>
<td>0.99 (0.99-1.00)</td>
<td>0.99 (0.99-1.00)</td>
<td>0.99 (0.99-1.00)</td>
<td>0.99 (0.99-1.00)</td>
<td>0.99 (0.99-1.00)</td>
</tr>
<tr>
<td>McNemar $\chi^2$ test (sensitivity)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>43.31</td>
<td>43.31</td>
<td>43.31</td>
</tr>
<tr>
<td>P value (sensitivity)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;.01</td>
<td>&lt;.01</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>McNemar $\chi^2$ test (specificity)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>16.70</td>
<td>16.70</td>
<td>16.70</td>
</tr>
<tr>
<td>P value (specificity)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;.01</td>
<td>&lt;.01</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Total patients identified, n</td>
<td>684</td>
<td>460</td>
<td>815</td>
<td>543</td>
<td>377</td>
<td>543</td>
</tr>
</tbody>
</table>

CCW = Chronic Condition Data Warehouse; DDC = Durham Diabetes Coalition; eMERGE = Electronic Medical Records and Genomics Network; HbA1c of NYC = Hemoglobin A1c of New York City; NA = not applicable; SUPREME-DM = Surveillance, Prevention, and Management of Diabetes Mellitus.

The ascertainment of true DM incidence as the reference standard was created by comparing results from the proposed method with those of the existing process, which is a manual review of the EHR for a DM diagnosis by bedside nurses. When the bedside nurse identifies DM, its documentation is noted in the preoperative evaluation patient flow sheet. This nurse preoperative evaluation patient flow sheet document serves as a surgery intake tool, nursing communication tool, and assessment tool. The nurse documents whether the patient has DM or a history of DM in the flow sheet. This is based on patient response, surgical listing information, the patient’s current medication regimen, and a clinical notes review. The nurse synthesizes this information and documents in the EHR if the nurse confirms that the patient has DM. To develop this reference standard, we first considered concordant cases with agreement between the results from the proposed method and the current manual process (eg, both suggest the presence of DM or neither suggests DM). Among such concordant cases, a randomized sample of 100 patients each (ie, a total of 200 concordant cases) was screened by an expert reviewer, who has 35 years of clinical nursing experience, including patient classification, and 16 years of experience in chart abstraction. This review was to confirm whether the patients indeed did or did not have DM (N=200). Second, we considered all discordant cases (ie, a total of 231 discordant cases) where the proposed method disagreed with the findings from the current...
manual process. All discordant pairs were screened manually by the independent reviewer to determine whether the patients had DM. Thus, a total of 431 cases were reviewed. Final determination of the presence or absence of DM involved manual review of a patient’s EHR by a research nurse specifically trained in the extraction of medical conditions of interest from the EHR. The decision to manually review a random sample of concordant cases was determined a priori and was arbitrarily based on the collective perception of the research team.

Algorithm Comparison
After developing a reference standard and validating the proposed phenotyping algorithm, we compared the performance of this proposed algorithm with the performance of existing comparator algorithms, as well as current manual imputation of a DM diagnosis by bedside nurses (Table 1).

Statistical Analyses
Measures of diagnostic accuracy (sensitivity, specificity, positive predictive value, and negative predictive value) and 95% CIs were computed for each case-definition method, and the best method that identified patients with DM most accurately was determined on the basis of McNemar test for sensitivities and specificities.29,30 All these analyses were conducted using open source R version 3.1.2 (R Foundation for Statistical Computing).31

RESULTS
A total of 4208 patients scheduled for surgery in August 2014 were considered in this study. Baseline characteristics of the study population are outlined in Table 2. Our proposed method classified 684 patients as having DM (16.25% of the full study cohort). Of these, 503 (73.53%) were classified as having DM because of the presence of 1 or more outpatient ICD-9-CM codes relating to DM

