Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2018 October ; 27(10): 1053–1059. doi:10.1002/pds.4377.

Validation of an Algorithm for Identifying Type 1 Diabetes in Adults Based on Electronic Health Record Data

Emily B Schroeder, MD, PhD^{1,2,3}, W Troy Donahoo, MD^{2,3,*}, Glenn K Goodrich, MS¹, and Marsha A Raebel, PharmD^{1,4}

¹Institute for Health Research, Kaiser Permanente Colorado, Denver, Colorado

²Division of Endocrinology, Colorado Permanente Medical Group, Denver, Colorado

³Division of Endocrinology, Metabolism and Diabetes, University of Colorado Denver, Aurora, Colorado

⁴Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Denver, Aurora, Colorado

Abstract

Purpose—Algorithms using information from electronic health records (EHR) to identify adults with type 1 diabetes have not been well studied. Such algorithms would have applications in pharmacoepidemiology, drug safety research, clinical trials, surveillance, and quality improvement. Our main objectives were to determine the positive predictive value for identifying type 1 diabetes in adults using a published algorithm (developed by Klompas et al), and to compare it to a simple requirement that the majority of diabetes diagnosis codes be type 1.

Methods—We applied the Klompas algorithm and the diagnosis code criterion to a cohort of 66,690 adult Kaiser Permanente Colorado members with diabetes. We reviewed 220 charts of those identified as having type 1 diabetes, and calculated positive predictive values.

Results—The Klompas algorithm identified 3,286 (4.9% of 66,690) adults with diabetes as having type 1 diabetes. Based on chart reviews, the overall positive predictive value was 94.5%. The requirement that the majority of diabetes diagnosis codes be type 1 identified 3,000 (4.5%) as having type 1 diabetes, and had a positive predictive value of 96.4%. However, the algorithm criterion involving dispensing of urine acetone test strips performed poorly, with a positive predictive value of 20.0%.

Conclusions—Data from EHRs can be used to accurately identify adults with type 1 diabetes. When identifying adults with type 1 diabetes, we recommend either a modified version of the Klompas algorithm without the urine acetone test strips criterion, or the requirement that the majority of diabetes diagnosis codes be type 1 codes.

Corresponding Author: Emily B Schroeder. Address: 10065 E Harvard Ave, Suite 300, Denver, CO, 80231, Phone: 303-614-1396, Fax: 303-614-1395, Emily.B.Schroeder@kp.org.

*Now at: Division of Endocrinology, Diabetes and Metabolism, University of Florida, Gainesville, Florida

Prior Presentations

Part of this manuscript was presented as a poster at the American Diabetes Association Scientific Sessions on June 11, 2017.

Keywords

Diabetes mellitus; diabetes mellitus; type 1; adult; electronic health records

INTRODUCTION

While algorithms using information from electronic health records (EHR) that can identify individuals with diabetes have been well studied,^{1–4} there has been little development of similar algorithms to distinguish adults with type 1 diabetes from adults with type 2 diabetes. Such algorithms could be used in in pharmacoepidemiology, drug safety research, clinical trial recruitment, surveillance, and quality improvement. For example, while the prevalence and incidence of type 1 diabetes in children has been well established through the SEARCH for Diabetes in Youth (SEARCH) study,^{5,6} the prevalence of type 1 diabetes in adults is largely based on data from the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS), which are based on current insulin use and age of diagnosis.⁷ However, an insulin use and age-based definition misclassifies adults with late-onset type 1 diabetes and individuals with early-onset type 2 diabetes who are quickly started on insulin, and cannot be applied in most studies involving EHR data, where the age of diagnosis is usually unknown for prevalent cases of diabetes (or, if known, is documented in an EHR location that is not readily extractable into an electronic database). As a result, a 2015 workshop sponsored by the NIH Diabetes Mellitus Interagency Coordinating Committee identified the “development and testing of a computable phenotype to identify people with type 1 diabetes in the electronic medical records” as an important priority area.⁸

While algorithms relying solely on type 1 and type 2 International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnostic codes are appealing, there is concern that inconsistent coding by health care providers could introduce errors. Recent reviews of the literature conducted as part of the FDA-sponsored Mini-Sentinel pilot program^{9,10} identified only two publications that presented algorithms designed to distinguish type 1 diabetes from type 2 diabetes in adults.^{11,12} The Lo-Ciganic algorithm is based on inpatient data elements.¹² We therefore we focused on the algorithm developed by Klompas et al.¹¹

