

A comparison of phenotype definitions for diabetes mellitus

Rachel L Richesson,¹ Shelley A Rusincovitch,² Douglas Wixted,³ Bryan C Batch,⁴ Mark N Feinglos,⁴ Marie Lynn Miranda,⁵ W Ed Hammond,^{2,6} Robert M Califf,^{3,7} Susan E Spratt⁴

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/amiajnl-2013-001952>).

¹Duke University School of Nursing, Durham, North Carolina, USA

²Applied Informatics Research, Duke Health Technology Solutions, Durham, North Carolina, USA

³Duke Translational Medicine Institute, Durham, North Carolina, USA

⁴Division of Endocrinology, Metabolism and Nutrition, Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA

⁵Department of Pediatrics, School of Natural Resources and Environment, University of Michigan, Ann Arbor, Michigan, USA

⁶Duke Center for Health Informatics, Durham, North Carolina, USA

⁷Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA

Correspondence to

Dr Rachel L Richesson, Duke University School of Nursing, 2007 Pearson Building, 307 Trent Drive, Durham, NC 27710, USA; rachel.richesson@duke.edu

Received 19 April 2013

Revised 30 July 2013

Accepted 20 August 2013

Published Online First

11 September 2013

ABSTRACT

Objective This study compares the yield and characteristics of diabetes cohorts identified using heterogeneous phenotype definitions.

Materials and methods Inclusion criteria from seven diabetes phenotype definitions were translated into query algorithms and applied to a population (n=173 503) of adult patients from Duke University Health System. The numbers of patients meeting criteria for each definition and component (diagnosis, diabetes-associated medications, and laboratory results) were compared.

Results Three phenotype definitions based heavily on ICD-9-CM codes identified 9–11% of the patient population. A broad definition for the Durham Diabetes Coalition included additional criteria and identified 13%. The electronic medical records and genomics, NYC A1c Registry, and diabetes-associated medications definitions, which have restricted or no ICD-9-CM criteria, identified the smallest proportions of patients (7%). The demographic characteristics for all seven phenotype definitions were similar (56–57% women, mean age range 56–57 years). The NYC A1c Registry definition had higher average patient encounters (54) than the other definitions (range 44–48) and the reference population (20) over the 5-year observation period. The concordance between populations returned by different phenotype definitions ranged from 50 to 86%. Overall, more patients met ICD-9-CM and laboratory criteria than medication criteria, but the number of patients that met abnormal laboratory criteria exclusively was greater than the numbers meeting diagnostic or medication data exclusively.

Discussion Differences across phenotype definitions can potentially affect their application in healthcare organizations and the subsequent interpretation of data.

Conclusions Further research focused on defining the clinical characteristics of standard diabetes cohorts is important to identify appropriate phenotype definitions for health, policy, and research.

INTRODUCTION

The ability to identify people with diabetes across healthcare organizations by using a common definition has value for clinical quality, health improvement, and research. Registries have been shown to improve care in diabetes, and are the cornerstone of the chronic disease care model.^{1 2} Standard phenotype definitions can enable direct comparison of population characteristics, risk factors, and complications, allowing decision makers to identify and target people for interventions demonstrated in similar

populations. Furthermore, standard phenotype definitions can streamline the development of patient registries from healthcare data, and enable consistent inclusion criteria to support regional surveillance and the identification of rare disease complications. An understanding of the populations generated from various phenotype definitions will inform standard methods for identifying diabetes cohorts, facilitate the rapid generation of patient registries and research datasets with uniform sampling criteria, and enable comparative and aggregate analysis. This descriptive study presents and compares the size and characteristics of patient populations retrieved using different phenotype definitions adopted from prominent diabetes registries and research networks, a large community intervention program in our county, and federal reporting standards.

BACKGROUND AND SIGNIFICANCE

Diabetes diagnosis and management

Diabetes is a complex disease with multiple subtypes associated with different etiologies, diagnostic indicators, and clinical management strategies. Type 2 diabetes mellitus (T2DM) is the most common (95%) type of diabetes in the USA and can be treated with diet and exercise, oral medication, or insulin. Type 1 diabetes mellitus (T1DM) is less common and requires treatment with insulin. Rare types of diabetes result from drug interactions, genetic defects of beta cell or insulin action function, pancreatic disorders, and inherited endocrine disorders. All types of diabetes manifest in high blood glucose, and laboratory values are the primary means for diagnosis and management.³

