Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Pediatr Diabetes. Author manuscript; available in PMC 2015 December 01.

Published in final edited form as:

Pediatr Diabetes. 2014 December ; 15(8): 573–584. doi:10.1111/pedi.12152.

Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study

Victor W. Zhong[a], **Emily R. Pfaff**[b], **Daniel P. Beavers**[c], **Joan Thomas**[a], **Lindsay M. Jaacks**[a], **Deborah A. Bowlby**[d], **Timothy S. Carey**[j], **Jean M. Lawrence**[e], **Dana Dabelea**[f], **Richard F. Hamman**[f], **Catherine Pihoker**[g], **Sharon H. Saydah**[h], and **Elizabeth J. Mayer-Davis**[a,i] **FOR THE SEARCH FOR DIABETES IN YOUTH STUDY GROUP**

[a]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

[b]North Carolina TraCS Institute, University of North Carolina, Chapel Hill, North Carolina, USA

[c]Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

[d]Division of Pediatric Endocrinology, Medical University of South Carolina, Charleston, South Carolina, USA

[e]Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, California, USA

[f]Department of Epidemiology, Colorado School of Public Health, University of Colorado, Denver, Aurora, Colorado, USA

[g]Department of Washington, University of Washington, Seattle, Washington, USA

[h]Centers for Disease Control and Prevention, Division of Diabetes Translation, Atlanta, Georgia, USA

[i]Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

[j]Cecil G. Sheps Center for Health Services Research, University of North Carolina, Chapel Hill, NC, USA

Author contributions: V.W.Z. participated in study design and coordination, reviewed medical records, analyzed data and wrote the manuscript. E.J.M.-D. conceived the study, participated in study design, reviewed and edited the manuscript and contributed to discussion. E.R.P. extracted and formatted the raw data from the CDW-H, reviewed and edited the manuscript and contributed to discussion. J.T. participated in study design and coordination, trained medical record reviewers, validated the record review work, reviewed and edited the manuscript. L.M.J. participated in study design and coordination, reviewed medical records, reviewed and edited the manuscript and contributed to discussion. D.P.B., D.A.B., D.D., R.F.H., J.M.L., C.P., and S.H.S. reviewed and edited the manuscript and contributed to discussion. V.W.Z. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

No potential conflicts of interest relevant to this article were reported.

This study was presented in part as two posters at the 73rd Scientific Session of the American Diabetes Association, Chicago, IL, June 21-25, 2013.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Abstract

Background—The performance of automated algorithms for childhood diabetes case ascertainment and type classification may differ by demographic characteristics.

Objective—This study evaluated the potential of administrative and electronic health record (EHR) data from a large academic care delivery system to conduct diabetes case ascertainment in youth according to type, age and race/ethnicity.

Subjects—57,767 children aged <20 years as of December 31, 2011 seen at University of North Carolina Health Care System in 2011 were included.

Methods—Using an initial algorithm including billing data, patient problem lists, laboratory test results and diabetes related medications between July 1, 2008 and December 31, 2011, presumptive cases were identified and validated by chart review. More refined algorithms were evaluated by type (type 1 versus type 2), age (<10 versus ≥10 years) and race/ethnicity (non-Hispanic white versus "other"). Sensitivity, specificity and positive predictive value were calculated and compared.

Results—The best algorithm for ascertainment of diabetes cases overall was billing data. The best type 1 algorithm was the ratio of the number of type 1 billing codes to the sum of type 1 and type 2 billing codes ≥0.5. A useful algorithm to ascertain type 2 youth with "other" race/ethnicity was identified. Considerable age and racial/ethnic differences were present in type-non-specific and type 2 algorithms.

Conclusions—Administrative and EHR data may be used to identify cases of childhood diabetes (any type), and to identify type 1 cases. The performance of type 2 case ascertainment algorithms differed substantially by race/ethnicity.

Keywords

childhood diabetes; case ascertainment; type classification; electronic health record; administrative data

---

Ongoing surveillance of childhood diabetes in the U.S. is needed to understand the trends in incidence and prevalence, and to anticipate health care delivery needs. The SEARCH for Diabetes in Youth Study (SEARCH) (1) documented an increase in the prevalence of type 1 and type 2 diabetes from 2001 to 2009 (2, 3). From 2010 to 2050, the number of youth with type 1 and type 2 diabetes is projected to increase by another 23% and 49%, respectively, even assuming no change in incidence since 2002 (4).

Surveillance of childhood diabetes is challenging. First, we are unable to employ existing national surveillance systems such as the National Health and Nutrition Examination Survey (NHANES) because childhood diabetes is uncommon; NHANES (1999-2002) yielded only18 self-reported cases of diabetes among youth aged 12 to 19 years (5). Second, ascertainment of childhood diabetes cases is often costly in terms of time and financial resources. Currently the SEARCH study conducts validation of potential cases by manual review of medical records which is expensive, although the resulting case ascertainment is estimated to be very complete (i.e., >90% for both prevalent and incident cases) based on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

capture-recapture analyses (6). Third, a useful childhood diabetes surveillance system should be able to discriminate between types of diabetes in different age, and racial/ethnic groups as the distribution of childhood diabetes varies by type, age and race/ethnicity (7, 8); and both etiology and treatment differ by diabetes type (7, 9-11).

The increasing utilization of computerized medical information systems may provide timely data for diabetes surveillance with substantially reduced cost relative to traditional approaches (12-14). Approaches to identify diabetes cases and classify type have been explored using administrative data (15-22), and electronic health record (EHR) data (23-26). Among all childhood diabetes algorithms in the literature, only two explored type-specific algorithms (21, 23). None of these studies evaluated algorithm performance according to age and race/ethnicity. In the U.S., it is not known whether administrative and EHR data from a large academic care delivery system can be used to accurately differentiate between childhood type 1 and type 2 diabetes (i.e., through use of type-sensitive algorithms) or whether such data can only identify cases without regard to type (i.e., through use of type-insensitive algorithms). It is also not known whether the performance of automated algorithms differs by age and race/ethnicity.

Our objective was to identify algorithms with high performance, as demonstrated by high sensitivity, specificity, and positive predictive value (PPV), with a goal to efficiently identify diabetes cases and classify type in youth, overall and by age and race/ethnicity in a large academic care delivery system caring for patients with all payment sources, utilizing administrative and EHR data from the University of North Carolina Health Care System (UNCHCS).