The Klompas algorithm was developed in a single health care setting in Eastern Massachusetts.¹¹ They identified an algorithm for distinguishing type 1 diabetes from type 2 diabetes based on 210 charts of individuals with diabetes. The algorithm was then validated on a sample of 100 charts from the same health care setting. The algorithm has not been externally validated. Candidate predictors in this algorithm included ICD-9 diagnosis codes, age, body mass index, laboratory test results (triglycerides, HDL, C-peptide, diabetes autoantibodies), and medications (diabetes medications, glucagon, and urine acetone test strips). Their final recommended algorithm included criteria based on diagnosis codes, medications, and laboratory test result values (C-peptide and diabetes autoantibodies). Their goal was to optimize sensitivity for type 1 diabetes while maintaining a high positive

predictive value (PPV). The recommended algorithm had a sensitivity of 100% and PPV of 96% in their derivation sample, and a sensitivity of 65% and PPV of 88% in their validation sample.

The main aim of the current study was to determine the PPV of the Klompas algorithm for identifying type 1 diabetes in a cohort of adults with diabetes at a separate site from where the algorithm was initially developed. As a secondary aim, we examined the difference between the Klompas algorithm and use of ICD-9 codes alone. We also updated the Klompas algorithm to include ICD-10 codes.

METHODS

Study design, setting, and data sources

Kaiser Permanente Colorado (KPCO) is an integrated healthcare delivery system that serves approximately 550,000 members in the Denver-Boulder metropolitan area. KPCO is a member of the Health Care Systems Research Network (HCSRN, formerly the HMO Research Network). Research institutions embedded in the health systems of the HCSRN have developed a distributed virtual data warehouse (VDW) that includes content areas such as demographics, outpatient pharmacy dispensing, laboratory tests and laboratory test results, and diagnosis and procedure codes from outpatient and inpatient health care encounters from their EHR and other clinical and administrative data system.¹³ For this project, we employed the published SUPREME-DM DataLink diabetes registry criteria for diabetes identification.^{3,4} We considered diabetes identification as the earlier of one inpatient diagnosis (ICD-9-CM 250.x, 357.2, 366.41, 362.01-362.07, either primary or secondary) or any combination of two of the following events: 1) HbA1c $\geq 6.5\%$; 2) fasting plasma glucose $\geq 126\text{mg/dl}$; 3) random plasma glucose $\geq 200\text{mg/dl}$; 4) outpatient diagnosis code (same codes as for inpatient); 5) any anti-hyperglycemic medication dispensing. When the two events were from the same source (e.g., two outpatient diagnoses or two elevated laboratory result values), we required them to occur on separate days no more than two years apart. Two dispensings of metformin or thiazolidinediones with no other indication in the database of diabetes were not included because these agents could be used for other conditions. Criteria ascertained during periods of pregnancy were excluded.

In this analysis, we included individuals identified with diabetes using the SUPREME-DM criteria who had at least one day of enrollment between 1/1/2006 and 9/30/2015 and were at least 20 years old (the age traditionally used by the Centers for Disease Control and Prevention to differentiate adults and youth with diabetes).¹⁴

Klompas algorithm

To identify individuals meeting the Klompas algorithm criteria for type 1 diabetes, we used all diagnosis codes, medication dispensings, and laboratory results during the study period (1/1/2006 through 9/30/2015). Individuals met the Klompas algorithm if they fulfilled any of the following four criteria:

1. ICD-Coded diagnoses and glucose medication: Over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3) were type 1 codes (ICD-9

250.x1, 250.x3), AND no dispensing for a non-insulin antidiabetic drug (excluding metformin)

2. Coded diagnoses and glucagon: Over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3) were type 1 codes (ICD-9 250.x1, 250.x3), AND a dispensing for glucagon
3. Urine test strips: Dispensing of urine acetone test strips
4. Labs: Negative C-peptide result or positive diabetes autoantibody result.

We also identified all individuals with over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3) as type 1 codes (ICD-9 250.x1, 250.x3). These individuals are described as meeting the ICD-9 criterion.

