

## RESEARCH ARTICLE

## Open Access

# Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm

Anil N Makam<sup>1\*</sup>, Oanh K Nguyen<sup>1</sup>, Billy Moore<sup>2</sup>, Ying Ma<sup>2</sup> and Ruben Amarasingham<sup>2,3</sup>**Abstract**

**Background:** Effective population management of patients with diabetes requires timely recognition. Current case-finding algorithms can accurately detect patients with diabetes, but lack real-time identification. We sought to develop and validate an automated, real-time diabetes case-finding algorithm to identify patients with diabetes at the earliest possible date.

**Methods:** The source population included 160,872 unique patients from a large public hospital system between January 2009 and April 2011. A diabetes case-finding algorithm was iteratively derived using chart review and subsequently validated ( $n = 343$ ) in a stratified random sample of patients, using data extracted from the electronic health records (EHR). A point-based algorithm using encounter diagnoses, clinical history, pharmacy data, and laboratory results was used to identify diabetes cases. The date when accumulated points reached a specified threshold equated to the diagnosis date. Physician chart review served as the gold standard.

**Results:** The electronic model had a sensitivity of 97%, specificity of 90%, positive predictive value of 90%, and negative predictive value of 96% for the identification of patients with diabetes. The kappa score for agreement between the model and physician for the diagnosis date allowing for a 3-month delay was 0.97, where 78.4% of cases had exact agreement on the precise date.

**Conclusions:** A diabetes case-finding algorithm using data exclusively extracted from a comprehensive EHR can accurately identify patients with diabetes at the earliest possible date within a healthcare system. The real-time capability may enable proactive disease management.

**Background**

Practice redesign efforts are shifting the paradigm from volume to value in healthcare in part by emphasizing care coordination, population health, and performance reporting. To this end, the National Committee for Quality Assurance (NCQA) requires practices to use patient tracking, disease registries and certified electronic health records (EHR) in order to qualify for patient-centered medical home (PCMH) and accountable care organization (ACO) accreditation [1,2].

Diabetes is well-suited to the principles of the PCMH and ACO, given that it affects 25.8 million people, [3] costs

\$174 billion annually, [4] and despite well-established and effective guidelines only 45% of diabetics receive recommended care [5,6]. Effective practice redesign efforts aimed at improving the care of diabetes will require proactive identification of patients with diabetes to narrow the knowledge-to-action gap. While existing diabetes case-finding algorithms are able to accurately identify patients with diabetes, [7-17] such algorithms rely on historical rather than real-time data. As a result, there may be a lag time between when a patient receives a diagnosis of diabetes in the clinical setting compared to when the patient is identified as a diabetic by a case-finding algorithm for the purpose of population management. Because preventing complications of diabetes depends critically on timely intervention, [18] this lag impedes the potential for case-finding algorithms to significantly affect prevention of such complications across diabetic populations.

\* Correspondence: [anil.makam@utsouthwestern.edu](mailto:anil.makam@utsouthwestern.edu)

<sup>1</sup>Division of General Internal Medicine, University of California San Francisco, Box 1211, Laurel Heights Campus, Room 383, 3333 California St., San Francisco, CA 94143, USA

Full list of author information is available at the end of the article

No published case-finding algorithm, that we know of to date, takes advantage of a comprehensive EHR that would allow for a fully automated and electronic system obviating the need for manual data entry, and importantly, real-time availability for data capture and identification [19]. Therefore, this study aims to derive and validate an electronic case-finding model (e-model) that could be used in real-time to identify patients who meet criteria for diabetes at the earliest possible date based on EHR data within a healthcare system.

## Methods

### Study population

The e-model was developed using historical data extracted from an EHR (Epic Systems Corporation, Verona, WI) deployed across inpatient and outpatient settings within Parkland Health & Hospital System (PHHS), a large urban safety-net health system in Dallas, TX. We used data from 160,872 unique adult patients (age  $\geq 18$  years) who had a first encounter within PHHS between January 1, 2009 and April 1, 2011.