Methods

The UNCHCS is a large not-for-profit integrated academic health care system located in central North Carolina, caring for a broad range of patients including those insured by Medicaid or without insurance. With a central 800-bed tertiary care center, and through its network of primary care and specialty physician practices located in 5 counties, UNC cares for over 800,000 people annually. Among insured patients, the predominant system of care in North Carolina is fee for service, with elements of care management.

Data sources

Three independent sources of data, presented below, were utilized from the Carolina Data Warehouse for Health (CDW-H), an enterprise-wide data warehouse within UNCHCS.

EHR data included demographics; outpatient medication lists; clinic, procedure, and hospitalization notes; patient problem lists; and laboratory test results. The outpatient medication list is a regularly updated record of active medications patients report being prescribed or taking, as well as inactive medications patients previously reported or were prescribed. The patient problem list includes ICD-9-CM codes for patients' past and current illnesses. Laboratory data are sourced from laboratories both internal and external to UNCHCS, from which we identified dates and results for tests of fasting or random blood glucose as well as HbA1c, diabetes auto-antibodies (GAD65, IAA and ICA), and C-peptide.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Inpatient medication order data were obtained from the Computerized Provider Order Entry.

Billing data included physician reimbursements and hospital (facility) charges accrued during outpatient and inpatient visits. Billing data, consisting of ICD-9-CM codes for each patient-visit, were captured separately for outpatient and inpatient visits. Diagnosis codes recorded at discharge, not at admission, were used for this analysis. The following diabetes-related ICD-9-CM codes were used: 250.xx (diabetes mellitus); 775.1 (neonatal diabetes); 648.0x (diabetes in pregnancy, non-gestational); 357.2 (diabetic neuropathy); 362.0x (diabetic retinopathy); and 366.41 (diabetic cataract).

This study was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill.

Study population

Study population—The population of interest was defined as all children < 20 years of age as of December 31, 2011 who were seen by a health care provider at UNCHCS at least once for any reason in 2011.

Case ascertainment—An initial algorithm was applied to the study population to identify presumptive diabetes cases who met any criteria in the initial algorithm at any time from July 1, 2008 to December 31, 2011. The initial algorithm was designed to ensure very high sensitivity; thus, children not identified by this initial algorithm were assumed to not have diabetes (i.e., true negatives). The initial algorithm included the following: 1) ≥1 HbA1c ≥ 6.0% (42 mmol/mol); or 2) ≥2 random blood glucose ≥ 200mg/dL on different days or ≥1 fasting blood glucose ≥126mg/dL; or 3) ≥1 patient problem list diabetes-related ICD-9-CM codes; or 4) ≥1 billing data diabetes-related ICD-9-CM codes; or 5) ≥1 diabetes-related medications, including insulin, glucagon, metformin, sulfonylurea, GLP-1 receptor agonists, thiazolidinediones and other hypoglycemic agents. The time period allowed to find evidence for diabetes reflected the SEARCH study protocol for prevalent case ascertainment.

Diabetes case validation

Using the SEARCH case validation approach (1), diabetes status of the presumptive cases and diabetes type of true cases were determined by presence of a diagnosis of diabetes in the EHR in one or more notes written by health care providers (gold standard). Five reviewers were trained using the SEARCH standardized protocol by a member of the SEARCH team (J.T.) who has 10 years of experience with the SEARCH case ascertainment protocol. Each week, 5% of the records reviewed by each reviewer were validated by the trainer. If any discrepancies were found, these were discussed immediately.

A case validation form was used to collect presumptive cases' demographics, prevalent diabetes status, initial and most recent diabetes type and associated diagnosis date, and presence of diabetic ketoacidosis (DKA) and diabetes auto-antibodies tests within 6 months of diagnosis. The most recent diabetes type recorded by the provider was used.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Criteria for evaluating algorithms' performance

True diabetes cases and their type validated by the standardized medical record review described above established our "gold standard". Diabetes type-insensitive algorithms were evaluated within the full study population while the type-sensitive algorithms were evaluated among true diabetes cases.

No established cutpoints were available to evaluate the usefulness of algorithms quantitatively. Some attributes considered crucial for public health surveillance systems include simplicity, timeliness, sensitivity and PPV (27). High sensitivity is crucial for identifying most of the cases. High PPV is preferred to reduce the number of false positives while high specificity is important to distinguish between type 1 and type 2 youth. Thus, sensitivity, specificity and PPV are our primary outcomes of interest, and a promising algorithm should yield values approaching or greater than 90%.

Evaluation of type-insensitive and type-sensitive case ascertainment algorithms

The sensitivity, specificity, and PPV were computed for each of the five criteria listed in the initial algorithm above, and combinations thereof, relative to ascertainment of diabetes regardless of type. The capacity to ascertain type 1 and type 2 cases was explored using various combinations of ICD-9-CM 250.xx billing codes, outpatient medications, and laboratory test results. Multiple instances of the same laboratory tests and billing codes on the same day were counted only once; only the highest laboratory value in a day was retained for analyses. Ratios of the number of type 1 or type 2 billing codes to all ICD-9-CM 250.xx billing codes identified within the whole 3.5 years surveillance window were also considered (24). Specifically, the ratio for type 1 algorithm was calculated as the number of type 1 billing codes divided by the total number of type 1 and type 2 billing codes; the ratio for type 2 algorithm was calculated as the number of type 2 billing codes divided by the total number of type 1 and type 2 billing codes.

Inpatient medication data included in the initial algorithm to ensure high sensitivity were excluded for evaluation of the algorithms, because among 59 presumptive cases whose only indication of diabetes was inpatient use of diabetes-related medications, none were true cases. Despite excluding inpatient medication data, these 59 presumptive cases were included in the analyses.

Given the importance of age and race/ethnicity in characterizing the prevalence of childhood diabetes, we evaluated the performance of the algorithms within two age groups (< 10 vs ≥ 10 years) since type 2 diabetes is rare under the age of 10 (8), and two racial/ethnic groups, non-Hispanic white (NHW) versus "other" (7, 8). The other reason to select age 10 as the cutpoint is due to its common use in the literature as a criterion to differentiate between type 1 and type 2 diabetes in children (21, 23). Nevertheless, we should be aware of potential bias that may result from this choice.