Chart review and gold standard

From the KPCO cohort of adults meeting any of the four criteria of the Klompas algorithm, we randomly selected 150 charts for review. We additionally oversampled individuals meetings criteria 2–4 in a hierarchical manner, such that each of the following groups had 20 charts for review:

1. Individuals who did not meet criterion 1 and did meet criterion 2
2. Individuals who did not meet criteria 1 and 2 and did meet criterion 3
3. Individuals who did not meet criteria 1, 2, and 3 and did meet criterion 4

We also reviewed all charts (n=26) of individuals who met the ICD-9 criterion and who did not meet any Klompas algorithm criteria.

The gold standard was based on chart review. We followed the same methods as Klompas to identify the “true” diabetes type. The following “rules” were applied sequentially: endocrinology provider diagnosis if available (type 1 or 2), never on insulin (type 2), C-peptide negative or diabetes autoantibodies present (type 1), currently on insulin but prior history of prolonged treatment with oral hypoglycemic alone (type 2), and nonendocrinologist provider diagnosis (type 1 or 2).

Chart review was conducted by authors W.T.D., M.A.R., and E.B.S (two endocrinologists and a pharmacist/pharmacoepidemiologist, all with extensive knowledge of diabetes). Each chart was reviewed by 2 members of the team, with any discrepancies resolved by the third team member.

ICD-10 codes

The United States transitioned to ICD-10 codes on 10/1/2015. To assess the performance of the Klompas algorithm updated to include ICD-10 codes, we identified a cohort of individuals with diabetes who had continuous enrollment for the 9 months before and 9 months after 10/1/2015. We then compared use of the ICD-9 codes in the 9 months before and after 10/1/2015. Type 1 diabetes codes were considered to be: ICD-9 250.x1, ICD-9 250.x3, and ICD-10 E10.xx. Type 2 diabetes codes were considered to be: ICD-9 250.x0, ICD-9 250.x2, and ICD-10 E11.xx. We also examined the use of ICD-10 codes for

secondary types of diabetes (E08.xx = diabetes mellitus due to underlying condition; E09.xx = drug or chemical induced diabetes mellitus; E13.xx = other specified diabetes mellitus).

Statistical methods

To obtain PPVs, we weighted the cases based on our sampling framework. For each sample we considered that entire sample as the numerator. The denominator was the number of charts in that sample that were reviewed. We then performed a domain analysis (Proc Survey means) in SAS® version 9.4 (SAS Institute Inc., Cary, NC), which allowed us to take into account the sampling while also simultaneously performing multiple comparisons.¹⁵

This study was approved by the Kaiser Permanente Colorado Institutional Review Board (IRB). The requirement for informed consent was waived.

RESULTS

Our cohort contained 66,690 adults with diabetes. Using the Klompas algorithm, 3,286 (4.9% of 66,690) individuals were identified as having type 1 diabetes. Using ICD-9 criterion alone identified 3,000 (4.5% of 66,690) individuals as having type 1 diabetes. Individuals identified as having type 1 diabetes were younger, were less likely to be Hispanic, and had a lower body mass index than the entire diabetes cohort (Table 1). The subset of 220 charts included in the chart review had similar characteristics as the full type 1 diabetes cohorts.

The PPV for the Klompas algorithm criteria 1–4 overall was 94.5% (Table 2). We examined the performance of the 4 Klompas criteria in a hierarchical manner (Table 2). All the criteria had very high PPV (>88%), except for urine test strips (Klompas algorithm criterion 3). Individuals who met the Klompas algorithm urine test strip criteria but did not meet the Klompas algorithm ICD-9 and glucose medication or ICD-9 and glucagon criteria (criteria 1 and 2) had a PPV of 71.2%. Further examination showed that individuals who only met the Klompas algorithm urine test strip criterion had a PPV of 20%, while those who met the Klompas algorithm urine acetone test strip criterion and the Klompas algorithm lab criteria (but not Klompas algorithm ICD-9 based criteria 1 or 2) had a PPV of 95.3%.