### Definition of algorithm variables

To determine the criteria used by the e-model to identify diabetes, we used a combination of diagnostic criteria from the American Diabetes Association (ADA) [5] and data elements identified by a panel of physician health service researchers, including a board certified endocrinologist (AM, CR, RA). Only those data elements which were routinely documented and extractable from structured data fields (e.g., encounter diagnosis, past medical history, problem list, medications, and laboratory results) within the EHR were included as variables in the e-model.

### Derivation of E-model

We used a point-based algorithm to identify the presence of diabetes and to determine the date that a diagnosis could have been made through information in the EHR. Each variable in the e-model was initially assigned a fractional point value, proportionate to its perceived relative contribution in diagnosing diabetes. Point totals of 0, between 0 and 1, and  $\geq 1$  were set as thresholds for e-model determination of 'no diabetes,' 'possible diabetes,' and 'diabetes,' respectively. The e-model determined the diagnosis date as the date when accumulated points reached or exceeded a threshold value of 1.

The perceived relative contribution of each variable was determined based on ADA diagnostic criteria, [5] existing diabetes case-finding algorithms, [10] and expert opinion. For example, since the ADA requires two fasting blood glucose values of  $\geq 126$  mg/dL for the diagnosis of diabetes, the presence of a single fasting blood glucose value of  $\geq 126$  mg/dL was assigned a point value of 0.5, such that two fasting blood glucose values would

give an individual a total of 1 point for an e-model identification of 'diabetes.'

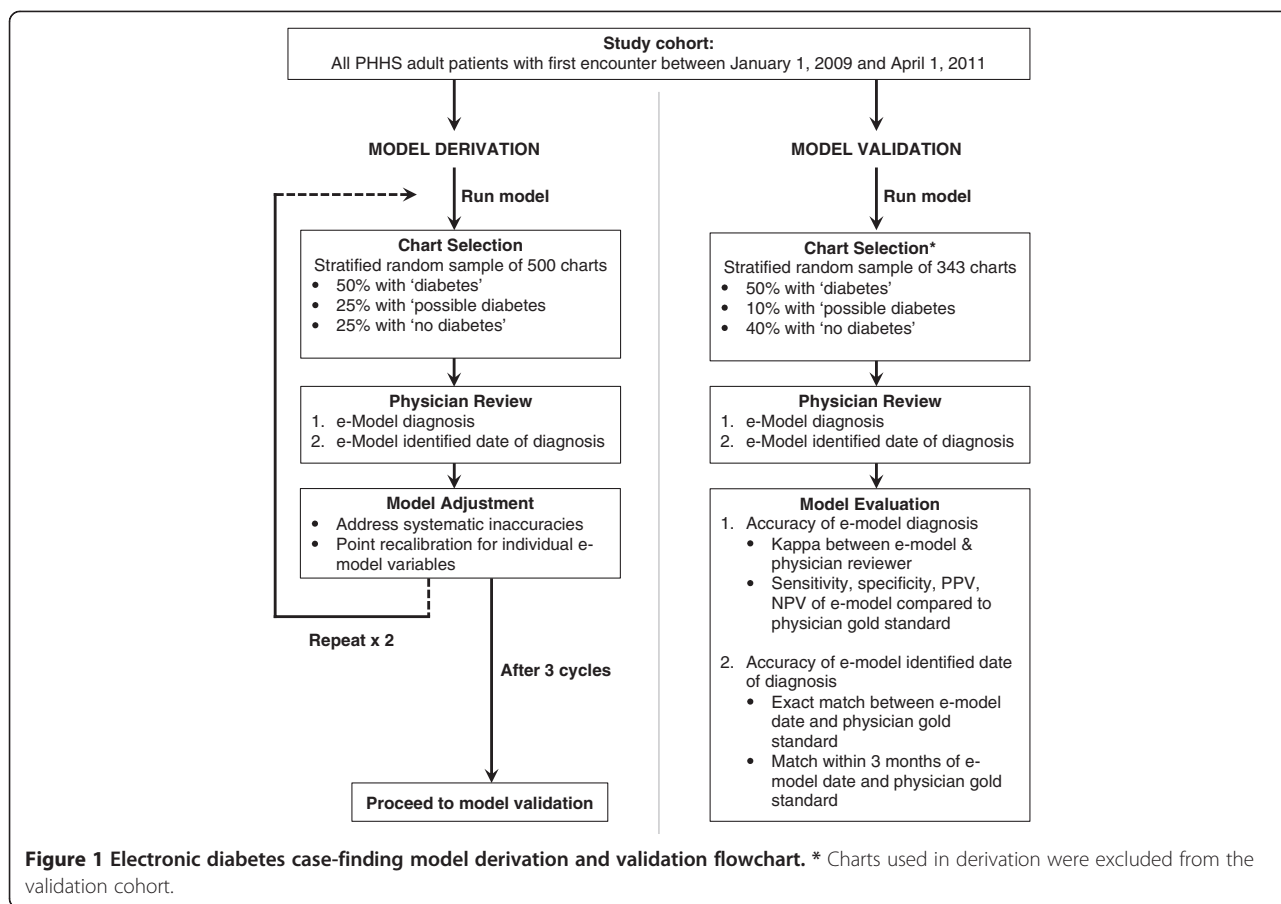
Point assignments were subsequently recalibrated through a clinically-guided strategy, consisting of an iterative, three-stage evaluation process (Figure 1). At each stage, a stratified random sample of up to 500 charts, with 50% 'diabetes,' 25% 'possible diabetes,' and 25% 'no diabetes' as determined by the e-model, underwent unblinded chart review by a physician to evaluate the accuracy of the e-model identification of diabetes and diagnosis date. To allow better evaluation of e-model performance, the 'no diabetes' group was restricted to individuals 50 years or older since the incidence of diabetes is strongly associated with age and may increase the potential to identify false negatives. To allow for evaluation of the accuracy of the e-model diagnosis date, 50% of the 'diabetes' group was selected to have accumulated  $\geq 1$  point(s) on a date more recent than the date of the first encounter. This allowed for a potentially earlier date of diagnosis to be determined by chart review.