Diabetes-relevant data available for electronic health record-based phenotyping

Data from three domains (International Classification of Disease, revision 9, clinical modification (ICD-9-CM) coded diagnoses, laboratory test results, and medication data) in varying combinations and thresholds constitute most phenotype definitions used for diabetes cohort identification. The ICD-9-CM coding system has more than 20 broad codes (and scores of higher precision codes) suggestive or indicative of diabetes (presented in the diabetes phenotype definition shared on Phenotype KnowledgeBase),⁴ and is a critical component of most queries and phenotypes. However, ICD-9-CM has been shown to be insufficient for capturing etiology, subtypes, or all cases of diabetes.^{5–7}

Diabetes-related medications are often included in phenotype definitions because medication data

To cite: Richesson RL, Rusincovitch SA, Wixted D, et al. *J Am Med Inform Assoc* 2013;20:e319–e326.

are generally available electronically, and the presence of specific medications can help distinguish patients with T1DM versus T2DM. Patients with T1DM require insulin to survive. T2DM, on the other hand, can be treated with diet and exercise, oral medications (eg, metformin), and/or insulin. With rare exceptions (eg, early diagnosis, latent auto-immune diabetes), diabetes patients who are not on insulin have type 2. However, some T2DM-associated medications are used for other disorders (eg, polycystic ovary syndrome, non-alcoholic fatty liver disease, impaired glucose tolerance for patients at higher risk of developing diabetes), limiting the value of medications alone as indicators of diabetes or specific subtypes.

Laboratory tests are routinely used for the diagnosis and management of diabetes and are readily available electronically. Most phenotypes include the concept of ‘abnormal glucose’, typically defined as random glucose greater than or equal to 200 mg/dl, fasting glucose greater than or equal to 126 mg/dl, or a hemoglobin A1c (HbA1c) laboratory test result of 6.5% or more. However, there are variations in the types and numbers of test observations required across different phenotype definitions⁸ that potentially limit the comparability of retrieved populations. Also, because the standardized logical observation identifiers names and codes coding system for laboratory test names has not been universally adopted, health systems have different capacities to identify test details easily (eg, random vs fasting glucose) from electronic health record (EHR) data.⁹

EHR-driven phenotype definitions may include other data sources, for example, clinical narratives using natural language processing specifications developed as part of the electronic medical records and genomics (eMERGE) consortium.¹⁰ In addition, any phenotype definition that is developed in a particular healthcare setting has additional explicit or implicit criteria (eg, patient age, enrollment in health plan, utilization of services, observation periods, encounter criteria, medical and billing coding practices) that affect the patient cohorts returned.

Existing phenotype definitions for diabetes cohort identification

The Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) diabetes registry¹¹ and the eMERGE consortium¹² are two prominent diabetes sampling efforts that reflect regional and multi-institutional communities. Both of these efforts engaged dozens of clinical experts to develop meaningful and practical definitions for identifying diabetes cohorts across multiple healthcare systems. The SUPREME-DM phenotype was developed for epidemiological study and public health intervention in (type 1 and 2) diabetes patients, and applied to 11 healthcare systems from the HMO Research Network (HMORN).^{13 14} The SUPREME-DM phenotype definition uses diagnosis, medication, and laboratory criteria that have been previously validated using manual chart review.¹⁵

The eMERGE definition for diabetes was developed by clinical investigators at Northwestern University to identify T2DM patients for genotype–phenotype correlation studies, and was designed additionally to identify and exclude T1DM cases.¹⁶ The eMERGE phenotype uses diagnosis codes supplemented by relevant laboratory results and medication prescriptions. The algorithm was validated at three of the five institutions participating in eMERGE, demonstrating a 98% positive predictive value for the identification of T2DM patients compared with clinician review.¹⁷ The eMERGE phenotype has since been used for positive and negative T2DM classification in a population-based study.¹⁸

A phenotype definition that is based exclusively on ICD-9-CM codes is the Chronic Condition Data Warehouse (CCW), developed by the Centers for Medicare and Medicaid Services (CMS). The CMS CCW provides disease-specific groups of ICD-9-CM codes for 26 conditions in a research database designed to make claims and assessment data from Medicare and Medicaid beneficiaries readily available for health services researchers.¹⁹ The CMS CCW definition for diabetes includes 68 ICD-9-CM codes for primary and secondary diabetes and for all types and complications. The CMS CCW diabetes algorithm was evaluated using records from participants of the NHANES I Epidemiologic Follow-up Study linked to Medicare claims data, and detected 69% of pre-existing diabetes cases during a 2-year observation period, and 77% of cases using a longer (9-year) observation period.²⁰

