Automated Detection and Classification of Type 1 Versus Type 2 Diabetes Using Electronic Health Record Data

MICHAEL KLOMPAS, MD, MPH[1]
EMMA EGGLESTON, MD, MPH[1]
JASON MCVETTA[2]

ROSS LAZARUS, MBBS[1]
LINGLING LI, PHD[1]
RICHARD PLATT, MD, MSC[1]

OBJECTIVE—To create surveillance algorithms to detect diabetes and classify type 1 versus type 2 diabetes using structured electronic health record (EHR) data.

RESEARCH DESIGN AND METHODS—We extracted 4 years of data from the EHR of a large, multisite, multispecialty ambulatory practice serving ~700,000 patients. We flagged possible cases of diabetes using laboratory test results, diagnosis codes, and prescriptions. We assessed the sensitivity and positive predictive value of novel combinations of these data to classify type 1 versus type 2 diabetes among 210 individuals. We applied an optimized algorithm to a live, prospective, EHR-based surveillance system and reviewed 100 additional cases for validation.

RESULTS—The diabetes algorithm flagged 43,177 patients. All criteria contributed unique cases: 78% had diabetes diagnosis codes, 66% fulfilled laboratory criteria, and 46% had suggestive prescriptions. The sensitivity and positive predictive value of ICD-9 codes for type 1 diabetes were 26% (95% CI 12–49) and 94% (83–100) for type 1 codes alone; 90% (81–95) and 57% (33–86) for two or more type 1 codes plus any number of type 2 codes. An optimized algorithm incorporating the ratio of type 1 versus type 2 codes, plasma C-peptide and autoantibody levels, and suggestive prescriptions flagged 66 of 66 (100% [96–100]) patients with type 1 diabetes. On validation, the optimized algorithm correctly classified 35 of 36 patients with type 1 diabetes (raw sensitivity, 97% [87–100], population-weighted sensitivity, 65% [36–100], and positive predictive value, 88% [78–98]).

CONCLUSIONS—Algorithms applied to EHR data detect more cases of diabetes than claims codes and reasonably discriminate between type 1 and type 2 diabetes.

Electronic health records (EHRs) are transforming public health surveillance. Systems that can automatically extract, analyze, organize, and communicate EHR data to public health agencies increase the breadth, clinical detail, and timeliness of public health surveillance (1). EHR-based systems have rich potential to improve public health surveillance for diabetes, but little work has been done thus far to characterize the accuracy of raw electronic data for diabetes surveillance or to create custom algorithms to accurately distinguish between type 1 versus type 2 diabetes.

Accurate discrimination between type 1 and type 2 diabetes is critical given the different pathophysiology, epidemiology, prevention, management, and prognosis of these two diseases (2–7). Traditional public health surveillance systems either do not distinguish between these two conditions at all or rely on self-reports to make the distinction. In addition, most existing public health surveys are too small to meaningfully track low-prevalence conditions such as type 1 in general and type 2 in youth.

We hypothesized that EHR data could substantially enrich diabetes surveillance by facilitating continuous evaluation of very large populations and leveraging clinical data to distinguish between type 1 and type 2 diabetes. We

therefore sought to develop EHR-based surveillance algorithms to detect and classify type 1 versus type 2 diabetes and then apply them to a live, prospective, EHR-based surveillance system to test their performance.

RESEARCH DESIGN AND METHODS

Setting and data sources
We used retrospective EHR data from Atrius Health to develop diabetes surveillance algorithms. Atrius Health is a large, ambulatory, multisite practice based in Eastern Massachusetts that provides primary and specialty care (including pediatrics, internal medicine, and endocrinology) to >700,000 patients of all ages using a single consolidated EHR (Epic Systems, Verona, WI). We extracted encounter data from the EHR on all patients seen in the practice between 1 June 2006 and 30 September 2010. The extract included all diagnosis codes, laboratory test results, and medication prescriptions. We included all patients with at least one encounter in the EHR in our analysis.

Detection of diabetes
We created and applied a surveillance algorithm for diabetes based upon American Diabetes Association (ADA) laboratory diagnostic criteria (8), suggestive medication prescriptions, and ICD-9 codes (Table 1). We required at least two instances of ICD-9 code 250.xx in patients without supporting laboratory or prescription flags. We did not include metformin in the algorithm as it may be prescribed for indications other than overt diabetes, such as polycystic ovarian syndrome and diabetes prevention.