Points for individual variables were reweighted after each stage based on commonly recurring e-model inaccuracies and were finalized after three successive stages (Table 1).

Through the derivation process, we adjusted the point values for diabetes medication, problem list and past medical history, and ICD-9 encounter diagnosis. The presence of any diabetes medication was initially assigned a point value of 1, since we considered this a surrogate for the presence of diabetes given that the primary and often only indication is the treatment of hyperglycemia. However, during chart review, metformin was found to be occasionally prescribed for pre-diabetes and polycystic ovarian syndrome, and individuals with only the presence of metformin were incorrectly identified by the e-model as 'diabetes.' Therefore, the point value for metformin was decreased to 0.75. The point value for the presence of diabetes in the past medical history or problem list fields were reduced to 0.4 because the data in these fields were found to be often inaccurate and outdated. Lastly, a single ICD-9 encounter diagnosis in the absence of other variables incorrectly identified patients as having diabetes in most cases. Existing case-finding algorithms have also found the presence of two ICD-9 codes across outpatient and inpatient settings to be highly sensitive and more specific than a single code for the diagnosis of diabetes [20]. Therefore, the encounter diagnosis variable was adjusted from 1 to 0.75 points.

### Validation of E-model

To validate the e-model, we compared the e-model identification of 'diabetes,' 'possible diabetes' and 'no diabetes' and date of diagnosis to the gold standard of physician chart review. Based on a conservative estimate



of 70% sensitivity +/- 10% error at a two-sided alpha of 0.05 and power of 80%, we determined that a stratified random subset of 343 patients (50% 'diabetes,' 10% 'possible' diabetes, and 40% 'no diabetes') was needed to adequately validate the e-model. Charts reviewed during the derivation process were not included in the validation cohort.

Blinded chart review for an initial 50 patients was completed by two board-certified internists working independently (AM and ON). Inter-rater agreement between reviewers for the classification of diabetes status was 0.80 with a linear weighted kappa statistic and 0.94 for the exact diagnosis date when both reviewers agreed that diabetes was present (n = 19). The remaining charts

**Table 1 Variables included in the electronic diabetes case-finding model**

Identification variable	Criteria*	Encounter location	Point value
ICD-9 encounter code	250.xx	Inpatient or outpatient	0.75
Hemoglobin A1c	≥ 6.5%	Inpatient or outpatient	1.00
Fasting Blood Glucose	≥ 126 mg/dL	Outpatient only	0.50
Random blood glucose	≥ 200 mg/dL	Outpatient only	0.50
2-hour OGTT	≥ 200 mg/dL	Inpatient or outpatient	0.75
Problem list or PMH	250.xx	Inpatient or outpatient	0.40
Diabetes medication**	Present	Outpatient only	1.00
Metformin	Present	Outpatient only	0.75

Abbreviations: ICD-9: International Classification of Diseases 9, OGTT: oral glucose tolerance test, PMH: past medical history, mg: milligram, dL: deciliter.