A new phenotype definition for diabetes has been developed to support the Durham Diabetes Coalition (DDC), a community-based demonstration project funded by the Bristol Meyers Squibb Foundation to identify and treat T2DM and reduce diabetes-related morbidity and mortality across Durham County, North Carolina. A grant from the Centers for Medicare and Medicaid Innovation is supporting expansion of the DDC model in Durham and three other counties, forming the Southeastern Diabetes Initiative. The DDC and Southeastern Diabetes Initiative are typical of many community programs across the USA that are struggling to define and implement useful strategies for identifying diabetes populations for public health interventions to optimize health outcomes. Informed in part by the analyses presented in this study, the phenotype definition for DDC has been iteratively expanded in order to represent a broad set of criteria that will identify the most diabetes cases, using ICD-9-CM codes, laboratory tests, or diabetes-associated medication data available in EHR. Because the focus of the DDC is T2DM, the DDC phenotype was developed with a preference for T2DM but without aggressive strategy to exclude T1DM. ICD-9-CM codes 250.x1 and 250.x3 (specific to T1DM) are not included in the DDC definition.

A phenotype definition based only on medications indicative or suggestive of diabetes was developed to identify the feasibility of using medications as a sole criterion to identify cases of T2DM from EHR. Because there are variations in disease management, the diabetes-associated medications phenotype definition includes all medications used to treat diabetes, including those that are standard for the management of T2DM (eg, metformin, thiazolidinediones and exenatide), and others that are used in both type 1 and 2 diabetes (eg, insulin, which is always prescribed for patients with T1DM). Because there are more patients with T2DM, the number of T2DM patients on insulin is greater than the number of T1DM patients on insulin.

We include a phenotype defined exclusively by HbA1c values, based on the A1c Registry of New York City (NYC), established in 2005 and populated by mandated laboratory reporting of HbA1c test results to the NYC Board of Health.²¹ The NYC A1c Registry is the first laboratory data-driven registry to address a non-communicable chronic disease. We define the criterion for abnormal HbA1c for the ‘NYC A1c Registry phenotype’ as an HbA1c value greater than or equal to 6.5%, based upon clinical guidelines.³ A comparison of the seven phenotype definitions is illustrated in table 1.

This study compares the yield and characteristics of patient populations returned by applying the above seven diabetes phenotype definitions in our healthcare system data warehouse.

MATERIALS AND METHODS

We applied inclusion and exclusion criteria from the seven phenotype definitions as query algorithms to select sets of

Table 1 Data domain criteria used in selected phenotype definitions

Phenotype definitions:	Data domain criteria							
	ICD-9-CM 250.xx	ICD-9-CM 250.x0 and 250.x2 (excludes type 1 specific codes)	Expanded ICD-9-CM Codes (249.xx, 357.2, 362.0x, 366.41)	HbA1c	Fasting glucose	Random glucose	Abnormal OGTT	Diabetes-associated medications*
ICD-9-CM 250.xx	●							
CMS CCW	▲*//		▲*//					
NYC A1c Registry				●				
Diabetes-associated medications								●
DDC		▲	▲	▲//	▲//	▲//	▲//	▲
SUPREME-DM	▲*//		▲*//	▲//	▲//	▲//	▲	▲
eMERGE†		●*//		▲	▲	▲		▲

*Medications vary by phenotype definition and are listed for each in the supplementary appendix (available online only).

†The eMERGE phenotype definition consists of five case scenarios with varying combinations of criteria. Any instance of type 1 specific codes (ie, 250.x1, 250.x3) results in the exclusion of the patient.

●=Sole criteria.

▲=Optional criteria, one of many.

*=Distinction made between inpatient and outpatient context.

// = Distinction made for multiple instances and/or time points.

CMS CCW, Centers for Medicare and Medicaid Services Chronic Condition Data Warehouse; DDC, Durham Diabetes Coalition; eMERGE, electronic medical records and genomics; HbA1c, hemoglobin A1c; ICD-9-CM, International Classification of Disease, revision 9, clinical modification; NYC, New York City; OGTT, oral glucose tolerance test; SUPREME-DM, Surveillance, Prevention, and Management of Diabetes Mellitus.

patients from the Duke Medicine Enterprise Data Warehouse (EDW). The EDW encompasses data generated in the care of patients within the Duke University Health System (DUHS), including three hospitals and an extensive network of outpatient clinics. More than 4.3 million patients and 35 million encounters are represented in the EDW.