Classification of type 1 versus type 2
We then began an iterative process to distinguish between type 1 and type 2 diabetes within the population of patients flagged by the general diabetes algorithm in Table 1. We started with a "straw man" algorithm designed to coarsely divide the population into pools of patients more

━━

Table 1—*Surveillance algorithm for diabetes*

Any of the following criteria at any time:

1. Hemoglobin A_{1c} ≥6.5%
2. Fasting glucose ≥126 mg/dL
3. Prescription for insulin outside of pregnancy
4. ICD-9 code 250.xx on two or more occasions
5. Prescription for one or more of the following medications:
   glyburide, gliclazide, glipizide, glimepiride
   pioglitazone, rosiglitazone
   repaglinide, nateglinide, meglitinide
   sitagliptin
   exenatide, pramlintide

likely to have type 1 and patients more likely to have type 2. We began with this preliminary algorithm in order to make chart reviews more efficient: type 1 diabetes is so rare compared with type 2 diabetes that random sampling of unselected patients yields very few type 1 cases. The straw man algorithm allowed us to enrich a study population with type 1 diabetes. A case of type 1 was defined as any patient with ICD-9 250.x1 or 250.x3 on two or more occasions, a current prescription for insulin, and no prescriptions for oral hypoglycemics at any time. A case of type 2 was defined as any patient with ICD-9 250.x0 or 250.x2 on two or more occasions or a prescription for an oral hypoglycemic at any time.

We randomly selected 210 charts classified by the straw man algorithm for review: 70 patients classified as type 1, 60 classified as type 2, and 80 left unclassified by the straw man algorithm. Charts were reviewed for diabetes using ADA diagnostic criteria and for diabetes type (8). We used the following rules to assign "true" diabetes type (applied sequentially): endocrinologist diagnosis if available, never on insulin (classify as type 2), C-peptide negative or diabetes autoantibodies present (classify as type 1), currently on insulin but prior history of prolonged treatment with oral hypoglycemic alone (classify as type 2), and nonendocrinologist physician diagnosis.

We then created a series of candidate algorithms based on ICD-9 code frequencies, laboratory test results, and suggestive prescriptions to optimize sensitivity for chart-confirmed type 1 diabetes while maintaining high positive predictive values. Patients who did not fulfill algorithm

criteria for type 1 diabetes were presumptively classified as type 2. The study endocrinologist (E.E.) and internist (M. K.) created the candidate algorithms based on clinical knowledge of diabetes management practices.

We calculated the sensitivity and positive predictive value of all candidate algorithms using inverse-probability weighting to correct for the sampling strategy. We generated 95% CIs for these estimates using Monte Carlo simulations. Specifically, we simulated the number of true type 1 and type 2 patients among reviewed charts for each sampling strata (straw man type 1, straw man type 2, and straw man unclassified) using multinomial distributions and probabilities estimated from the observed data. We repeated this process 1 million times and derived 95% CIs from the resulting 2.5 and 97.5 percentiles. Calculations were executed using SAS version 9.3 (SAS Institute, Cary, NC).

We then created a final algorithm with optimized sensitivity and positive predictive value for type 1 diabetes by combining the candidate algorithms with the highest positive predictive values using "or" statements. We focused the optimized algorithm on sensitivity to type 1 diabetes because even slight misclassification of type 1 patients as type 2 is substantially magnified after weighting for the greater size of the type 2 population and therefore exerts considerable cost in net sensitivity for type 1.

Validation

We validated the optimized algorithm within a live, prospective EHR-based surveillance system based in the same practice that contributed derivation data (1,9). The algorithm was implemented on 1 December 2011 and retroactively applied to all data resident in the surveillance system (1 June 2006 through November 30, 2011). We applied the reference criteria described above to 100 hitherto unreviewed charts: 40 classified by the optimized algorithm as type 1 and 60 classified as type 2. We calculated the raw and population-weighted sensitivity and positive predictive values of the final algorithm.

Comparison of diagnosis codes alone versus the optimized algorithm

We compared the sensitivity and positive predictive value of ICD-9 codes alone versus the optimized algorithm for detection of type 1 diabetes. We compared

algorithm performance by calculating 95% CIs for the differences in algorithm performance using multinomial distributions and Monte Carlo simulations with 1 million repetitions per comparison in a manner analogous to the method described above for calculating 95% CIs for each candidate algorithm.

RESULTS

Diabetes criteria

The diabetes criteria flagged 43,177 individuals. Characteristics of the population and percentages of patients flagged by each criterion are shown in Table 2. On review, diabetes was confirmed in 298 of 310 patients (95% positive predictive value after correcting for sampling). The 12 false-positives included one clinician

Table 2—*Characteristics of patients flagged by the diabetes surveillance algorithm described in Table 1*