Data analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results

Demographic and clinical characteristics of the study population

The initial algorithm identified 1,348 presumptive diabetes cases from the population of 57,767 children with any health care encounters in 2011 (Table 1). Review of the medical records for these 1,348 presumptive cases yielded 537 true cases: 405 type 1, 86 type 2 and 46 cases of other type. The mean age and standard deviation (SD) of true diabetes cases, 14.4 (4.1) years, was greater than that of the overall study population, 8.6 (6.2) years, and presumptive cases, 11.1 (6.5) years. Mean diagnosis age (SD) was 7.9 (4.3) years for type 1 and 12.9 (2.3) years for type 2. The prevalence of DKA within 6 months of diagnosis was 25% among diabetes cases. Nearly half of type 2 youth used insulin and 9% of type 1 youth used metformin.

Performance of diabetes type-insensitive algorithms

Full sample—The specificity of all type-insensitive algorithms was >99% (Table 2 and Table 3). Of the 5 individual criteria evaluated, the glucose criterion captured the most presumptive cases, but also the most false positives indicated by the lowest PPV (95% confidence interval (CI)) of 45.9% (42.5%, 49.2%) (Table 2). Interestingly, of the 435 children who met only the glucose criterion, only 1 was diagnosed with diabetes (Supplemental Table 1, specific data combinations). Billing data was the best single criterion. The sensitivity (95% CI) was 97.0% (95.6%, 98.5%) while the PPV (95% CI) was 82.2% (79.2%, 85.2%). The patient problem list had the highest PPV of 97.0% (94.9%, 99.0%), but the sensitivity was only 48.0% (43.8%, 52.3%). Combining the patient problem list and billing data together (sensitivity=97.2% (95.8%, 98.6%); PPV=81.8% (78.8%, 84.8%)) did not result in improved performance compared to use of billing data alone as 262 out of 266 diabetes cases captured by the patient problem list were flagged by billing data as well (data not shown). The sensitivity and PPV of the algorithm requiring ≥ 2 criteria met were 90.7% (88.2%, 93.1%) and 89.0% (86.4%, 91.7%), respectively. The performance of each specific combination of the data was listed in the Supplemental Table 1. The PPV was only 6.7% (4.9%, 8.5%) for children who met only one criterion.

Age and racial/ethnic subgroups—Although billing data was the best single criterion in the full sample, its PPV was 62.5% (54.1%, 70.9%) in children < 10 years of age compared to 87.2% (84.2%, 90.1%) in children ≥10 years of age (Table 2). The PPV of outpatient medications was higher in the younger group 84.1% (76.4%, 91.7%) compared to the older group 74.5% (70.9%, 78.2%). The algorithm that required meeting ≥2 criteria performed similarly in the two age groups. The greatest difference between age groups was seen using the algorithm that included fasting or random glucose values (PPV: 16.6% (13.1%, 20.1%) versus 75.1% (71.0%, 79.2%)). Regarding the racial/ethnic subgroups, the algorithms' performances were generally better in the NHW group than in the "other" group (Table 3). The greatest difference between racial/ethnic groups was seen using HbA1c, which had a PPV of 91.4% (88.0%, 94.8%) in the NHW group compared to 65.3% (59.0%, 71.6%) in the "other" group.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Performance of diabetes type-sensitive algorithms

Type 1 algorithms—The algorithm requiring the ratio of the number of type 1 billing codes to the sum of type 1 and type 2 billing codes ≥0.5 had sensitivity, specificity and PPV >90% in both the full sample, and in all age and racial/ethnic subgroups (Table 4 and Table 5), with the exception of specificity of 83.3% in children <10 years of age (95% CI was not computed due to only 6 true non-cases, so it's not reliable). The algorithms' performance was not improved with addition of laboratory data and outpatient medications data.

Type 2 algorithms—The algorithm requiring the ratio of the number of type 2 billing codes to the sum of type 1 and type 2 billing codes ≥0.4 had a sensitivity of 91.9% (86.1%, 97.6%) and a PPV of only 57.2% (49.0%, 65.5%) in the full sample, but it had a sensitivity of 93.1% (86.6%, 99.6%), a specificity of 87.2% (81.8%, 92.6%) and a PPV of 74.0% (63.9%, 84.0%) in youth with "other" race/ethnicity (Table 5). With the addition of medication data, the PPV increased to 78.2% (69.0%, 87.4%), but the sensitivity decreased to 70.9% (61.3%, 80.5%) in the full sample. Generally, the type 2 algorithms had better performance in the "other" race/ethnicity group. Age subgroups were not evaluated because only 2 children < 10 years of age had type 2 diabetes.

Discussion

This study supports the use of automated algorithms from administrative and EHR data, with billing data as the best single data source, to ascertain cases of childhood diabetes overall, cases of type 1 diabetes, and cases of type 2 diabetes in youth who were not NHW. We found considerable age and racial/ethnic differences in performance of the five individual diabetes criteria and considerable racial/ethnic differences in performance of type 2 algorithms. Minimal age and racial/ethnic differences were observed for type 1 algorithms. However, no internal and external validation of algorithms were performed for those algorithms.

Our results suggest that using billing data alone may facilitate identifying diabetes cases regardless of type. In fact, billing data was the best single criterion, with little gained from the use of additional data. Based on our 90% evaluation criteria, the algorithm that required meeting ≥2 criteria could also be used. Our study is consistent with most of the literature in regards to billing data being the best single criterion in identifying diabetes cases (17, 22).

Additionally, our study revealed that youth with diabetes whose race/ethnicity was other than NHW or those <10 years of age were more difficult to identify. Billing data may not be sufficient in those subpopulations. For younger youth, depending on the surveillance goals, outpatient medications or the algorithm requiring ≥2 criteria met may be more viable options. The caveat with the latter algorithm was that the individual criteria used between systems should be the same in order to have comparable accuracy. Similarly, Zgibor et al (28) pointed out the best algorithm was the one with two or more criteria met or an outpatient diagnosis code. However, this study was conducted in an adult population and with different individual criteria. When accounting for racial/ethnic differences, although billing data may still be the best, the accuracy of ascertainment (i.e., PPV) was less

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

preferable for the “other” race/ethnicity group. Similar decreases in PPV were observed for the algorithm requiring ≥2 criteria met in the “other” race/ethnicity group.