We reviewed 159 charts out of 3,000 individuals meeting the ICD-9 criterion (majority of diabetes codes being type 1 diabetes codes), and 154 were found to have type 1 diabetes, with a PPV 96.4% (95% CI 93.5, 99.4). We reviewed all 26 charts for the individuals who met the ICD-9 criterion only (and did not meet any of the Klompas algorithm criteria), and 19 were found to have type 1 diabetes, with a PPV of 73.1%. For the chart reviews based upon the Klompas algorithm or the ICD-9 criterion, we relied upon a non-endocrinologist provider diagnosis in 11% of charts after appropriate weighting (12% unweighted).

Table 3 shows the estimated prevalence of type 1 diabetes among adults with diabetes using several permutations of the Klompas algorithm criteria, as well as using ICD-9 codes alone. The PPV ranged from 94.5% to 96.4%, and the prevalence from 4.4% to 4.9%. Including a lab results criterion (criterion 4 in the Klompas algorithm) increases the estimated prevalence of type 1 diabetes among adults above ICD-9 criteria based prevalence estimates

with little change in PPV (decrease from 96.4% to 94.5% or 95.1%, depending on the specific criteria met).

As shown in Table 4, the conversion between the ICD-9 and ICD-10 codes was quite good. Among individuals with a majority of type 1 ICD-9 codes in the 9 months prior to the ICD-10 transition, 92.4% had a majority of type 1 ICD-10 codes in the 9 months after the transition, and only 2.5% had a majority of type 2 ICD-10 codes in the 9 months after the transition (5.1% had no diabetes codes in that 9 month period). Among individuals with a majority of type 2 ICD-9 codes in the 9 months prior to the ICD-10 transition, only 1.1% had a majority of type 1 ICD-10 codes in the 9 months after the transition. Among individuals who met the Klompas algorithm criteria for type 1 diabetes prior to the ICD-10 transition, 85.9% had a majority of type 1 ICD-10 codes in the 9 months after the transition. Less than 1% (288 out of 36,861) had any secondary diabetes ICD-10 codes (E08, E09, E13).

DISCUSSION

We conducted an external validation study of a published algorithm to identify adults with type 1 diabetes using data elements commonly available in EHRs. For the published Klompas algorithm, we found a PPV of 94.5% and an estimated 4.9% prevalence of type 1 diabetes among adults with diabetes. A simpler criterion that only required a majority of diabetes diagnosis codes to be type 1 diabetes codes had a PPV of 96.4% and an estimated prevalence of type 1 diabetes of 4.5%.

Our estimate of the prevalence of type 1 diabetes in adults corresponds closely to the widely applied “rule of thumb” that approximately 5% of the adult American population with diagnosed diabetes has type 1 diabetes.^{14,16,17} This estimate is largely based on NHANES/NHIS data that considers only current insulin use and age of diagnosis,⁷ which misclassifies adults with late-onset type 1 diabetes and individuals with early-onset type 2 diabetes who are quickly started on insulin, and cannot be applied in most EHR systems where age of diagnosis is not usually available.

Most published diabetes algorithms for adults have been developed to either identify any diabetes or type 2 diabetes and were not designed to identify type 1 diabetes.^{1,2,18–21} The Klompas algorithm, which we considered here, had a sensitivity of 100% and PPV of 96% in their original derivation sample, and a sensitivity of 65% and PPV of 88% in their internal validation sample. The lower sensitivity in their validation dataset was due to the misclassification of a single individual.¹¹ We did find one other type 1 diabetes algorithm for use in adults, but it required inpatient data elements,¹² and was not applicable to adults with diabetes who had not been hospitalized. Others have developed algorithms for use in children and young adults.^{21,22} One of these algorithms was based on prescriptions (at least one prescription for insulin with at most one oral hypoglycemic prescription),²² while the other was based on diagnosis codes.²¹ In addition, the SEARCH study has published algorithms to identify children with type 1 diabetes.^{5,6} Zhong and colleagues considered billing data, patient problem lists, laboratory test results, and diabetes related medications, and found that the criterion of the majority of diabetes billing codes being type 1 codes

performed the best (sensitivity of 95.6% and PPV of 97.5%), but the authors did note considerable age and racial/ethnic differences.⁶ Lawrence and colleagues considered diagnosis codes, laboratory test results (A1c, glucose), and medication dispensings. They recommended the criteria of 1 or more outpatient type 1 diabetes diagnosis or 1 or more insulin dispensings, which had a sensitivity of 95.9% and a PPV of 95.5%.⁵

Strengths of this study include the large cohort size, use of a robust, well-curated electronic data warehouse and a previously developed diabetes registry, a gold standard based on chart reviews, an independent validation setting, and a comparison of different potential algorithm definitions.