	Frequency, *n* (%)
Age-group	
0–19 years	519 (1.2)
20–59 years	18,169 (42)
60–79 years	18,104 (42)
≥80 years	6,264 (15)
Sex	
Male	20,852 (48)
Female	22,325 (52)
Race/ethnicity	
Caucasian	24,746 (57)
African American	5,286 (12)
Asian	1,538 (3.6)
Hispanic	1,268 (2.9)
Unknown or other	10,339 (24)
Method of detection	
Hemoglobin A_{1c} ≥6.5	27,959 (65)
Fasting glucose ≥126 mg/dL	1,647 (3.8)
Hemoglobin A_{1c} ≥6.5 or fasting glucose ≥126 mg/dL	28,492 (66)
Prescription for insulin outside of pregnancy	10,275 (24)
Prescription for an oral hypoglycemic other than metformin	13,727 (32)
Prescription for insulin or oral hypoglycemic other than metformin	20,013 (46)
ICD-9 250.xx on two or more occasions	33,876 (78)

coding error (ICD-9 code 250.xx used for diabetes screening rather than diagnosis), two EHR coding errors (oral glucose tolerance test 2-hour results miscoded as fasting glucoses), three algorithm programming errors (prescriptions for "insulin syringes" parsed as prescriptions for insulin), four patients with gestational diabetes mellitus (pregnancy not coded as a discrete data element in the EHR), one patient with impaired fasting glucose, and one patient with impaired glucose tolerance. Among the 298 patients with chart-confirmed diabetes, 193 had type 2 diabetes, 102 had type 1 diabetes, 1 had diabetes secondary to recurrent pancreatitis, and 2 could not be classified (prescribed insulin but insufficient clinical data in the EHR to clarify the underlying diagnosis). Two patients with latent autoimmune diabetes in adults were included in the count of type 1 patients for the sake of simplicity.

Straw man algorithm
The straw man algorithm for classifying patients as type 1 versus type 2 flagged 1,095 patients as type 1 and 29,410 as type 2. The remaining patients were not classified. Type 1 diabetes was confirmed in 57 of 70 randomly selected type 1 patients (population-weighted sensitivity, 32%; positive predictive value, 81%). Type 2 diabetes was confirmed in 54 of 60 randomly selected type 2 patients (population-weighted sensitivity, 76%; positive predictive value, 90%). Among 80 patients left unclassified by the straw man algorithm, 7 had type 1 diabetes, 65 had type 2 diabetes, 2 had gestational diabetes mellitus, 1 was being prescribed insulin but lacked sufficient data to classify diabetes type, and 5 did not have diabetes.

Candidate algorithms
Candidate algorithms to distinguish type 1 versus type 2 diabetes are presented in Table 3 along with their performance rates and 95% CIs. ICD-9 codes for type 1 diabetes performed variably. Requiring an exclusive history of type 1 codes alone had high positive predictive value for type 1 diabetes (94%) but poor sensitivity (26%). Requiring at least one ICD-9 code for type 1 with any number of type 2 codes increased sensitivity to 90% but lowered positive predictive value to 41%. Requiring at least two ICD-9 codes for type 1 and any number of type 2 codes preserved sensitivity at 90% but increased positive predictive value to 57%. Seeking a plurality of ICD-9 codes for type 1 diabetes (i.e., ratio of type 1 to type 2 ICD-9 codes >50%) had

intermediate sensitivity (63%) but high positive predictive value (94%).

Algorithms based on medication prescriptions performed with similar variability. Presumptively classifying patients as type 1 based upon a prescription for insulin had excellent sensitivity (95%) but poor positive predictive value (23%), indicating that many type 2 patients are treated with insulin. A single false-negative accounted for the less than perfect sensitivity of insulin prescriptions for type 1; the patient was prescribed an insulin pump that was recorded in the EHR as free text rather than as a structured prescription. Likewise, classifying patients without any history of oral hypoglycemics as type 1 was very sensitive (95%) but positive predictive value was low (15%). Creating an exception for metformin increased sensitivity to 98% but lowered positive predictive value to 11%. The one medication with reasonable positive predictive value for type 1 was glucagon (sensitivity, 68%; positive predictive value, 79%). Notably, prescriptions for urine acetone tests had outstanding positive predictive value for type 1 (sensitivity, 33%; positive predictive value, 100%).

Combining ICD-9 and medication criteria increased positive predictive values. For example, a plurality of ICD-9 codes and a prescription for glucagon had a sensitivity of 31% but a positive predictive value of 100%. Similarly, a plurality of type 1 ICD-9 codes and no history of oral hypoglycemic prescriptions had a sensitivity of 59% and a positive predictive value of 99%.

Optimized algorithm
We optimized an algorithm to maximize sensitivity and positive predictive value for type 1 diabetes by combining all the high positive predictive value candidate algorithms using "or" statements (Table 4). The final algorithm sought patients with any of the following: a plurality of ICD-9 codes for type 1 diabetes and a prescription for glucagon, a plurality of ICD-9 codes for type 1 diabetes and a negative history of prescriptions for oral hypoglycemics other than metformin, a negative plasma C-peptide, positive diabetes autoantibody tests, or a prescription for urine acetone test strips. The final algorithm flagged 73 patients, including all 66 patients with type 1 diabetes (raw sensitivity, 100% [95% CI 96–100]; positive predictive value, 90% [82–96]). Correcting for the sampling strategy yielded a net population-weighted sensitivity of 100%

(100–100) and a positive predictive value of 96% (91–99).