A useful diabetes surveillance system should be able to ascertain cases efficiently according to diabetes type. Our findings suggest that the best type 1 algorithm was the ratio of the number of type 1 billing codes to the sum of type 1 and type 2 billing codes ≥0.5. There were minimal differences in performance of type 1 algorithms across age and racial/ethnic groups. These results are consistent with previous studies that have reported that type 1 youth are easier to accurately identify relative to type 2 youth (21, 23). Interestingly, the performance of type 1 algorithms was not improved with addition of medication and laboratory data. Our best type 1 algorithm satisfies the crucial attributes for public health surveillance systems including simplicity, sensitivity, PPV (27), and high specificity.

Type 2 algorithms were far from optimal in ascertaining type 2 cases overall. Age differences were not considered due to the fact that type 2 diabetes is exceedingly rare in youth with age <10 years (8); only two type 2 cases were identified in our population. The algorithms performed considerably better in the “other” race/ethnicity group compared to the NHW group. This may be related to the higher prevalence of type 2 diabetes in population subgroups other than NHW (8). The algorithm with the ratio of the number of type 2 billing codes to the sum of type 1 and type 2 billing codes ≥ 0.4 was the best with a sensitivity 93.1% (86.6%, 99.6%), a specificity of 87.2% (81.8%, 92.6%) and a PPV of 74.0% (63.9%, 84.0%) among youth with “other” race/ethnicity. This performance seems acceptable if manual review of medical records is not an option in a system, given the evidence that childhood type 2 cases are more difficult to ascertain (21, 23). There were no useful type 2 algorithms for NHW youth. Thus, relying solely on automated algorithms for ascertaining type 2 youth from an overall sample of youth with diabetes or for NHW youth with diabetes may not be possible with current EHR and billing data. The low performance may be attributable to the fact that type 2 ICD-9-CM codes include unspecified diabetes type (29); insulin is commonly used in type 2 diabetes (9) and metformin is also used to treat polycystic ovarian syndrome and for weight control, leading to false positives (24). Together, these factors represent a significant barrier to developing a useful type 2 algorithm.

Although we have highlighted some useful algorithms, the purpose was not to advocate using them in health systems directly without any validation study at present. Careful consideration should be given (e.g., the goal of the study) in terms of selecting an algorithm or developing a new one. Particularly, a small pilot validation study may be conducted within a system before implementing the algorithm in the whole system.

Differences in case ascertainment performance of HbA1c and fasting/random glucose criteria were observed according to race/ethnicity and age. Racial/ethnic differences were greatest for the HbA1c criterion. The PPV was considerably lower in the “other” race/ethnicity group. When a cutpoint of 6.5% was used, a diagnostic value for diabetes in adults, instead of 6.0%, the sensitivity was similar, but the PPV was significantly improved for the “other” race/ethnicity group (93.2% versus 65.3%). The explanation may be that NHW youth had significantly lower HbA1c level compared to youth with “other” racial/ethnicity

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in the general U.S. population (30), and diabetes population (31). There is currently no consistency about which HbA1c value should be used for surveillance; studies previously used 6.7% (32), 6.5% (19, 25), or a presence of HbA1c test regardless of value (26, 28). Our study used 6.0% to ensure very high sensitivity of the initial algorithm. The appropriate cutpoint of HbA1c value for surveillance should be further evaluated in other populations.

As for fasting/random glucose, the performance differed significantly by age groups. The PPV of the fasting/random glucose criterion (45.9%) was lower than that in the recent analysis by Lawrence et al. (>60%) (23); however, only six to twelve months of laboratory data were used in their study. In our study, the PPV increased to a comparable level of 63.2% when one year of data was used while the sensitivity remained similar (data not shown). Our original low PPV may be driven by the younger age group rather than by the length of surveillance window, because the PPV was 75.1% in the older group in our study. In fact, the PPV increased only from 16.6% to 31.0% for age <10 years group using one year of data. Our population had a proportion of youth < 5 years of age twice that of the sample used in the study by Lawrence et al. (23) The findings of the present study suggest that glucose criterion may not accurately ascertain childhood diabetes cases in youth < 10 years of age. The reasons may involve the fact that younger youth visit health system more frequently and a qualified glucose test result can frequently be triggered by conditions other than diabetes. Also, children with fasting glucose results pulled from EHR may not be truly fasting at the blood draw. Further, a fasting glucose result may be mislabeled as the random glucose result.

This study has several strengths. First, to the best of our knowledge, this is the first U.S. study to evaluate demographic differences in the performance of diabetes case ascertainment algorithms by diabetes type. Also, case ascertainment and type classification approaches for childhood diabetes have not been well explored, especially outside of Canada where health care is universal and the Kaiser Permanente, a large integrated health system (23). Our study identified useful algorithms that could greatly facilitate surveillance efforts. Furthermore, all presumptive cases were validated individually by review of the complete medical records, which served as our gold standard.

Limitations should be noted. First, our study population included all children < 20 years old who were seen at least once at UNCHCS in 2011. Therefore, the study population may not represent all youth served by UNCHCS. However, the DKA prevalence (25.0%) was the same as found in SEARCH (25.5%) (33), and use of diabetes medications was also similar (9). Hence, sampling bias likely posed little, if any, influence, with the assumption that youth with diabetes are regularly seen in a health care setting. Second, we made an assumption that children not selected by the initial algorithm were true negatives. It is possible that some true cases were not identified by our initial algorithm. However, in our data, children identified only by inpatient medications were all true negatives and only 1 of the 435 children captured only by glucose criterion had diabetes. In fact, the overall false positive proportion was 93.3% among youth who met only 1 criterion. With the very high proportion of false positives among children with only one criterion met and the very low prevalence of diabetes in children, it is exceedingly unlikely that false negative case numbers would be sufficiently high to impact our findings. Finally, this work applies to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

estimation of prevalence of childhood diabetes, not incidence which would require ascertainment of diagnosis date.