| TABLE 2. Patients’ Characteristics at Baseline and Predisposing Factors for DMa,b |
|----------------------------------|----------------------------------|----------------------------------|
| Characteristic                   | All patients, (N=4208)           | Patients with DM present (n=684) | Patients with DM absent (n=3524) |
| Age (y)                          |                                  |                                  |                                 |
| <18                              | 501 (11.91)                      | 12 (1.75)                       | 489 (13.88)                     |
| 18-21                            | 83 (1.97)                        | 1 (0.15)                        | 82 (2.33)                      |
| 22-29                            | 201 (4.77)                       | 10 (1.46)                       | 191 (5.42)                     |
| 30-39                            | 337 (8.00)                       | 29 (4.24)                       | 308 (8.74)                     |
| 40-49                            | 455 (10.81)                      | 94 (13.17)                      | 361 (10.24)                    |
| 50-65                            | 1256 (29.84)                     | 284 (41.52)                     | 972 (27.58)                    |
| 66-80                            | 1129 (26.82)                     | 230 (33.63)                     | 899 (25.51)                    |
| >80                              | 246 (5.84)                       | 24 (3.51)                       | 222 (6.30)                     |
| BMI                              |                                  |                                  |                                 |
| Underweight                      | 378 (8.98)                       | 5 (0.73)                        | 373 (10.58)                    |
| Normal                           | 1105 (26.26)                     | 63 (9.21)                       | 1042 (29.57)                   |
| Overweight                       | 1199 (28.49)                     | 176 (25.73)                     | 1023 (29.03)                   |
| Obese                            | 1489 (35.38)                     | 439 (64.18)                     | 1050 (29.80)                   |
| Unavailable                      | 37 (0.88)                        | 1 (0.15)                        | 36 (1.02)                      |
| Sex                              |                                  |                                  |                                 |
| Female                           | 2152 (51.14)                     | 298 (43.57)                     | 1854 (52.61)                   |
| Male                             | 2056 (48.86)                     | 386 (56.43)                     | 1670 (47.39)                   |
| Race                             |                                  |                                  |                                 |
| American Indian/American Native  | 25 (0.59)                        | 14 (2.14)                       | 11 (0.31)                      |
| Non-Hispanic white               | 3848 (91.44)                     | 615 (90.05)                     | 3233 (91.73)                   |
| African American                 | 66 (1.57)                        | 12 (1.68)                       | 54 (1.53)                      |
| African                          | 11 (0.26)                        | 1 (0.15)                        | 10 (0.28)                      |
| Asian                            | 67 (1.59)                        | 9 (1.23)                        | 58 (1.65)                      |
| Native Hawaiian/Pacific Islander | 5 (0.12)                        | 1 (0.15)                        | 4 (0.11)                       |
| Other                            | 186 (4.42)                       | 32 (4.59)                       | 154 (4.37)                     |

aBMI = body mass index; DM = diabetes mellitus.
bValues are n (%).
diagnosis, the number of cases identified by other criteria such as DM-related medications (Supplemental Appendix 2, available online at http://www.mcpiqojournal.org), abnormal laboratory values, and metformin use. DM annotations in patient notes are outlined in Figure 2. A single false-positive finding was noted with the keyword searches because of documentation noting “denis [sic] diabetes” instead of “denies diabetes.” Of the other 3524 patients identified as not having DM, 2 patients were incorrectly classified as not having DM by the proposed algorithm (false-negative results). This misclassification was due to a lack of access to source systems that contained DM-related information for these 2 patients. The details are graphically represented using the classification tree diagram of Figure 2.

Of note, the rules implemented in the proposed DM phenotyping algorithm are not mutually exclusive. Nineteen patients were identified as having DM because of the coexistence of metformin and an abnormal DM laboratory value and also met the criteria for DM on the basis of the presence of a DM-related keyword in their clinical notes. Similarly, the cases identified because of overlapping inclusion criteria are depicted in Figure 3.

Compared with the 684 patients identified by our proposed approach, the number of DM cases identified by the comparator methods varied considerably: eMERGE, 377 cases; CCW, 460 cases; DDC, 815 cases; SUPREME-DM, 543 cases; HbA1c of NYC, 350 cases, and the Harvard Medical School method, 1043 cases. The current manual approach identified 626 patients as having DM (Table 1).

Our proposed method resulted in a sensitivity of 0.9971 (95% CI, 0.9895-0.9996) and a specificity of 0.9997 (95% CI, 0.9984-1.00). The sensitivity of the present phenotyping algorithm exceeded the sensitivity of all other existing methods considered in our study (Table 1). In contrast, the McNemar test for comparison of specificities failed to detect a statistically significant difference between the
specificity on the basis of the present approach and specifications proposed by either CCW (0.99; 95% CI, 0.99-1.00; McNemar test $\chi^2$, 1.00; $P = .32$) or eMERGE (1.00; 95% CI, 0.99-1.00; McNemar test $\chi^2$, 1.00; $P = .32$). However, statistically significant differences in specificities were observed between our approach and the approaches recommended by DDC, SUPREME-DM, HbA1c of NYC, and Harvard Medical School and the manual approach.