Comparison of diagnosis codes alone versus the optimized algorithm
The optimized algorithm was more sensitive for type 1 diabetes compared with surveillance for patients with one or more type 1 diabetes ICD-9 codes and any number of type 2 codes (100% vs. 90%, 10% absolute difference [95% CI 4.9–19]), surveillance for patients with two or more type 1 diabetes ICD-9 codes and any number of type 2 codes (100% vs. 90%, 10% absolute difference [4.9–19]), and surveillance for patients with type 1 diabetes ICD-9 codes alone and no type 2 codes (100% vs. 26%, 74% absolute difference [50–88]). The optimized algorithm's positive predictive value for type 1 diabetes was also high compared with surveillance for one or more type 1 codes (96% vs. 41%, 55% absolute difference [31–72]) or two or more type 1 codes (96% vs. 57%, 39% absolute difference [11–62]), but equivalent to surveillance for patients with type 1 codes alone and no type 2 codes (96% vs. 94%, 2% absolute difference [−4.9 to 7.6]).

Validation and implementation
The final algorithm was applied to a live, prospective surveillance system for validation. A total of 100 hitherto unreviewed charts were evaluated: 40 random charts flagged as type 1 and 60 random charts flagged as type 2. Chart review confirmed 35 of 40 patients flagged as type 1. Of the remaining five, four had type 2 diabetes and one had gestational diabetes mellitus. Notably, three of the four false-positive cases were type 2 patients with a plurality of type 1 ICD-9 codes; these were all longstanding type 2 patients managed for years on insulin alone and labeled as type 1 diabetic in their physicians' notes. Among the 60 patients classified as type 2, 57 were confirmed on chart review. Of the remaining three, one had type 1 diabetes, one had gestational diabetes mellitus, and one was prescribed insulin but had insufficiently detailed notes to determine diabetes type. All told, the final algorithm correctly identified 35 of 36 patients with type 1 diabetes (raw sensitivity, 97% [95% CI 87–100]) and 57 of 61 patients with type 2 diabetes (raw sensitivity, 93% [85–98]). Projecting from these samples to the full study population using inverse probability sample weighting yielded net sensitivity and positive

Table 3—Candidate algorithms to classify patients fulfilling the diabetes criteria in Table 1 as type 1 or type 2

Candidate algorithms (Classify as type 1 if algorithm positive; classify as type 2 if algorithm negative)	Flagged (classified as type 1)	Confirmed as type 1	Corrected sensitivity for type 1 % (95% CI)	Corrected PPV for type 1 % (95% CI)	Not flagged (classified as type 2)	Confirmed as type 2	Corrected sensitivity for type 2 % (95% CI)	Corrected PPV for type 2 % (95% CI)
Age ≤20 years	12	11	14 (3.7–32)	76 (33–100)	198	131	100 (99–100)	86 (80–92)
BMI ≤25 kg/m²	74	33	61 (32–80)	14 (4.8–27)	136	98	75 (65–83)	90 (84–96)
Triglycerides ≤200 mg/dL	171	61	97 (93–99)	8.5 (4.5–14)	39	34	25 (16–35)	99 (98–100)
High density lipoproteins ≥60 mg/dL	53	31	68 (41–83)	21 (8.1–38)	157	112	83 (74–90)	91 (86–95)
ICD-9s for type 1 alone	10	8	26 (12–49)	94 (83–100)	200	130	100 (100–100)	87 (81–92)
≥1 ICD-9 for type 1 (and any number of type 2 ICD-9s)	70	49	90 (81–95)	41 (22–65)	140	111	90 (83–96)	91 (84–96)
≥2 ICD-9s for type 1 (and any number of type 2 ICD-9s)	65	49	90 (81–95)	57 (33–86)	145	116	95 (89–99)	91 (85–96)
Ratio of type 1 ICD-9s to type 2 ICD-9s >0.5	68	61	63 (38–99)	94 (88–98)	142	125	100 (99–100)	90 (83–95)
C-peptide <0.8	8	8	43 (4.7–66)	100 (100–100)	202	132	100 (100–100)	87 (81–92)
Diabetes autoantibodies detected	3	3	1.7 (0–4.7)	100 (100–100)	207	132	100 (100–100)	86 (77–91)
Prescription for insulin	101	64	95 (83–100)	23 (12–37)	109	99	77 (67–85)	99 (86–97)
Prescription for metformin	65	9	5.1 (2.0–11)	0.8 (0.3–1.4)	145	77	50 (40–60)	76 (66–85)
Prescription for sulfonylurea	45	3	1.7 (0–4.7)	0.3 (0–0.7)	165	90	55 (44–65)	77 (67–86)
No record of any oral hypoglycemics	130	62	95 (89–98)	15 (7.6–23)	80	70	64 (54–73)	99 (97–100)
No record of any oral hypoglycemics other than metformin	160	62	98 (94–100)	11 (6.1–18)	50	46	49 (38–59)	100 (99–100)
Prescription for glucagon	33	31	68 (41–83)	79 (43–100)	177	131	100 (100–100)	91 (86–95)
Prescription for urine acetone test strips	13	13	33 (7.4–59)	100 (100–100)	197	132	100 (100–100)	88 (82–93)
Ratio of type 1 ICD-9s to type 2 ICD-9s >0.5 and prescription for insulin	66	59	58 (34–96)	94 (87–98)	144	125	100 (99–100)	89 (83–95)
Ratio of type 1 ICD-9s to type 2 ICD-9s >0.5 and prescription for glucagon	26	26	31 (15–57)	100 (100–100)	184	132	100 (100–100)	88 (81–93)
Ratio of type 1 ICD-9s to type 2 ICD-9s >0.5 and no record of any oral hypoglycemics	54	53	59 (35–93)	99 (97–100)	156	131	100 (100–100)	89 (83–95)
Ratio of type 1 ICD-9s to type 2 ICD-9s >0.5 and no record of oral hypoglycemics other than metformin	63	58	61 (37–97)	96 (91–99)	147	127	100 (100–100)	90 (86–97)