In conclusion, automated algorithms from administrative and EHR data may be useful to ascertain cases of diabetes without regard to type, as well as cases of type 1 diabetes and cases of type 2 diabetes among youth with race/ethnicity other than NHW. Detailed review of medical records may be needed to ascertain type 2 cases in NHW youth accurately. Future work will be required to replicate our case ascertainment methodologies in other health systems to determine the generalizability and to inform the development of low-cost sustainable public health surveillance systems of childhood diabetes in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The SEARCH for Diabetes in Youth Study is indebted to the many youth and their families, and their health care providers, whose participation made this study possible.

Grant Support: SEARCH for Diabetes in Youth is funded by the Centers for Disease Control and Prevention (PA numbers 00097, DP-05-069, and DP-10-001) and supported by the National Institute of Diabetes and Digestive and Kidney Diseases.

Site Contract Numbers: Kaiser Permanente Southern California (U48/CCU919219, U01 DP000246, and U18DP002714), University of Colorado Denver (U48/CCU819241-3, U01 DP000247, and U18DP000247-06A1), Kuakini Medical Center (U58CCU919256 and U01 DP000245), Children's Hospital Medical Center (Cincinnati) (U48/CCU519239, U01 DP000248, and 1U18DP002709), University of North Carolina at Chapel Hill (U48/CCU419249, U01 DP000254, and U18DP002708), University of Washington School of Medicine (U58/CCU019235-4, U01 DP000244, and U18DP002710-01), Wake Forest University School of Medicine (U48/CCU919219, U01 DP000250, and 200-2010-35171).

This study was also supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Award Number UL1TR000083.

The authors wish to acknowledge the involvement of General Clinical Research Centers (GCRC) at the South Carolina Clinical & Translational Research (SCTR) Institute, at the Medical University of South Carolina (NIH/NCRR Grant number UL1RR029882); Seattle Children's Hospital (NIH CTSA Grant UL1 TR00423 of the University of Washington); University of Colorado Pediatric Clinical and Translational Research Center (CTRC) (Grant Number UL1 TR000154) and the Barbara Davis Center at the University of Colorado at Denver (DERC NIH P30 DK57516); and the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant 8 UL1 TR000077; and the Children with Medical Handicaps program managed by the Ohio Department of Health.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention and the National Institute of Diabetes and Digestive and Kidney Diseases.

References

1. SEARCH Study Group. SEARCH for Diabetes in Youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. Control Clin Trials. 2004; 25:458–471. [PubMed: 15465616]
2. Dabelea D, et al. Is prevalence of type 2 diabetes increasing in youth? the SEARCH for diabetes in youth study [abstract]. Diabetes. 2012; 66(Suppl. 1):A61.
3. Mayer-Davis EJ, et al. Increase in prevalence of type 1 diabetes from the SEARCH for diabetes in youth study: 2001 to 2009 [abstract]. Diabetes. 2012; 66(Suppl. 1):A322.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4. Imperatore G, Boyle JP, Thompson TJ, et al. Projections of type 1 and type 2 diabetes burden in the U.S. population aged <20 years through 2050: dynamic modeling of incidence, mortality, and population growth. Diabetes Care. 2012; 35:2515–2520. [PubMed: 23173134]
5. Duncan GE. Prevalence of diabetes and impaired fasting glucose levels among US adolescents: National Health and Nutrition Examination Survey, 1999-2002. Arch Pediatr Adolesc Med. 2006; 160:523–528. [PubMed: 16651496]
6. Hamman, RF.; Dabelea, D.; Liese, AD., et al. [5 January 2014] SEARCH TECHNICAL REPORT: Estimation of Completeness of Case Ascertainment Using Capture-Recapture 2013. https://searchfordiabetes.org/publications.cfm
7. Amed S, Dean HJ, Panagiotopoulos C, et al. Type 2 diabetes, medication-induced diabetes, and monogenic diabetes in Canadian children: a prospective national surveillance study. Diabetes Care. 2010; 33:786–791. [PubMed: 20067956]
8. Liese AD, D'Agostino RB Jr, et al. SEARCH for Diabetes in Youth Study Group. The burden of diabetes mellitus among US youth: prevalence estimates from the SEARCH for Diabetes in Youth Study. Pediatrics. 2006; 118:1510–1518. [PubMed: 17015542]
9. Bell RA, Mayer-Davis EJ, Beyer JW, et al. Diabetes in non-Hispanic white youth: prevalence, incidence, and clinical characteristics: the SEARCH for Diabetes in Youth Study. Diabetes Care. 2009; 32(Suppl 2):S102–11. [PubMed: 19246575]
10. Dabelea D, Pihoker C, Talton JW, et al. Etiological approach to characterization of diabetes type: the SEARCH for Diabetes in Youth Study. Diabetes Care. 2011; 34:1628–1633. [PubMed: 21636800]
11. Shield JP, Lynn R, Wan KC, et al. Management and 1 year outcome for UK children with type 2 diabetes. Arch Dis Child. 2009; 94:206–209. [PubMed: 18838418]
12. Jha AK. Meaningful use of electronic health records: the road ahead. JAMA. 2010; 304:1709–1710. [PubMed: 20959581]
13. Klompas M, McVetta J, Lazarus R, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. Am J Prev Med. 2012; 42(6 Suppl 2):S154–62. [PubMed: 22704432]
14. Virnig BA, McBean M. Administrative data for public health surveillance and planning. Annu Rev Public Health. 2001; 22:213–230. [PubMed: 11274519]
15. Amed S. Validation of diabetes case definitions using administrative claims data. Diabetic Med. 2011; 28:424–427. [PubMed: 21392063]
16. Chen G, Khan N, Walker R, et al. Validating ICD coding algorithms for diabetes mellitus from administrative data. Diabetes Res Clin Pract. 2010; 89:189–195. [PubMed: 20363043]
17. Dart AB, Martens PJ, Sellers EA, et al. Validation of a pediatric diabetes case definition using administrative health data in manitoba, Canada. Diabetes Care. 2011; 34:898–903. [PubMed: 21378211]
18. Guttmann A, Nakhla M, Henderson M, et al. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. Pediatr Diabetes. 2010; 11:122–128. [PubMed: 19500278]
19. Harris SB, Glazier RH, Tompkins JW, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. BMC Health Serv Res. 2010; 10:347-6963–10-347. [PubMed: 21182790]
20. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. Diabetes Care. 2004; 27(Suppl 2):B10–21. [PubMed: 15113777]
21. Vanderloo SE, Johnson JA, Reimer K, et al. Validation of classification algorithms for childhood diabetes identified from administrative data. Pediatr Diabetes. 2012; 13:229–234. [PubMed: 21771232]
22. Wilson C, Susan L, Lynch A, et al. Patients with diagnosed diabetes mellitus can be accurately identified in an Indian Health Service patient registration database. Public Health Rep. 2001; 116:45–50. [PubMed: 11571407]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