An important limitation of this study is that we did defer to the clinical assessment found in the chart. We prioritized assessments by the endocrinology provider, and then considered medication use and lab result values. However, in 11% of the charts reviewed, the individual had not been seen by an endocrinology provider, and the diagnosis of type 1 or type 2 diabetes was still not clear after review of medication use and lab result values. In those cases, we used the documentation from the most recent non-endocrinologist provider encounter. At times, this documentation was simply reflected in the diagnosis code chosen, without elaboration in the chart. Although this approach reflects the best available information, this methodology will result in the criteria involving diagnosis codes appearing more accurate than they actually are.

In addition, the generalizability of our results depends on coding practices in different settings, laboratory test ordering and the availability of laboratory test result values in electronic databases, as well as healthcare practices regarding prescription of glucagon and acetone test strips. In particular, we found the dispensing of urine test strips to be low in our system. We did not think that criterion alone had sufficient face validity to recommend including in the algorithm. As a result, we do not recommend use of this component of the algorithm. However, it should be noted that such post hoc alterations to the algorithm may potentially lead to overfitting. Further, C-peptide and diabetes autoantibody result values are not currently available in a standardized, easily extractable form in many EHRs or in the laboratory result tables of distributed databases such as the FDA Sentinel Distributed Database and PCORnet.^{23,24}

This algorithm has a very high PPV for identifying type 1 diabetes. Because we did not perform chart reviews of individuals classified by the algorithm as not having type 1 diabetes, we were unable to determine sensitivity of the algorithm, and thus the algorithm may not correctly identify all individuals with type 1 diabetes. Additional validation studies could be done to accurately determine the negative predictive value, sensitivity, and specificity of this algorithm.

When identifying type 1 diabetes in adults in settings in which lab result values for C-peptide and diabetes autoantibodies are not available (or are not electronically accessible), we recommend that either the first two criteria of the Klompas algorithm be used (the ICD code and glucose medication criterion and the ICD code and glucagon criterion), or the simple criterion of a majority of diabetes diagnosis codes being type 1 codes. In our cohort.

these options performed very similarly, with both having PPV of 96.4% and prevalence of 4.4 and 4.5%, respectively. The simple criterion of a majority of diabetes diagnosis codes being type 1 codes is easy to implement and only requires knowledge of codes. However, it has the drawback of being more sensitive to coding practices and coding errors.

In setting where C-peptide and diabetes autoantibodies lab results values are available, we recommend using the Klompas algorithm without the urine test strips criterion to identify type 1 diabetes in adults:

1. Over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3; or ICD-10 E9.xx, E10.xx) were type 1 codes (ICD-9 250.x1, 250.x3, or ICD-10 E10.xx), AND no dispensing for a non-insulin antidiabetic drug (excluding metformin)
2. Over 50% of diabetes codes were type 1 codes (same codes as in #1), AND a dispensing for glucagon
3. Negative C-peptide result or positive diabetes autoantibodies lab test result.

In conclusion, we have confirmed the external validity and usefulness of three of four criteria in a published algorithm for identifying type 1 diabetes in adults. Further, we have demonstrated the utility and performance of using only coded diagnoses for identifying type 1 diabetes in adults. Data from EHRs can be used to accurately identify adults with type 1 diabetes without the need for chart review. These criteria could be used to identify type 1 diabetes for purposes of cohort identification, confounder or covariate definition, or as an outcome in safety surveillance studies. They thus have important potential applications in pharmacoepidemiology, drug safety research, clinical trial recruitment, surveillance, and quality improvement.

Acknowledgments

Dr. Schroeder was supported by grant number 1K23DK099237-01 from the National Institute of Diabetes and Digestive and Kidney Diseases, and the project was supported by an internal grant from Kaiser Permanente Colorado.