\*Criteria are counted only once except for ICD-9 codes (maximum twice) and random and fasting blood glucose (maximum twice each) as long as repeated glucose values are ≥ 3 months apart.

\*\*Insulin, sulfonylurea, thiazolidinedione, alpha glucosidase, DPP-4 inhibitor, meglitinide, amylin mimetic, incretin mimetic, combination medication.

were reviewed by a single physician after establishing adequate inter-rater reliability.

To examine model performance, the “possible diabetes” group was combined with the “no diabetes” group to allow for a dichotomous classification. Two sensitivity analyses were performed on the validation cohort to 1) determine the optimal point threshold for identification of diabetes and 2) determine the effect of an alternate dichotomous grouping on e-model performance. For the first sensitivity analysis, we varied the point threshold for the identification of diabetes to test for the optimal cutoff point to maximize e-model performance. For the second, we combined the “possible diabetes” group with the “diabetes” group to examine the effect on e-model sensitivity and specificity.

We then evaluated the performance of the e-model on correct identification of the earliest diagnosis date. The e-model date was considered to be in agreement with the physician date of diagnosis if it was within a 3 month window of the physician-determined date. The time interval of 3 months was chosen based on the ADA consensus opinion which recommends repeat glycemic testing in 3 months for newly diagnosed or poorly controlled diabetes patients [5]. Lastly, we compared the performance of the e-model in identifying the date of diagnosis to the performance of a simplified claims-based case-finding algorithm that would not require the presence of an EHR. This method utilized the presence of only two diabetes encounter codes (ICD-9 250) to identify the diagnosis date.

### Statistical analysis

The diagnostic performance of the e-model on identification of diabetes status was evaluated first by the inter-rater agreement between e-model classification and physician classification of diabetes status using the kappa statistic and second, by using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The optimal point threshold was evaluated by the receiver-operator curve. We evaluated the performance of the e-model on correct identification of the diagnosis date by the inter-rater agreement between the e-model date of diagnosis and physician determination of date of diagnosis for individuals identified as ‘diabetes’ by the physician.

Analyses were conducted using STATA statistical software (version 12.0; STATA Corp, College Station, TX). The University of Texas Southwestern Medical Center Institutional Review Board approved the research protocol.

### Results

Using the e-model to characterize the overall study cohort, the source population included 14,025 (8.7%) patients

identified as having ‘diabetes,’ 1,882 (1.2%) patients as ‘possible diabetes,’ and 144,965 (90.1%) patients as ‘no diabetes.’ In the overall diabetic population (n = 14,205), the mean age was 52 years (+/- 13 years), 44% were Hispanic, 27% black, 22% white, 53% male, 68% had a primary payer of self-pay or charity, and the mean number of healthcare encounters was 6.5. Of the 1,500 charts for model derivation, 83 were excluded because of duplicated patients or age less than 18 years. Of the 343 charts for validation, 2 were excluded from analysis because the chart did not exist (n = 1) or was a duplicate (n = 1). Patients in the derivation and validation cohorts were similar with respect to age, race and ethnicity, sex, and primary payer, but patients in the derivation group had a slightly greater number of encounters over a one-year period (Table 2).