**DISCUSSION**

With the rapid adoption of EHRs and the development of innovative data extraction strategies, many institutions have initiated or have completed the development of automated computable case-definition methods. Our results indicate that most of these preexisting methods achieved a high sensitivity of patient identity. Yet, the degradation in the sensitivity of these preexisting methods compared with the current proposed method has many causes. For instance, SUPREME-DM requires that a patient have at least 2 abnormal laboratory values or outpatient ICD-9-CM diagnosis codes not more than 2 years apart (Table 3). The DDC requires 2 abnormal laboratory values not more than 1 year apart. This requirement excludes patients with a single abnormal laboratory value or ICD-9-CM diagnosis codes. Such data result in increased false negatives and reduced sensitivity. CCW and HbA1c of NYC are single-element case-definition methods. The CCW requires only patient DM-related ICD-9-CM codes, and HbA1c of NYC requires only abnormal HbA1c values and does not identify patients with other indicators, such as DM-related medications or DM-related ICD-9-CM diagnostic codes and resulting in large false negatives and lower sensitivity.

**TABLE 3. Overview of Differences Among Other Preexisting Methods Proposed by Other Authors**

<table>
<thead>
<tr>
<th>Existing diabetes phenotype</th>
<th>Data elements used</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCW</td>
<td>Chronic Condition Data Warehouse</td>
<td>x</td>
<td>Only ICD-9 codes</td>
</tr>
<tr>
<td>DDC</td>
<td>Spratt et al 12</td>
<td>x x x</td>
<td>$\geq 2$ abnormal lab values, not more than 1 y apart</td>
</tr>
<tr>
<td>SUPREME-DM</td>
<td>Desai et al 15</td>
<td>x x x</td>
<td>$\geq 2$ abnormal lab values or ICD-9, $\leq 2$ y apart; insured patients only</td>
</tr>
<tr>
<td>eMERGE</td>
<td>Kho et al 18</td>
<td>x x x</td>
<td>Mainly designed to identify patients with T2DM and must satisfy $\geq 2$ criteria</td>
</tr>
<tr>
<td>HbA1c NYC</td>
<td>Chamary et al 14</td>
<td>x</td>
<td>Only HbA1c</td>
</tr>
<tr>
<td>Harvard Medical School</td>
<td>Klompas et al 19</td>
<td>x x x</td>
<td>Only abnormal lab value with no metformin requirement</td>
</tr>
</tbody>
</table>

CCW = Chronic Conditions Data Warehouse; DDC = Durham Diabetes Coalition; eMERGE = Electronic Medical Records and Genomics Network; HbA1c = hemoglobin A1c; HbA1c NYC = HbA1c of New York City; ICD-9 = International Classification of Diseases, Ninth Edition; lab = laboratory; SUPREME-DC = Surveillance, Prevention and Management of Diabetes Mellitus; T2DM = type 2 diabetes mellitus.
The eMERGE DM algorithm requires at least 2 different DM-related data elements to classify patients as having DM and ignores patients with a single instance of either DM-related ICD-9-CM code or medication or metformin and abnormal values. The algorithm does not categorize these patients as having DM, thus lowering its sensitivity. In addition, neither method considered exploring the DM cases mentioned by the physicians in free-text patient notes data. This point further contributed to lower sensitivity among these different comparator methods.

Similar concerns exist when considering the varied specificities of the alternative algorithms. The rules proposed by the SUPREME-DM, DDC, and Harvard Medical School methods resulted in more false positives than with our proposed method because these rules classified patients as having DM solely on the basis of abnormal laboratory values. Concerns about the use of abnormal laboratory values as the singular most important indicator of DM have been raised by several investigators. Sacks et al32 have suggested that neither random nor fasting glucose concentrations measured in an accredited laboratory should be used as the primary indicator of DM. Instead, they recommended supplementing these criteria with other information. Similarly, Monnier et al33 stated that no unifying argument supports the accuracy of random or fasting glucose test when identifying patients who may have DM. The methods proposed by SUPREME-DM, DDC, and Harvard Medical School that identified patients with DM through only abnormal laboratory values resulted in large false positives and lower specificity compared with our proposed method.

Strengths and Limitations

A major strength of this investigation is its inclusion of information from multiple domains relevant to DM. This included ICD-9-CM diagnostic codes, medication information, laboratory findings, and keyword searches mining unstructured text contained within clinical notes.

Several investigators have emphasized the importance of incorporating information buried within a patient’s clinical notes when attempting to identify patients with specific disease conditions.34-36 Sohn et al37 noted that rule-based algorithms performed well when NLP-extracted attributes from clinical text were incorporated into the algorithm design. Peissig et al38 studied cataract cases and asserted that incorporating concepts from clinical notes increased the case identification by a factor of 3 compared with using a sole mode-structured data approach. We used a combination of complex Structured Query Language embedded with Boolean logic (“and” “or”) to detect either negation or positive annotation indicative of DM in patient notes (Supplemental Appendix 3, available online at http://www.mcpiqojournal.org). Within such a construct, we were able to identify several dominant patterns and develop a set of inclusion and exclusion criteria that indicated the presence or absence of DM. In addition to identifying various affirmation statuses, negation had a notable role in identifying patients who did not have DM and in improving the accuracy of our algorithm. Unlike confirmatory avowals, negations are commonly indicated by such terms as no, NA, and unknown, so initially we started with these 3 basic terms and iteratively refined our algorithm by adding more negation clauses, such as brothers, one of whom has diabetes type, insignificant for diabetes mellitus, doesn’t have a history of diabetes, diabetes screen: fasting glucose = NA, and diabetes screen: unknown if ever checked, and many others during the iterative validation process.