References

1. Zgibor JC, Orchard TJ, Saul M, et al. Developing and validating a diabetes database in a large health system. *Diabetes Res Clin Pract.* 2007; 75:313–319. [PubMed: 16934906]
2. Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroschowski MC, O'Connor PJ. Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. *Am J Med Qual.* 2006; 21:238–245. [PubMed: 16849780]
3. Nichols GA, Desai J, Elston LJ, et al. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM Project. *Prev Chronic Dis.* 2012; 9:E110. [PubMed: 22677160]
4. Nichols GA, Schroeder EB, Karter AJ, et al. Trends in diabetes incidence among 7 million insured adults, 2006–2011: the SUPREME-DM Project. *Am J Epidemiol.* 2015; 181:32–39. [PubMed: 25515167]
5. Lawrence JM, Black MH, Zhang JL, et al. Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. *Am J Epidemiol.* 2014; 179:27–38. [PubMed: 24100956]

6. Zhong VW, Pfaff ER, Beavers J, et al. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes care ascertainment and type classification: the SEARCH for Diabetes in Youth Study. *Pediatric Diabetes*. 2014; 15:573–584. [PubMed: 24913103]
7. Menke A, Orchard TJ, et al. Imperatore G. The prevalence of type 1 diabetes in the United States. *Epidemiology*. 2013; 24:773–774. [PubMed: 23903880]
8. National Institutes of Health. [Accessed on 8 November 2016] Diabetes Mellitus Interagency Coordinating Committee (DMICC) workshop on research supported by the Special Statutory Funding Program for Type 1 Diabetes Research. 2015. Available at: <https://www.niddk.nih.gov/about-niddk/advisory-coordinating-committees/diabetes-mellitus-interagency-coordinating-committee/meeting-agendas-summaries-presentations/Pages/dmicc-reports.aspx>
9. Leonard CE, Freeman CP, Razzaghi H. , et al. [Accessed on 8 November 2016] 15 cohorts of interest for surveillance preparedness. 2014. Available at: <https://www.sentinelssystem.org/sentinel/methods/332>
10. Raebel MA, Schroeder EB, Goodrich G. , et al. [Accessed on 8 November 2016] Validating type 1 and type 2 diabetes mellitus in the Mini-Sentinel Distributed Database using the Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) DataLink. 2016. Available at: <https://www.sentinelssystem.org/sentinel/methods/validating-type-1-and-type-2-diabetes-mellitus-mini-sentinel-distributed-database>
11. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care*. 2013; 36:914–921. [PubMed: 23193215]
12. Lo-Ciganic W, Zgibor JC, Ruppert K, Arena VC, Stone RA. Identifying type 1 and type 2 diabetic cases using administrative data: a tree-structured model. *J Diabetes Sci Technol*. 2011; 5:486–493. [PubMed: 21722564]
13. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: a public data model to support collaboration. *eGEMs*. 2014; 2:1049. [PubMed: 25848584]
14. Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States. Atlanta, GA: US Department of Health and Human Services; 2014. 2014
15. Lohr S. Sampling: Design and Analysis. 2. Boston, MA: Brooks/Cole, Cengage Learning; 2010.
16. American Diabetes Association. Standards of medical care in diabetes - 2017. *Diabetes Care*. 2017; 40(Suppl 1):S1–S135. [PubMed: 27979885]
17. JDRF. [Accessed 29 August 2016] Type 1 diabetes facts. Available from: <http://www.jdrf.org/about/fact-sheets/type-1-diabetes-facts>
18. Holt TA, Gunnarsson CL, Cload PA, Ross SD. Identification of undiagnosed diabetes and quality of diabetes care in the United States: cross-sectional study of 11.5 million primary care electronic records. *CMAJ Open*. 2014; 2:E248–255.
19. O'Connor PJ, Rush WA, Pronk NP, Cherney LM. Identifying diabetes mellitus or heart disease among health maintenance organization members: sensitivity, specificity, predictive value, and cost of survey and database methods. *Am J Manag Care*. 1998; 4:335–342. [PubMed: 10178496]
20. Pacheco JA, Thompson W. [Accessed on 25 October 2016] Type 2 diabetes mellitus electronic medical record case and control selection algorithms. 2011. Available from: <https://phekb.org/sites/phenotype/files/T2DM-algorithm.pdf>
21. Rhodes ET, Laffel LM, Gonzalez TV, Ludwig DS. Accuracy of administrative coding for type 2 diabetes in children, adolescents, and young adults. *Diabetes Care*. 2007; 30:141–143. [PubMed: 17192348]
22. Bobo WV, Cooper WO, Stein CM, et al. Positive predictive value of a case definition for diabetes mellitus using automated administrative health data in children and youth exposed to antipsychotic drugs or control medications: a Tennessee Medicaid study. *BMC Med Res Methodol*. 2012; 12:128. [PubMed: 22920280]
23. Corley DA, Feigelson HS, Lieu TA, McGlynn EA. Building data infrastructure to evaluate and improve quality: PCORnet. *J Oncol Pract*. 2015; 11:204–206. [PubMed: 25980016]