Algorithm-positive patients presumptively classified as type 1; algorithm-negative patients presumptively classified as type 2. PPV, positive predictive value.

Table 4—Optimized algorithm to detect type 1 diabetes among patients fulfilling the criteria for diabetes in Table 1

|	Flagged (classified as type 1)	Confirmed as type 1	Sensitivity for type 1 (corrected for sample weights) % (95% CI)	PPV for type 1 (corrected for sample weights) % (95% CI)	Not flagged (classified as type 2)	Confirmed as type 2	Sensitivity for type 2 (corrected for sample weights) % (95% CI)	PPV for type 2 (corrected for sample weights) % (95% CI)
Classify as type 1 if any of the following present:								
1. Ratio of type 1 to type 2 ICD-9 codes >0.5 and prescription for glucagon								
2. Ratio of type 1 to type 2 ICD-9 codes >0.5 and no record of an oral hypoglycemic other than metformin								
3. C-peptide negative								
4. Diabetes autoantibodies positive								
5. Prescription for urine acetone test strips								
Classify as type 2 if algorithm negative								
Derivation (n = 210)	73	66	100 (100–100)	96 (91–99)	137	125	100 (99–100)	92 (86–97)
Validation (n = 100)	40	35	65 (36–100)	88 (78–98)	60	57	100 (99–100)	95 (88–100)
Derivation and validation combined (n = 310)	113	101	86 (67–100)	89 (83–95)	197	182	100 (99–100)	92 (88–96)

Algorithm-negative patients presumptively classified as type 2.

predictive values of 65% (95% CI 36–100) and 88% (78–98) for type 1 diabetes, and net sensitivity and positive predictive values of 100% (99–100) and 95% (88–100) for type 2 diabetes. Figure 1 depicts a heat map of the prevalence of type 2 diabetes by zip code that was automatically generated by our EHR-based live surveillance system using the final algorithm.

CONCLUSIONS—This work demonstrates the feasibility and potential utility of EHR data for automated diabetes surveillance. Leveraging the full array of data captured by EHRs, including laboratory test results and prescriptions in addition to diagnosis codes, increases both the sensitivity and granularity of surveillance compared with diagnosis codes alone. Our integrated algorithm, including laboratory, prescription, and ICD-9 criteria, captured more patients with diabetes than any one criterion alone; diagnosis codes flagged 78% of the final population, laboratory tests flagged 66%, and prescriptions for insulin or oral hypoglycemics (excluding metformin) flagged 46%. A combination algorithm seeking patients with a plurality of type 1 diagnosis codes and suggestive prescriptions, diagnostic laboratory test results (negative plasma C-peptide or positive diabetes autoantibodies), or prescriptions for urine acetone test strips classified patients as type 1 versus type 2 reasonably well.

The application of these surveillance algorithms to EHR data has the potential to provide timely, clinically detailed information on large numbers of patients at low marginal cost per patient. Furthermore, in contrast to claims-based surveillance, EHR data streams can also provide rich contextual data about patients, including their demographics, clinical parameters (e.g., blood pressure, BMI, hemoglobin A_{1c}, and lipids), medications, patterns of care, and how things are changing over time. EHR-based surveillance systems capable of extracting, analyzing, and reporting these indicators to public health departments in almost real time are emerging (1). EHRs are likely to become increasingly important sources for public health surveillance as more practices adopt EHRs and vendors begin to add functionality for public health surveillance. The federal government's "meaningful use" incentives provide substantial financial encouragement for both of these goals (10).