### E-model performance on identification of diabetes

The kappa statistic between the e-model and physician reviewer on the question of whether diabetes was present was 0.76 (p < 0.001) with 86% overall agreement. Combining the “possible diabetes” group with the “no diabetes” group, the sensitivity, specificity,

**Table 2 Baseline cohort characteristics for the electronic diabetes case-finding model\***

Characteristics	Derivation	Validation	P-value**
n	1417	341	
Age, mean years (SD)	49.3 (14.4)	45.0 (14.6)	<.001
Race/ethnicity,%			.83
Hispanic	44	40	
White	23	24	
Black	25	27	
Other	9	9	
Male,%	47	52	.13
Primary payer,%			.15
Commercial	18	16	
Medicare	9	6	
Medicaid	13	13	
Self-pay	30	29	
Charity	30	36	
Encounters, mean no. (SD)			
All	7.86 (9.03)	6.65 (9.06)	.01
Primary care	2.25 (3.90)	1.99 (3.60)	.45
Specialty care	4.31 (6.86)	3.62 (7.08)	.02
Urgent care and ED	0.81 (2.58)	0.61 (0.89)	.48
Inpatient	0.49 (0.97)	0.43 (0.88)	.25

\*Derivation and validation cohorts defined as per Figure 1.

\*\*Student t-test for age;  $\chi^2$  tests for race/ethnicity, male sex, and primary payer; Wilcoxon rank-sum test for encounters.

positive predictive value, and negative predictive value of the e-model were 97%, 90%, 89%, and 97% respectively. Alternatively, when the “possible diabetes” group was combined with the “diabetes” group, the sensitivity and specificity of the e-model were 99% and 81% respectively.

The performance of the e-model by different point thresholds for the identification of diabetes corroborated using a threshold of 1 point to optimize sensitivity and specificity (Figure 2).

#### E-model performance on earliest date of diagnosis

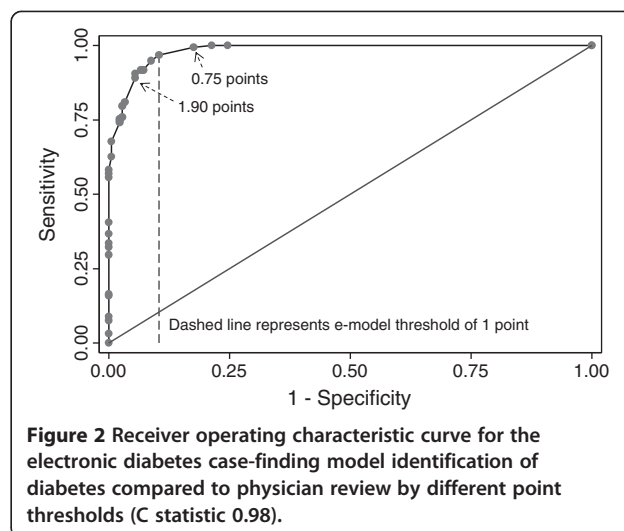
The kappa score between e-model and physician on the date of diagnosis was 0.94 with agreement on the exact date in 76% of the cases. Among the cases where both the physician and e-model made a diagnosis of diabetes, only 4 observations (2.6%) were diagnosed by the e-model more than 3 months after the correct date (Figure 3).

However, a more simplified claims-based algorithm that used only two ICD-9 encounter diagnoses (e.g. the diagnosis date equals the date of the second ICD-9 code) did not accurately identify the diagnosis date compared to a physician reviewer. Using this approach, the kappa statistic was 0.62 with only 4.6% agreement on the exact date. Notably, 27% of the cases ( $n = 33$ ) did not have a second ICD-9 encounter code during the study period.

#### Discussion

We developed an electronic case-finding algorithm that accurately identified patients with diabetes at their earliest possible date within a healthcare system using data extracted from an EHR. The performance of our model in identifying patients with diabetes is comparable to other diabetes case-finding algorithms [10-17]. However, the distinct advantage of our automated, real-time algorithm is the timely recognition of diabetes. Relying on only two ICD-9 encounter codes to establish the diagnosis date, a quarter of the cases in our cohort would have been missed and another 11% would have had a delayed diagnosis. By using multiple data elements we were able to identify the date of diagnosis within three months of a physician’s chart review date in 94% of case, with three-quarters of cases having a perfect date match.