23. 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]
24. 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]
25. Nichols GA, Desai J, Elston Lafata J, 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]
26. Wilke RA, Berg RL, Peissig P, et al. Use of an electronic medical record for the identification of research subjects with diabetes mellitus. Clin Med Res. 2007; 5:1–7. [PubMed: 17456828]
27. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001; 50:1–35. quiz CE1-7. [PubMed: 18634202]
28. 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]
29. Rhodes ET, Laffel LM, Gonzalez TV, et al. Accuracy of administrative coding for type 2 diabetes in children, adolescents, and young adults. Diabetes Care. 2007; 30:141–143. [PubMed: 17192348]
30. Saaddine JB, Fagot-Campagna A, Rolka D, et al. Distribution of HbA(1c) levels for children and young adults in the U.S.: Third National Health and Nutrition Examination Survey. Diabetes Care. 2002; 25:1326–1330. [PubMed: 12145229]
31. Kirk JK, D'Agostino RB Jr, Bell RA, et al. Disparities in HbA1c levels between African-American and non-Hispanic white adults with diabetes: a meta-analysis. Diabetes Care. 2006; 29:2130–2136. [PubMed: 16936167]
32. Southern DA, Roberts B, Edwards A, et al. Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. Can J Public Health. 2010; 101:61–64. [PubMed: 20364541]
33. Rewers A, Klingensmith G, Davis C, et al. Presence of diabetic ketoacidosis at diagnosis of diabetes mellitus in youth: the Search for Diabetes in Youth Study. Pediatrics. 2008; 121:e1258–12. [PubMed: 18450868]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Demographic and clinical characteristics of the total study population, presumptive diabetes cases and diabetes cases in the Carolina Data Warehouse for Health in 2011

	Study population %	Presumptive cases %	True diabetes cases, %			
			Total	Type 1	Type 2	Other*
All, n	57,767	1348	537	405	86	46
Gender						
Male	51.9	47.8	44.3	45.2	39.5	45.7
Female	48.1	52.2	55.7	54.8	60.5	54.3
Age (years)†						
0-4.9	36.3	26.5	2.8	3.2	0	4.3
5.0-9.9	22.2	12.0	12.5	15.6	2.3	4.3
10.0-14.9	19.8	23.6	32.6	34.6	34.9	10.9
15.0-17.9	12.7	21.8	30.2	26.4	37.2	50.0
18.0-19.9	9.0)	16.1	22.0	20.3	25.6	30.4
Mean age, years ± SD	8.6 ± 6.2	11.1 ± 6.5	14.4 ± 4.1	13.9 ± 4.3	15.8 ± 2.6	16.0 ± 3.7
Age at diagnosis, years ± SD			8.8 ± 4.6	7.9 ± 4.3	12.9 ± 2.3	12.8 ± 5.3
Race/ethnicity‡						
Non-Hispanic White	48.5	52.5	61.6	67.9	32.6	60.9
Black	21.0	29.2	25.5	19.8	50.0	30.4
Hispanic	14.6	11.4	7.1	6.7	9.3	6.5
Asian Pacific Islanders	1.8	1.3	0.9	1.0	1.2	0
American Indian	0.6	1.3	1.3	0.7	3.5	2.2
Other/unknown/missing	13.6	4.3	3.5	4.0	3.5	0
Health insurance type						
Private	43.2	39.5	50.8	55.3	27.9	54.3

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	Study population %	Presumptive cases %	True diabetes cases, %			
			Total	Type 1	Type 2	Other*
Government	42.4	48.1	34.1	29.1	57.0	34.8
Tricare§	8.3	9.0	11.0	12.1	5.8	10.9
Not insured/missing/other	6.2	3.3	4.1	3.5	9.3	0
DKA within 6 months of diagnosis						
Yes			25.0	31.9	2.3	6.5
No/unknown			75.0	68.1	97.7	93.5
DAA within 6 months of diagnosis						
Yes			49.3	54.1	38.4	28.3
No/unknown			50.7	45.9	61.6	71.7
DAA test results from CDW-H						
Yes		13.1	30.0	29.4	31.4	32.6
Positive		61.0	67.1	88.2	3.7	13.3
Negative		39.0	32.9	11.8	96.3	86.7
No		86.9	70.0	70.6	68.6	67.4
C-peptide test results from CDW-H						
Yes		12.2	27.6	27.4	30.2	23.9
High		12.1	8.8	0	38.5	27.3
Normal		34.5	33.8	26.1	57.7	54.5
Low		53.3	57.4	73.9	3.8	18.2
No		87.8	72.4	72.6	69.8	76.1
Medications from EHR and COPE						
Insulin prescription		44.7	82.3	91.9	46.5	65.2
Metformin prescription		17.8	20.1	8.9	72.1	21.7
Glucagon prescription		27.6	66.1	77.0	32.6	32.6
Other OHA prescription		1.3	2.2	0.7	5.8	8.7

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Abbreviations: COPE, computerized provider order entry; DAA, diabetes autoantibodies; DKA, diabetic ketoacidosis; EHR, electronic health record; OHA, oral hypoglycemic agents; SD, standard deviation.

[*] Includes 25 secondary diabetes cases, 2 cases of maturity onset diabetes of the young and 19 diabetes type unspecified cases.

[†] Age was calculated based on December 31, 2011.

[‡] Race/ethnicity data were from medical records.