24. Brown J, Beaulieu N, Curtis L, Raebel MA, Haynes K. [Accessed on 25 October 2016] Sentinel common data model V6.0. 2016. Available from: <https://www.sentinel-system.org/sentinel/data/distributed-database-common-data-model/106>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Key Points

- Algorithms using information from electronic health records (EHRs) to identify adults with type 1 diabetes have not been well studied. Such algorithms have applications in pharmacoepidemiology, drug safety research, clinical trials, surveillance, and quality improvement.
- Using a cohort of 66,690 adults with diabetes, we determined the positive predictive value (PPV) for identifying type 1 diabetes in adults using a published algorithm (developed by Klompas et al).
- Both the Klompas algorithm and the requirement that the majority of diabetes diagnosis codes be for type 1 diabetes performed very well, with PPVs of 94.5% and 96.4%, respectively.
- Data from EHRs can be used to accurately identify adults with type 1 diabetes.

Table 1

Demographics of 66,690 adult members of Kaiser Permanente Colorado with diabetes mellitus, 1/1/2006 and 9/30/2015.

	Full diabetes cohort N=66,690	Type 1 diabetes identified through the Klompas algorithm N=3,286 [†]	Type 1 diabetes identified through the ICD-9 criterion N=3,000 [‡]	Chart review sample N=220
Age, years (mean, SD)	58.8 (13.7)	42.4 (14.4)	41.5 (14.2)	45.3 (14.5)
Female	31,999 (48.0%)	1,572 (47.8%)	1,425 (47.5%)	106 (48.2%)
Race				
Hispanic	12,229 (18.3%)	278 (8.5%)	234 (7.8%)	20 (9.1%)
NH Black	4,206 (6.3%)	127 (3.9%)	101 (3.4%)	11 (5.0%)
NH White	37,903 (56.8%)	2,198 (66.9%)	2,005 (66.8%)	154 (70.0%)
Other	4,373 (6.6%)	118 (3.6%)	108 (3.6%)	8 (3.6%)
Missing	7,979 (12.0%)	565 (17.2%)	552 (18.4%)	27 (12.3%)
BMI, kg/m ² (mean, SD)	32.5 (7.5)	26.9 (5.4)	26.7 (5.2)	27.4 (6.0)
Length of follow-up after diabetes recognition, years (mean, SD)	4.1 (3.3)	4.0 (3.4)	3.9 (3.4)	4.3 (3.5)
≥ 1 type 1 ICD-9 code	7399 (11.1%)	3176 (96.7%)	3,000 (100%)	210 (95.5%)
> 50% of diabetes codes as type 1 codes (or no type 2 codes)	3012 (4.5%)	2974 (90.5%)	3,000 (100%)	185 (84.1%)
No record of oral hypoglycemic other than metformin	25492 (38.2%)	2910 (88.6%)	2721 (90.7%)	145 (65.9%)
Glucagon dispensing	1515 (2.3%)	1126 (34.3%)	1070 (35.7%)	72 (32.7%)
Urine acetone strip dispensing	596 (0.9%)	594 (18.1%)	544 (18.1%)	58 (26.4%)
C-peptide negative and/or diabetes autoantibody positive	934 (1.4%)	929 (28.3%)	650 (21.7%)	87 (39.5%)

Mean;SD for continuous variables. N;% for categorical variables.