Very few public health surveys currently distinguish between type 1 and
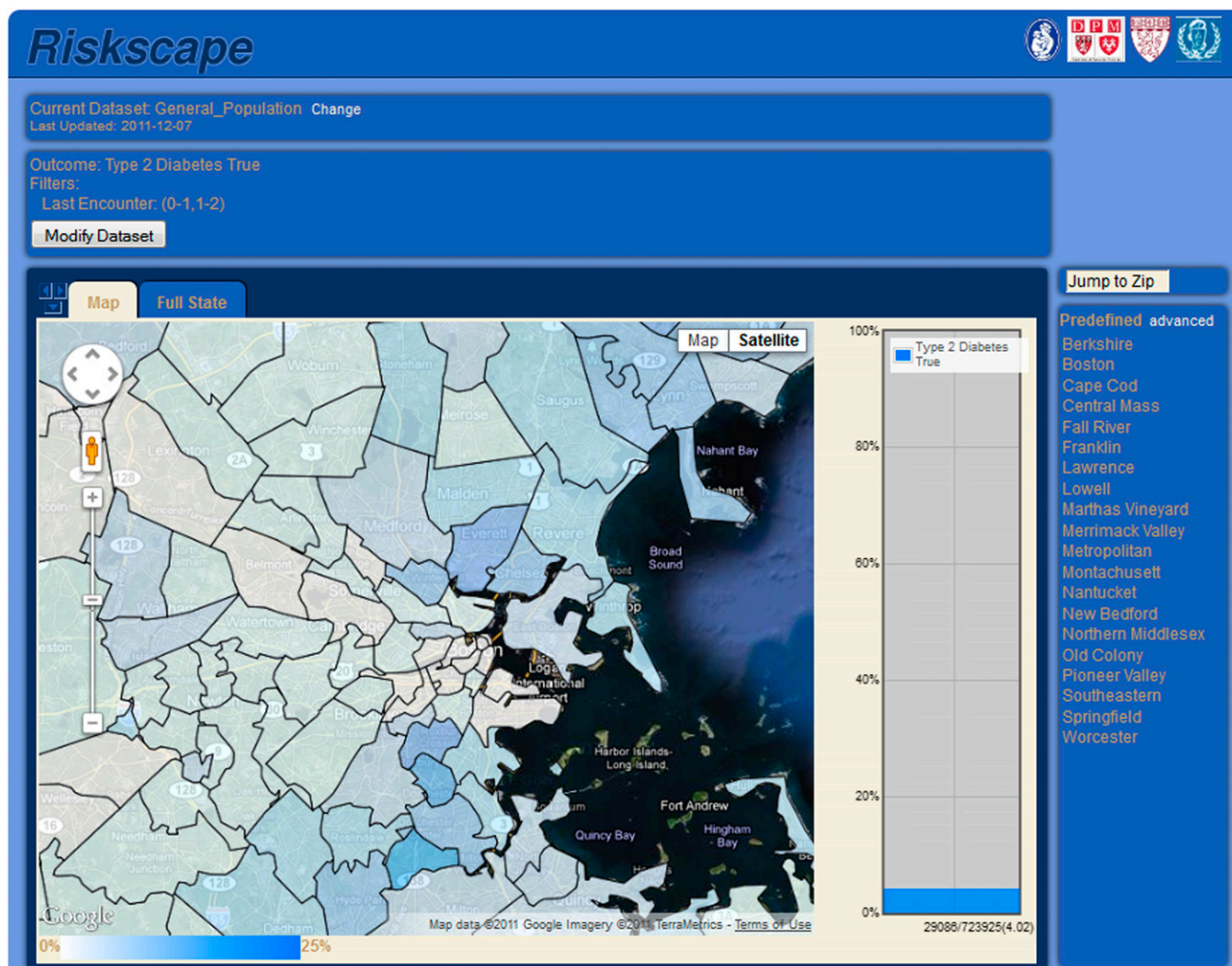
Figure 1—*Screenshot of the Electronic medical record Support for Public health (ESP) live EHR-based public health surveillance and reporting system. The map depicts the prevalence of type 2 diabetes by zip code among Atrius Health patients in Eastern Massachusetts. The system automatically detects diabetes and classifies patients as type 1 versus type 2 using data refreshed and analyzed nightly using the algorithms described in this article.*

type 2 diabetes despite their very different demographics, risk factors, prevention strategies, and management pathways. The Canadian Community Health Survey imputes diabetes type based on self-reported age at diagnosis, pregnancy history, and medication usage (11). Some U.S. states add an extra question on diabetes type to the Behavioral Risk Factor Surveillance System (12). Both of these classifications, however, depend upon self-reports rather than clinical and laboratory data. Vanderloo et al. (13) proposed a claims-based classification scheme for children based upon age, ethnicity, and medication history, but the age and ethnicity criteria limit the generalizability of their algorithm for general population surveillance. Finally, some research surveys, such as the SEARCH for Diabetes in Youth, gather detailed

physiological data to help classify patients as type 1 versus type 2 (14). This approach provides invaluable insights but expense and complexity preclude its use for routine public health operations. EHRs' clinically rich data and efficient coverage of large populations constitute an additional, attractive option to improve public health capacity to routinely track type 1 and type 2 diabetes.