[§] Military health insurance plan

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Performance of type-insensitive algorithms by age applied to the total study population (N=57,767) in the Carolina Data Warehouse for Health in 2011

diabetes criteria* met	N†			Sensitivity (%)			Specificity (%)			PPV (%)		
	Total	<10 y	≥10 y	Total	<10 y	≥10 y	Total	<10 y	≥10 y	Total	<10 y	≥10 y
HbA1c	486	85	401	72.1	82.9	70.1	99.8	99.9	99.7	79.6	80.0	79.6
Fasting/random glucose	857	428	429	73.2	86.6	70.8	99.2	99.2	99.5	45.9	16.6	75.1
Patient problem list	266	52	214	48.0	59.8	45.9	99.9	99.9	99.9	97.0	94.2	97.7
Billing data	634	128	506	97.0	97.6	96.9	99.8	99.8	99.7	82.2	62.5	87.2
Outpatient medications	626	88	538	88.5	90.2	88.1	99.7	99.7	99.4	75.9	84.1	74.5
At least 1 criteria	1289	503	786	100.0	100.0	100.0	98.7	98.7	98.6	41.7	16.3	57.9
At least 2 criteria	547	89	458	90.7	91.5	90.5	99.9	99.9	99.8	89.0	84.3	90.0
At least 3 criteria	453	75	378	81.3	87.8	80.2	99.9	99.9	99.9	96.5	96.0	96.6
At least 4 criteria	377	69	308	69.3	82.9	66.8	99.9	99.9	99.9	98.7	98.6	98.7
5 criteria	203	45	158	37.4	54.9	34.3	99.9	99.9	99.9	99.0	100.0	98.7

Abbreviation: PPV, positive predictive value; y, years

True diabetes cases (N=537) confirmed by medical record review established our “gold standard” for evaluation of algorithms’ performance.

*Criteria are ≥1 HbA1c ≥ 6.0% (42mmol/mol) OR ≥2 random blood glucose ≥ 200mg/dL on different days or ≥1 fasting blood glucose ≥ 126mg/dL OR ≥1 patient problem list diabetes-related ICD-9-CM codes OR ≥1 billing data diabetes-related ICD-9-CM codes OR ≥1 diabetes-related outpatient medications, including insulin, glucagon, metformin, sulfonylurea, GLP-1 receptor agonists, thiazolidinediones and other hypoglycemic agents.

†Number of youth captured.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Performance of type-insensitive algorithms by race/ethnicity applied to the total study population (N=57,767) in the Carolina Data Warehouse for Health in 2011

diabetes criteria* met	N†			Sensitivity (%)			Specificity (%)			PPV (%)		
	Total	NHW	Other	Total	NHW	Other	Total	NHW	Other	Total	NHW	Other
HbA1c	486	267	219	72.1	73.7	69.4	99.8	99.9	99.7	79.6	91.4	65.3
Fasting/random glucose	857	488	369	73.2	77.0	67.0	99.2	99.2	99.2	45.9	52.3	37.4
Patient problem list	266	165	101	48.0	48.6	47.1	99.9	99.9	99.9	97.0	97.6	96.0
Billing data	634	378	256	97.0	97.9	95.6	99.8	99.8	99.8	82.2	85.7	77.0
Outpatient medications	626	367	259	88.5	88.8	87.9	99.7	99.7	99.7	75.9	80.1	69.9
At least 1 criteria	1289	683	606	100.0	100.0	100.0	98.7	98.7	98.6	41.7	48.5	34.0
At least 2 criteria	547	329	218	90.7	91.5	89.3	99.9	99.9	99.9	89.0	92.1	84.4
At least 3 criteria	453	284	169	81.3	83.7	77.7	99.9	99.9	99.9	96.5	97.5	94.7
At least 4 criteria	377	243	134	69.3	73.1	63.1	99.9	99.9	99.9	98.7	99.6	97.0
5 criteria	203	126	77	37.4	37.8	36.9	99.9	99.9	99.9	99.0	99.2	98.7

Abbreviation: NHW, non-Hispanic white; PPV, positive predictive value;

True diabetes cases (N=537) confirmed by medical record review established our “gold standard” for evaluation of algorithms’ performance.

*Criteria are ≥1 HbA1c ≥ 6.0% (42mmol/mol) OR ≥2 random blood glucose ≥ 200mg/dL on different days or ≥1 fasting blood glucose ≥ 126mg/dL OR ≥1 patient problem list diabetes-related ICD-9-CM codes OR ≥1 billing data diabetes-related ICD-9-CM codes OR ≥1 diabetes-related outpatient medications, including insulin, glucagon, metformin, sulfonylurea, GLP-1 receptor agonists, thiazolidinediones and other hypoglycemic agents.

†Number of youth captured.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Performance of type-sensitive algorithms by age applied to the true cases (N=537) in the Carolina Data Warehouse for Health in 2011

	diabetes algorithms	N*			Sensitivity (%)			Specificity (%)			PPV (%)		
		Total	<10 y	≥10 y	Total	<10 y	≥10 y	Total	<10 y	≥10 y	Total	<10 y	≥10 y
Type 1 diabetes	≥1 type 1 codes†	443	76	367	97.0	98.7	96.7	62.1	83.3	61.1	88.7	98.7	86.6
	≥2 type 1 codes	415	76	339	93.6	98.7	92.4	72.7	83.3	72.2	91.3	98.7	89.7
	0 type 2 codes and ≥1 type 1 codes	141	18	123	34.6	23.7	37.1	99.2	100.0	99.2	99.3	100.0	99.2
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.3†‡	413	76	337	96.8	98.7	96.4	84.1	83.3	84.1	94.9	98.7	94.1
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.4	402	76	326	96.3	98.7	95.7	90.9	83.3	91.3	97.0	98.7	96.6
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5	397	76	321	95.6	98.7	94.8	92.4	83.3	92.9	97.5	98.7	97.2
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.6	383	75	308	93.3	97.4	92.4	96.2	83.3	96.8	98.7	98.7	98.7
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and insulin use OR ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and glucagon use	372	73	299	89.9	94.7	88.8	93.9	83.3	94.4	97.8	98.6	97.7
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and insulin use OR ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and glucagon use OR DAA positive OR C-peptide negative	378	73	305	90.4	94.7	89.4	90.9	83.3	91.3	96.8	98.6	96.4
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5 OR ≥1 type 1 code and glucagon and no oral medications § OR ≥1 type 1 code and insulin and no oral medications § OR DAA positive OR C-peptide negative	412	76	336	96.5	98.7	96.0	84.1	83.3	84.1	94.9	98.7	94.0
Type 2 diabetes	≥1 type 2 codes	375	62	313	91.9	N/A	92.9	34.4	N/A	36.7	21.1	N/A	24.9
	≥1 type 2 codes and 0 type 1 codes and no insulin use	57	4	53	46.5	N/A	46.4	96.2	N/A	96.2	70.2	N/A	73.6
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.3	151	10	141	91.9	N/A	92.9	84.0	N/A	83.0	52.3	N/A	55.3
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.4	138	6	132	91.9	N/A	92.9	86.9	N/A	85.4	57.2	N/A	59.1
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.5	128	4	124	88.4	N/A	89.3	88.5	N/A	86.8	59.4	N/A	60.5
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.6	116	4	112	84.9	N/A	85.7	90.5	N/A	89.2	62.9	N/A	64.3