The phenotype definitions applied include: (1) the 250.xx ICD-9-CM diabetes codes; (2) the CMS CCW definition (a full set of 68 ICD-9-CM diabetes codes); (3) NYC A1c Registry criteria (HbA1c value $\geq 6.5\%$); (4) a diabetes-associated medications definition; and the diabetes phenotype definitions for the (5) DDC; (6) SUPREME-DM; and (7) eMERGE projects. Inclusion criteria from the different definitions were translated to align with data availability and representation in the Duke Medicine EDW. A senior data warehouse analyst (SAR) interpreted and translated the concepts specified in each definition into metadata and human-readable pseudocode that defined variables and rules (eg, number of observations, temporal conditions), essentially providing a detailed map for programming the data extraction code. The pseudocode was independently reviewed by an informatician (DW) for accuracy. The details of the phenotype definitions (including specific ICD-9-CM codes and medication names), metadata, and pseudocode are presented in the supplementary appendix (available online only).

Using the pseudocode as a guide, each definition was programmatically executed (by SAR) as a series of logical database queries to the EDW. The queries were independently reviewed by a data warehouse quality assurance specialist for accuracy. Our base or reference population was defined as adults at least 18 years old when the observation period began on 1 January 2007 that resided in Durham County, North Carolina (determined by geocoding methods)²², and had one or more encounters within DUHS during the 5-year observation period (1 January 2007 to 31 December 2011); this observation period was selected for convenience and availability of data relevant to the study. The yield and characteristics of the populations generated using each phenotype definition were compared using

aggregate statistics, and the proportions of overlap between phenotype-generated populations were computed.

RESULTS

The features of cohorts returned by the different definitions, as well as those of the reference population (n=173, 503), are presented in table 2. The DDC phenotype definition returned the most patients (22 050), followed by the SUPREME-DM (18 958), the ICD-9-CM 250.xx codes (18 893), the CMS CCW (16 320), NYC A1c Registry (12 182), diabetes-associated medications (11 800), and eMERGE (11 620) definitions. The average age in the reference population was 41.7 years; the average age of diabetes cohorts generated from each definition was older, ranging from 55.6 (DDC) to 57.3 years (eMERGE). The percentage of women returned for each cohort and for the reference population was similar (56–59%). The mean number of encounters (visits to DUHS over 5 years for clinical, laboratory, or radiology appointments) for the source population was 20; the mean numbers of encounters for each of the phenotype cohorts were higher (range 44–54 encounters/5 years). The mean number of encounters (54) for both the NYC A1c Registry and diabetes-associated medications phenotype definitions was higher than for the other phenotype definitions, which ranged from 45 to 49 encounters. The mean number of days between first and last patient encounter was lower (861) for the reference population than for all seven of the phenotype-identified populations, which ranged from 1224 (DDC) to 1394 (diabetes-associated medications) mean days between first and last patient encounter with DUHS during our observation period. Using a two-sided t test ($\alpha=0.05$), all cohorts show statistically significant results versus the reference population for age, number of encounters, and length of time between first and last patient encounter. We would expect this pattern given that patients with diabetes would tend to be older than a healthier population and have greater exposure to health services within Duke Medicine. SD and significance were calculated using JMP Pro, V10, and are presented in table 2.

Table 2 Demographic characteristics of cohorts returned from selected diabetes phenotype definitions

Characteristic	DUHS reference population	Phenotype definition						
		ICD-9-CM 250.xx Codes	CMS CCW (full ICD-9 set)	NYC A1c Registry	Diabetes-associated medications	DDC phenotype	SUPREME-DM	eMERGE
Purpose for phenotype	–	Billing	Health services research	Care management	–	Community-wide intervention	Epidemiology; community-wide intervention	Genetic research
Type of diabetes targeted	–	All	All	All	*T2DM preferred	*T2DM preferred	All	†T2DM exclusive
Age—year† (mean±SD)‡	41.7±17.5	56.1±15.8	56.8±15.5	56.2±15.1	54.1±15.3	55.6±16.3	56.6±15.9	57.3±15.4
Female sex: # and (%)	99 695 (57%)	10 644 (56%)	9185 (56%)	6812 (56%)	6933 (59%)	12 603 (57%)	10 681 (56%)	6524 (56%)
No of encounters§ (mean±SD)†	20±33.5	46±57.1	49±58.6	54±59.3	54±60.4	46±56.3	48±57.9	45±52.5
Length of time (in days) between first and last patient encounter (mean±SD)†	861±675.9	1252±587.6	1295±558.4	1365±524.5	1394±500.5	1224±595.9	1257±576.1	1258±579.4
Total patients identified	173 503	18 893	16 320	12 182	11 800	22 050	18 958	11 620
% Reference population identified	n/a	11%	9%	7%	7%	13%	11%	7%

*Project focus or intent is for T2DM populations, but phenotype does not aggressively eliminate T1DM patients.