This combination attributed 79 additional cases identified by our algorithm, resulting in performance superior to that of other approaches that did not include unstructured data. Thus, it follows that such a text-parsing NLP algorithm that uses simple keywords to negate or validate the presence of DM will be an EHR system—agnostic and highly portable that can be implemented in other EHR environments.

Inclusion of metformin as an additional requirement for patients with abnormal laboratory values is another distinctive aspect of our algorithm that may have resulted in higher specificity than with other available phenotyping methods. Our method also did not mandate multiple observations per patient at various time intervals. This proposed method has the utility of identifying patients before their surgical encounter rather than after it.
Finally, implementation of the manual review for final determination of DM by a trained study coordinator enhanced our confidence in accuracy regarding the presence or absence of the condition of interest.

Several study limitations merit discussion. First, this investigation represents a single institution's experience in designing, testing, and implementing a DM case-definition algorithm compared with multi-institution case-definition methods, such as eMERGE, SUPREME-DM, and Harvard Medical School. Further validation and confirmation of the present findings in alternate health care settings across various health care institutions are necessary to understand the generalizability of our proposed method. Many health care institutions will not have access to some of the data elements proposed by our case-definition strategy—most notably, the ability to query unstructured data for key terms of interest. Although we were able to validate the results from our algorithm using an independent nurse reviewer, we were unable to completely mask the details to the reviewer during the validation because of the iterative nature of algorithm development.

In addition to these concerns, we acknowledge the potential for bias due to the use of a reference standard and not a criterion standard developed by an independent reviewer to whom results were blinded completely while evaluating the accuracy of our proposed method in our study. However, the CIs around our sensitivity and specificity estimates support the stability of our results to our results and mitigate concerns related to the use of such reference standards during the validation process. Currently, a more formal validation of the performance of our proposed method in a clinical environment at our institution is under way.

Although the use of International Classification of Diseases, Tenth Edition codes is the current standard, this study used ICD-9-CM codes. This course was taken because the standard at the time of study onset was ICD-9-CM. Moreover, the comparator algorithms (DDC, CCW, eMERGE, Harvard Medical College, Supreme-DM) were also developed using ICD-9-CM codes. Nonetheless, we acknowledge the use of ICD-9-CM codes in our study as a limitation, and we have implemented a strategy to map the ICD-9-CM codes associated with this study to their corresponding International Classification of Diseases, Tenth Edition codes.

CONCLUSION
We developed an efficient and accurate DM phenotypic algorithm that outperformed other available approaches. This algorithm facilitates the early recognition of DM, thereby permitting the implementation of improved workflows for optimal DM-related care in the perioperative period.

SUPPLEMENTAL ONLINE MATERIAL
Supplemental material can be found online at http://www.mcpiqojournal.org. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronyms. CCW = Chronic Condition Data Warehouse; DDC = Durham Diabetes Coalition; DM = diabetes mellitus; eMERGE = Electronic Medical Records and Genomics; EHR = electronic health record; HbA1c = hemoglobin A1c; HbA1c of NYC = Hemoglobin A1c of New York City; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification; MICS = Mayo Integrated Clinical Systems; NLP = natural language processing; SUPREME-DM = Surveillance, Prevention, and Management of Diabetes Mellitus; T1DM = type 1 diabetes mellitus; T2DM = type 2 diabetes mellitus

Affiliations (Continued from the first page of this article): and Robert D. and Patricia E. Kern Center for Science of Health Care Delivery (R.R.C.), Mayo Clinic, Rochester, MN.

Grant Support. The work was supported in part by the Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery and a grant from the National Institutes of Health (NIH) (grant number R01 GM105688-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funding sources for this work had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Correspondence. Address to Daryl J. Kor, MD, Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (kdaryl@mayo.edu).

REFERENCES
1. Ponce BA, Menendez ME, Oladeji LO, Soldado F. Diabetes as a risk factor for poorer early postoperative outcomes


29. Carter RE. I can’t make heads or tails out of what you are saying, so let’s just agree to be fair. Teach Stat. 2001;23(3):127-130.