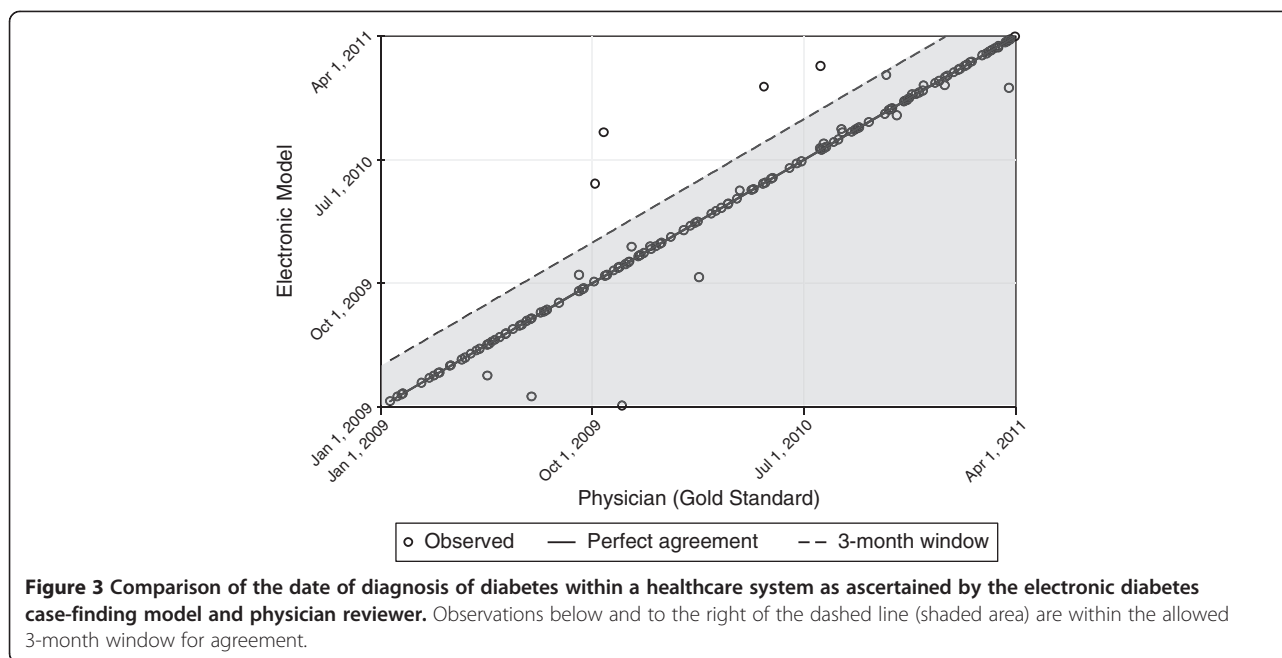
Achieving early glycemic control in patients with newly diagnosed diabetes reduces the risk of microvascular complications, myocardial infarction, and all-cause mortality [18]. Attaining the benefits of instituting early treatment requires timely diagnosis. In the ARIC cohort, a population-based prospective study of middle-aged adults, Samuels et al. found that even with an effective screening program the median delay from the



onset of diabetes to physician diagnosis was 2.4 years, with more than 7% of incident cases remaining undiagnosed for at least 7.5 years [21]. In addition, delayed diagnoses are more widespread in safety-net settings where patients may have more fragmented, episodic care [22]. Real-time, automated patient identification and tracking can help healthcare systems close this gap and facilitate the delivery of timely, effective therapy at the point-of-care at the earliest possible date [19].

Improving care for diabetics is increasingly important for healthcare systems in today’s pay-for-performance climate. The high cost, rising prevalence, and documented quality gap has positioned diabetes in the forefront of policies benchmarking performance. To qualify for financial incentives and avoid public scrutiny, healthcare systems are increasingly faced with the challenge to achieve acceptable rates in their diabetic population for targeted metrics such as glycated hemoglobin, low-density lipoprotein, and screening for microalbuminuria. Our electronic case-finding algorithm, which leverages real-time data in the EHR, can enable proactive management of these quality measures. Healthcare systems may additionally apply this algorithm to provide feedback to providers on the quality of their care, generate letters to patients, identify underperforming clinics for quality improvement initiatives, link clinical decision support tools to inform decision making at the point-of-care, and risk stratify diabetic patients to direct limited resources to patients at greatest risk for developing complications.