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

diabetes algorithms	N*			Sensitivity (%)			Specificity (%)			PPV (%)		
	Total	<10 y	≥10 y	Total	<10 y	≥10 y	Total	<10 y	≥10 y	Total	<10 y	≥10 y
0 type 1 codes and ≥3 type 2 codes and oral medications §	22	1	21	25.6	N/A	25.0	100.0	N/A	100.0	100.0	N/A	100.0
Ratio of type 2 to type 1 and type 2 codes ≥ 0.4 and oral medications §	78	1	77	70.9	N/A	71.4	96.2	N/A	95.4	78.2	N/A	77.9

Abbreviation: DAA, diabetes auto-antibodies; NA, not available; PPV, positive predictive value; y, years

True diabetes cases (N=537) and type confirmed by medical record review established our "gold standard" for evaluation of algorithms' performance. Only outpatient medications were included for analyses.

*Number of youth captured;

†Only 250.xx billing codes were used

‡The use of ratio was inspired by Klompas et al.'s study [24]

§Include metformin, sulfonylurea, GLP-1 receptor agonists, thiazolidinediones and other hypoglycemic agents, but not insulin and glucagon.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 5

Performance of type-sensitive algorithms by race/ethnicity applied to the true cases (N=537) in the Carolina Data Warehouse for Health in 2011

	diabetes algorithms	N*			Sensitivity (%)			Specificity (%)			PPV (%)		
		Total	NHW	Other	Total	NHW	Other	Total	NHW	Other	Total	NHW	Other
Type 1 diabetes	≥1 type 1 codes†	443	288	155	97.0	97.1	96.9	62.1	62.5	61.8	88.7	92.7	81.3
	≥2 type 1 codes	415	276	139	93.6	94.2	92.3	72.7	69.6	75.0	91.3	93.8	86.3
	0 type 2 codes and ≥1 type 1 codes	141	103	38	34.6	37.1	29.2	99.2	98.2	100.0	99.3	99.0	100.0
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.3†‡	413	277	136	96.8	96.7	96.9	84.1	80.4	86.8	94.9	96.0	92.6
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.4	402	273	129	96.3	96.7	95.4	90.9	87.5	93.4	97.0	97.4	96.1
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5	397	269	128	95.6	96.0	94.6	92.4	91.1	93.4	97.5	98.1	96.1
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.6	383	262	121	93.3	93.8	92.3	96.2	92.9	98.7	98.7	98.5	99.2
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and insulin use OR ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and glucagon use	372	251	121	89.9	89.8	90.0	93.9	92.9	94.7	97.8	98.4	96.7
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and insulin use OR ratio of type 1 to type 1 and type 2 codes ≥ 0.5 and glucagon use OR DAA positive OR C-peptide negative	378	252	126	90.4	90.2	90.8	90.9	92.9	89.5	96.8	98.4	93.7
	Ratio of type 1 to type 1 and type 2 codes ≥ 0.5 OR ≥1 type 1 code and glucagon and no oral medications § OR ≥1 type 1 code and insulin and no oral medications § OR DAA positive OR C-peptide negative	412	276	136	96.5	96.7	96.2	84.1	82.1	85.5	94.9	96.4	91.9
Type 2 diabetes	≥1 type 2 codes	375	219	156	91.9	89.3	93.1	34.4	36.0	31.1	21.1	11.4	34.6
	≥1 type 2 codes and 0 type 1 codes and no insulin use	57	23	34	46.5	35.7	51.7	96.2	95.7	97.3	70.2	43.5	88.2
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.3	151	75	76	91.9	89.3	93.1	84.0	83.5	85.1	52.3	33.3	71.1
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.4	138	65	73	91.9	89.3	93.1	86.9	86.8	87.2	57.2	38.5	74.0
	Ratio of type 2 to type 1 and type 2 codes ≥ 0.5	128	58	70	88.4	82.1	91.4	88.5	88.4	88.5	59.4	39.7	75.7

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

diabetes algorithms	N*			Sensitivity (%)			Specificity (%)			PPV (%)		
	Total	NHW	Other	Total	NHW	Other	Total	NHW	Other	Total	NHW	Other
Ratio of type 2 to type 1 and type 2 codes ≥ 0.6	116	51	65	84.9	75.0	89.7	90.5	90.1	91.2	62.9	41.2	80.0
0 type 1 codes and ≥3 type 2 codes and oral medications §	22	5	17	25.6	17.9	29.3	100.0	100.0	100.0	100.0	100.0	100.0
Ratio of type 2 to type 1 and type 2 codes ≥ 0.4 and oral medications §	78	27	51	70.9	71.4	70.7	96.2	97.7	93.2	78.2	74.1	80.4

Abbreviation: DAA, diabetes auto-antibodies; NHW, non-Hispanic white; PPV, positive predictive value.

True diabetes cases (N=537) and type confirmed by medical record review established our "gold standard" for evaluation of algorithms' performance.
Only outpatient medications were included for analyses.

*Number of youth captured;

†Only 250.xx billing codes were used

‡The use of ratio was inspired by Klompas et al.'s study [24]

§Include metformin, sulfonylurea, GLP-1 receptor agonists, thiazolidinediones and other hypoglycemic agents, but not insulin and glucagon.