[†]The Klompas algorithm is fulfilled if any of the following four criteria are met: 1) Over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3) were type 1 codes (ICD-9 250.x1, 250.x3), AND no dispensing for a non-insulin antidiabetic drug (excluding metformin); 2) Over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3) were type 1 codes (ICD-9 250.x1, 250.x3), AND a dispensing for glucagon; 3) Dispensing of urine acetone test strips; 4) Negative C-peptide result or positive diabetes autoantibody result.

[‡]The ICD-9 criterion is fulfilled if over 50% of diabetes codes (ICD-9 250.x0, 250.x1, 250.x2, and ICD-9 250.x3) are type 1 codes (ICD-9 250.x1, 250.x3).

Table 2

Positive predictive value for Klompas criteria for identification of type 1 diabetes in adults

Group	Number	Number of charts reviewed	Number confirmed as type 1 diabetes	Positive predictive value (95% CI) [†]
Klompas algorithm criteria 1 – 4 overall	3,286	194	177	94.5 (91.3, 97.6)
Met criterion 1 [ICD-9 codes and glucose lowering meds]	2904	134	129	96.3% (93.1, 99.4)
Met criterion 2 [ICD-9 codes and glucagon], did not meet criterion 1	42	20	20	100
Met criteria 3 [urine test strips] and 4 [labs], did not meet criterion 1 or 2 [‡]	21	10	9	95.3% (85.8, 100.0)
Met criterion 3 [urine test strips], did not meet criterion 1, 2, or 4 [‡]	34	10	2	20.0% (0, 44.2)
Met criterion 4 [labs], did not meet criterion 1 or 2 or 3	285	20	17	88.9 (76.9, 100)

[†] Positive predictive value does not equal number confirmed as type 1 diabetes/number of charts reviewed due to sampling weights.

[‡] These 55 individuals met criterion 3 (urine test strips), with or without meeting criterion 4. Across these 55 individuals the PPV was 72.1 (95% CI 49.5, 92.9).

Table 3

Prevalence and positive predictive values for methods of identifying type 1 diabetes in adults

	Number	Prevalence	Positive predictive value
Majority type 1 ICD-9 codes	3,000	4.5%	96.4
Klompas algorithm (criteria 1 and 2)	2,946	4.4%	96.4
Klompas algorithm (criteria 1, 2, and 4)	3,252	4.9%	95.1
Klompas algorithm (criteria 1–4)	3,286	4.9%	94.5

Table 4

Agreement between ICD-9 and ICD-10 diagnosis codes for type 1 and type 2 diabetes

	N=36,861				N=36,861	
	> 50% type 1 ICD-9 codes N=1,269	>= 50% type 2 ICD-9 codes N=30,373	No type 1 or type 2 ICD-9 codes N=5,219	Type 1 diabetes per Klompas algorithm N=1514	Not type 1 diabetes per Klompas algorithm N=35,347	
At least one type 1 ICD-10 code	1,190 (93.8%)	321 (1.1%)	60 (1.1%)	1,292 (85.3%)	279 (0.8%)	
At least one type 2 ICD-10 code	247 (19.5%)	26,011 (85.6%)	2,620 (50.2%)	368 (24.3%)	28,510 (80.7%)	
At least one "other" diabetes ICD-10 code	26 (2.0%)	241 (0.8%)	21 (0.4%)	35 (2.3%)	253 (0.7%)	
>50% type 1 ICD-10 codes	1,172 (92.4%)	93 (0.3%)	50 (1.0%)	1,255 (85.9%)	60 (0.2%)	
>=50% type 2 ICD-10 codes	32 (2.5%)	25,971 (85.5%)	2,614 (50.0%)	130 (8.6%)	28,487 (80.6%)	
No type 1 or type 2 ICD-10 codes	65 (5.1%)	4,309 (14.2%)	2,555 (49.0%)	129 (8.5%)	6,800 (19.2%)	

Type 1 ICD-9 codes = 250.x1, 250.x3. Type 2 ICD-9 codes = 250.x0, 250.x2.

Type 1 ICD-10 codes = E10.xx. Type 2 ICD-10 codes = E11.xx.

"Other" diabetes ICD-10 codes = E08.xx, E09.xx, E13.xx.

Percentage of type 1 codes was calculated as: # type 1 codes / (# type 1 codes + # type 2 codes).

Percentage of type 2 codes was calculated as: # type 2 codes / (# type 1 codes + # type 2 codes).