Our study has several limitations. First, as with other registries, the limitations of miscoding and misclassification prohibited subtype distinction between type 1, type 2, and secondary diabetes [23]. Second, due to limits in study costs, we established an enriched prevalence of diabetes of 50% in our validation cohort to reduce the



number of charts required for manual chart review. While this inflated the positive and negative predictive values, the sensitivity and specificity of our algorithm remained unaffected and were comparable to other diabetes case-finding algorithms. Third, by using the e-model algorithm to select our validation cohort, we were unable to evaluate how individual data elements performed in identifying diabetes. Fourth, the direct applicability of our algorithm to other settings is unknown because of differences in practice style, EHR integration across outpatient and inpatient settings, and EHR documentation. Systems with greater accuracy in EHR documentation may need to increase the relative weight of the problem list and past medical history field to maximize the model's performance. With proper weight adjustments we expect our algorithm to be suitable to a wide range of healthcare settings. Automated machine learning techniques may provide approaches to model adjustment that could minimize manual recalibration and allow larger scales of dissemination. Lastly, in clinical settings transitioning from paper-based records to an EHR, the e-model may not accurately distinguish between newly established versus preexisting cases of diabetes within a healthcare system [24].

## Conclusion

Our electronic case-finding algorithm can accurately identify patients with diabetes at the earliest possible date within a healthcare system. We believe this algorithm can be used by healthcare systems with comprehensive EHRs

to build real-time diabetes identification systems. This is foundational to diabetes "system awareness," or building information systems that are able to construct and maintain awareness of a patient's status across time, setting, provider, and context.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

All authors in this study have been involved in concept and design, critical revision of the manuscript for important intellectual content, and giving final approval of the manuscript for submission. AM and RA designed the study. AM and ON served as physician chart reviewers. AM performed data analysis, interpretation and drafting of the manuscript, with significant contributions from ON, YM, and RM. YM participated in refinement of derivation process and data analytics. All authors read and approved the final manuscript.

## Acknowledgments

The authors would like to acknowledge Dr. Chang Rhee, a board-certified endocrinologist at UT Southwestern, who provided content expertise in selecting and weighting candidate variables. This study was supported by the grant, "Dallas Social-Health Information Exchange Project," from the W.W. Caruth Jr. Foundation Fund of the Communities Foundation of Texas. AM had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. AM and ON were supported by a federal training grant from the National Research Service Award (NRSA T32HP19025-07-00).

## Author details

<sup>1</sup>Division of General Internal Medicine, University of California San Francisco, Box 1211, Laurel Heights Campus, Room 383, 3333 California St., San Francisco, CA 94143, USA. <sup>2</sup>Parkland Center for Clinical Innovation, Dallas, TX, USA. <sup>3</sup>Division of General Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390, USA.