The persistent sources of error in our algorithms provide a telling window into the limitations of EHR data for public health surveillance. Sources of error included physician miscoding (type 2 diagnosis codes assigned to patients with type 1 diabetes, diabetes diagnosis codes used for screening rather than frank disease, and prescription for insulin pump recorded as free text rather than as structured data), EHR miscoding (oral glucose

tolerance tests coded as fasting glucoses), and algorithm programming errors (prescription for "insulin syringe" incorrectly parsed as prescription for insulin and failure to recognize some patients as pregnant, leading to misclassification of gestational diabetes mellitus as frank diabetes).

Imperfections of EHR data also account for the large discrepancy in the performance of the classification algorithm in derivation versus validation. Algorithm sensitivity dropped from 100% in derivation to 65% on validation. This decrease was entirely attributable to misclassification of a single case of type 1 diabetes as type 2 diabetes. The patient in question was a 45-year-old woman with type 1 diabetes since age 11. The EHR data on record included 33 encounters with ICD-9 codes for diabetes but 27 of the 33 encounters were miscoded as type

2 diabetes. In theory, the algorithm could still have classified her as type 1 since she was prescribed urine acetone test strips (as well as insulin and glucagon). These prescriptions were not recorded in the EHR prescription fields, however, because they were provided by an endocrinologist outside of our study practice.

The marked drop in sensitivity from a single misclassification demonstrates the challenge of electronic surveillance for rare outcomes in large populations. The raw sensitivity of our algorithm on chart review appeared very good (35 of 36, 97% [95% CI 87–100]), but correcting for the sampling strategy multiplied the effect of the missing case many-fold. Inverse probability sample weighting increased the estimate of missed type 1 cases by a factor proportionate to the size of the type 2 population. The net population-weighted sensitivity of the final algorithm for type 1 diabetes therefore dropped to 65% (36–100).

In addition, this missed case highlights the ongoing risk of misclassification in surveillance studies using EHR data due to miscoding and the balkanization of patients' data between different EHRs that do not participate in data exchange. Two factors, however, may mitigate these sources of error over time. First, our live EHR-based surveillance system reclassifies patients anew every time fresh EHR data are uploaded. It is possible that this false-positive case may be reclassified correctly in the future if her EHR-affiliated providers begin coding more accurately or enter a suggestive prescription. Likewise, health information exchanges that consolidate data from multiple EHR systems across a region may help diminish missing data (15). A recent survey of U.S. hospitals found that 10.7% currently exchange data with unaffiliated providers (16). Current federal funding initiatives for states to develop regional information exchanges and meaningful use incentives for clinicians to exchange data are likely to encourage further growth of electronic information sharing between unaffiliated organizations in years to come (17).

Limitations of our study include the risk of incorrect reference classifications and algorithm overtraining. Our reference standard classifications were made using retrospective chart review data rather than comprehensive physiological studies. We attempted to optimize reference classifications by setting out a hierarchy of classification rules that gave greatest weight to endocrinologist diagnoses,

consistent laboratory studies, and suggestive medication prescription histories rather than accepting any diagnosis from any clinician at face value. Nonetheless, misclassifications could have occurred due to endocrinologist error, false-positive or -negative laboratory test results, incomplete clinical histories and prescription records, or primary care physician errors. Laboratory data in particular need to be interpreted with caution; some type 1 patients, for example, can have negative antibodies and others may have remnant β-cell function reflected in detectable C-peptide on standard clinical assays (18). Fewer than 5% of our classifications, however, were based upon C-peptide or autoantibody measures. It would be informative to validate and possibly refine our suggested algorithm for type 1 versus type 2 diabetes using data from studies with comprehensive physiological measurements on all patients (although even these studies sometimes suffer from controversy regarding their classification schemes). In addition, our classification algorithm is heavily dependent upon clinicians' coding, testing, and prescribing decisions. Even if the algorithm is optimized using physiological data as reference standard, there is still a future risk of misclassifying patients if clinicians make incorrect diagnoses and then code and prescribe accordingly.