Received: 7 April 2013 Accepted: 26 July 2013

Published: 1 August 2013

## References

- Berwick DM: **Launching accountable care organizations—the proposed rule for the Medicare shared savings program.** *N Engl J Med* 2011, **364**:e32.
- Stange KC, Nutting PA, Miller WL, Jaen CR, Crabtree BF, Flocke SA, Gill JM: **Defining and measuring the patient-centered medical home.** *J Gen Intern Med* 2010, **25**:601–612.
- Centers for Disease Control and Prevention: *National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011*, Book centers for disease control and prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. City: U.S: Department of Health and Human Services, Centers for Disease Control and Prevention; 2011.
- American Diabetes Association: **Economic costs of diabetes in the U.S. In 2007.** *Diabetes Care* 2008, **31**:596–615.
- American Diabetes Association: **Standards of medical care in diabetes—2011.** *Diabetes Care* 2011, **34**(1):S11–S61.
- McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA: **The quality of health care delivered to adults in the United States.** *N Engl J Med* 2003, **348**:2635–2645.
- Chin MH, Cook S, Drum ML, Jin L, Guillen M, Humikowski CA, Koppert J, Harrison JF, Lippold S, Schaefer CT: **Improving diabetes care in midwest community health centers with the health disparities collaborative.** *Diabetes Care* 2004, **27**:2–8.
- Engelgau MM, Geiss LS, Manninen DL, Orians CE, Wagner EH, Friedman NM, Hurley JS, Trinkaus KM, Shatin D, Van Vorst KA: **Use of services by diabetes patients in managed care organizations. Development of a diabetes surveillance system. CDC diabetes in managed care work group.** *Diabetes Care* 1998, **21**:2062–2068.
- Grant RW, Cagliero E, Sullivan CM, Dubey AK, Estey GA, Weil EM, Gesmundo J, Nathan DM, Singer DE, Chueh HC, Meigs JB: **A controlled trial of population management: diabetes mellitus: putting evidence into practice (DM-PEP).** *Diabetes Care* 2004, **27**:2299–2305.
- Khan L, Mincemoyer S, Gabbay RA: **Diabetes registries: where we are and where are we headed?** *Diabetes Technol Ther* 2009, **11**:255–262.
- Pogach LM, Hawley G, Weinstock R, Sawin C, Schiebe H, Cutler F, Zieve F, Bates M, Repke D: **Diabetes prevalence and hospital and pharmacy use in the veterans health administration (1994). Use of an ambulatory care pharmacy-derived database.** *Diabetes Care* 1998, **21**:368–373.
- Pollard C, Bailey KA, Petite T, Baus A, Swim M, Hendryx M: **Electronic patient registries improve diabetes care and clinical outcomes in rural community health centers.** *J Rural Health* 2009, **25**:77–84.
- Selby JV, Karter AJ, Ackerson LM, Ferrara A, Liu J: **Developing a prediction rule from automated clinical databases to identify high-risk patients in a large population with diabetes.** *Diabetes Care* 2001, **24**:1547–1555.
- Stroebel RJ, Scheitel SM, Fitz JS, Herman RA, Naessens JM, Scott CG, Zill DA, Muller L: **A randomized trial of three diabetes registry implementation strategies in a community internal medicine practice.** *Jt Comm J Qual Improv* 2002, **28**:441–450.
- Wilke RA, Berg RL, Peissig P, Kitchner T, Sijercic B, McCarty CA, McCarty DJ: **Use of an electronic medical record for the identification of research subjects with diabetes mellitus.** *Clin Med Res* 2007, **5**:1–7.
- Wilson C, Susan L, Lynch A, Saria R, Peterson D: **Patients with diagnosed diabetes mellitus can be accurately identified in an indian health service patient registration database.** *Public Health Rep* 2001, **116**:45–50.
- Zgibor JC, Orchard TJ, Saul M, Piatt G, Ruppert K, Stewart A, Siminerio LM: **Developing and validating a diabetes database in a large health system.** *Diabetes Res Clin Pract* 2007, **75**:313–319.
- Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA: **10-year follow-up of intensive glucose control in type 2 diabetes.** *N Engl J Med* 2008, **359**:1577–1589.
- Gabbay RA, Khan L, Peterson KL: **Critical features for a successful implementation of a diabetes registry.** *Diabetes Technol Ther* 2005, **7**:958–967.
- Saydah SH, Geiss LS, Tierney E, Benjamin SM, Engelgau M, Brancati F: **Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data.** *Ann Epidemiol* 2004, **14**:507–516.
- Samuels TA, Cohen D, Brancati FL, Coresh J, Kao WH: **Delayed diagnosis of incident type 2 diabetes mellitus in the ARIC study.** *Am J Manag Care* 2006, **12**:717–724.
- Fisher-Hoch SP, Vatcheva KP, Laing ST, Hossain MM, Rahbar MH, Hanis CL, Brown HS 3rd, Rentfro AR, Reiningger BM, McCormick JB: **Missed opportunities for diagnosis and treatment of diabetes, hypertension, and hypercholesterolemia in a Mexican American population, Cameron County hispanic cohort, 2003–2008.** *Prev Chronic Dis* 2012, **9**:110298.
- Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K: **Incorrect and incomplete coding and classification of diabetes: a systematic review.** *Diabet Med* 2010, **27**:491–497.
- Kudyakov R, Bowen J, Ewen E, West SL, Daoud Y, Fleming N, Masica A: **Electronic health record use to classify patients with newly diagnosed versus preexisting type 2 diabetes: infrastructure for comparative effectiveness research and population health management.** *Popul Health Manag* 2012, **15**:3–11.

doi:10.1186/1472-6947-13-81

**Cite this article as:** Makam et al.: Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. *BMC Medical Informatics and Decision Making* 2013 **13**:81.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