The algorithms in this study were developed using data from a single EHR serving a single practice. Algorithm performance may be different if implemented in different settings with different clinical practice and coding patterns (19). The algorithm attempts to mitigate this risk by allowing multiple criteria for type 1 diabetes, thereby increasing the probability that local practice and coding styles will still trigger the algorithm. The algorithms need to be validated in new settings, however, to test their robustness and portability.

We created the algorithms in this paper for EHR-based surveillance systems but they may also be useful for the analysis of claims-based datasets. Claims data typically include many of the same elements as EHRs, including diagnosis codes, prescriptions, and laboratory tests, but usually only indicate that a test was performed, not the specific result. A major advantage of claims datasets, however, is that they include data on all of each patient's encounters, not just those within a given practice. This feature

could mitigate the problem of missing data due to patients seeking care from multiple practices with unconnected EHR systems. Many of the candidate algorithms in Table 3 are suitable for application to claims datasets already insofar as they only include diagnosis codes and/or medication prescriptions. In addition, we assessed two potential adaptations of our final algorithm suitable for claims data. In the first adaptation, we retained the full algorithm but simply looked to see whether C-peptide and/or diabetes autoantibody tests were performed rather than requiring a specific result. This preserved sensitivity but lowered positive predictive value for type 1 diabetes (100 and 79%, respectively). In the second adaptation, we removed the C-peptide and diabetes autoantibody criteria altogether. This decreased sensitivity for type 1 diabetes but preserved a high positive predictive value (82 and 95%, respectively).

In sum, we demonstrate the utility of EHR-based algorithms to detect and classify patients with diabetes. Integration of these algorithms into EHR-based surveillance systems could provide a rich resource for automated public health surveillance, practice management, and patient recruitment for clinical studies. In addition, we identify current sources of misclassification that limit algorithm performance. Knowing these sources of error could help practices and/or public health agencies design strategies to improve data quality and/or further refine our algorithm to mitigate these errors. Further work validating, enhancing, and applying these algorithms to public health surveillance systems is merited.

of the data and the accuracy of the data analysis.

Parts of this study were presented in poster form at the 72nd Scientific Sessions of the American Diabetes Association, Philadelphia, Pennsylvania, 8–12 June 2012.

References

1. 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–S162

2. Vehik K, Hamman RF, Lezotte D, et al. Increasing incidence of type 1 diabetes in 0- to 17-year-old Colorado youth. Diabetes Care 2007;30:503–509

3. Borchers AT, Uibo R, Gershwin ME. The geoepidemiology of type 1 diabetes. Autoimmun Rev 2010;9:A355–A365

4. Dabelea D, Bell RA, D'Agostino RB Jr, et al.; Writing Group for the SEARCH for Diabetes in Youth Study Group. Incidence of diabetes in youth in the United States. JAMA 2007;297:2716–2724

5. TODAY Study Group; Zeitler P, Hirst K, Pyle L, et al. A clinical trial to maintain glycemic control in youth with type 2 diabetes. N Engl J Med 2012;366:2247–2256

6. Petitti DB, Klingensmith GJ, Bell RA, et al. Glycemic control in youth with diabetes: the SEARCH for diabetes in Youth Study. J Pediatr 2009;155:668–672.e1–3

7. 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–e1266

8. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2011;34(Suppl. 1):S62–S69

9. Lazarus R, Klompas M, Campion FX, et al. Electronic Support for Public Health: validated case finding and reporting for notifiable diseases using electronic medical data. J Am Med Inform Assoc 2009;16:18–24

10. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010;363:501–504

11. Ng E, Dasgupta K, Johnson JA. An algorithm to differentiate diabetic respondents in the Canadian Community Health Survey. Health Rep 2008;19:71–79

12. Chowdhury P, Balluz L, Town M, et al.; Centers for Disease Control and Prevention (CDC). Surveillance of certain health behaviors and conditions among states and selected local areas - Behavioral Risk Factor Surveillance System, United States, 2007. MMWR Surveill Summ 2010;59:1–220

13. 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

14. Dabelea D, Pihoker C, Talton JW, et al.; SEARCH for Diabetes in Youth Study. Etiological approach to characterization of diabetes type: the SEARCH for Diabetes in Youth Study. Diabetes Care 2011;34: 1628–1633

15. Adler-Milstein J, Bates DW, Jha AKUS. U.S. Regional health information organizations: progress and challenges. Health Aff (Millwood) 2009;28:483–492

16. Adler-Milstein J, DesRoches CM, Jha AK. Health information exchange among US hospitals. Am J Manag Care 2011;17: 761–768

17. Adler-Milstein J, Jha AK. Sharing clinical data electronically: a critical challenge for fixing the health care system. JAMA 2012; 307:1695–1696

18. Greenbaum CJ. Dead or alive? Diabetes Care 2012;35:459–460

19. Amed S, Vanderloo SE, Metzger D, et al. Validation of diabetes case definitions using administrative claims data. Diabet Med 2011;28:424–427