†Patients with indications of T1DM are specifically excluded.

‡Age at the beginning of the observation period, 1 January 2007.

§Within observation period, 1 January 2007–31 December 2011.

CMS CCW, Centers for Medicare and Medicaid Services Chronic Condition Data Warehouse; DDC, Durham Diabetes Coalition; DUHS, Duke University Health System; eMERGE, electronic medical records and genomics; HbA1c, hemoglobin A1c; ICD-9-CM, International Classification of Disease, revision 9, clinical modification; NYC, New York City; SUPREME-DM, Surveillance, Prevention, and Management of Diabetes Mellitus; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Table 3 illustrates the number and proportion of patients from the DUHS sample that meet various single criteria that are the components (diagnosis, laboratory, and medications) of the seven definitions applied. A total of 24 520 (14%) patients in the DUHS sample population met criteria in at least one of the categories; 18 893 patients had one or more instances of ICD-9-CM 250.xx codes, of which 16 456 met those criteria and no others in the diagnosis category (ie, those patients did not have any ICD-9-CM codes other than 250.xx). Of those, 3441 patients were identified only by this single criterion, and no other criteria in any category (ie, no laboratory or medication criteria were met). A total of 2524 patients had instances of ICD-9-CM codes from the expanded list, and of those, 25 patients were identified as having diabetes using only these criteria. Because patients often had multiple instances of these codes, the numbers for ‘any’ ICD-9-CM code are not the sum of the patients with 250.xx codes and codes from the expanded set of ICD-9-CM codes.

A total of 18 833 patients showed at least one instance of any abnormal laboratory result (HbA1c, fasting or random glucose), with more patients having an abnormal random glucose test result (15 320) than an abnormal HbA1c (12 182) or fasting glucose (507) test result. Fewer patients were identified solely by an abnormal laboratory test (3332 for random glucose, 578 for HbA1c, and 52 for fasting glucose) with no other diagnosis or medication criteria. A total of 11 800 patients had instances of one or more diabetes medications, but just over 1000 were identified exclusively by diabetes-related medications, including 638 on metformin.

Figure 1 depicts the overlap of patients meeting inclusion criteria presented in table 3 by category: diagnosis (18 980), laboratory (18 833), and medications (11 800). Of the 24 520 patients who met criteria in any of the three categories, 9441 (39%) were identified by all three criteria, and 15 652 patients (64%) were identified by at least two criteria.

A total of 8868 (36%) patients met criteria in one data category exclusively. As shown in figure 1, laboratory test results (n=4203) identify a greater number of patients exclusively than do ICD-9-CM diagnosis codes (n=3536) and diabetes-related medications (1129). A total of 4981 patients had diabetes-specific ICD-9-CM codes and abnormal laboratory test values (with no documented diabetes medications), and 1022 had ICD-9-CM codes and diabetes-related medications with no laboratory test criteria. Only 208 patients had laboratory and medication codes suggestive of diabetes with no relevant ICD-9-CM codes.

The overlap between pairs of phenotype definitions is further illustrated in table 4, as a percentage of patients retrieved by one definition in a pair to the number of unique patients retrieved by both definitions. The following pairs of phenotype definitions showed the highest proportion of overlapping populations: CMS CCW versus ICD-9-CM250.xx codes (86%), CMS CCW versus SUPREME-DM (86%), DDC versus ICD-9-CM250.xx codes (85%), and DDC versus SUPREME-DM (85%). The eMERGE definition showed the lowest concordance with all other definitions (range 44–61%) as did the NYC A1c Registry definition (range 50–60%).

DISCUSSION

In our sample, ICD-9-CM codes and abnormal laboratory test results identified more patients than did medications. For the subset of patients that met criteria in any category (diagnosis, laboratory results, medications), laboratory test criteria was the largest single-domain criteria and identified more patients than ICD-9-CM diagnoses or medications alone, probably because

Table 3 Specifics of DUHS sample population by data category and specific criteria

Categories and specific criteria	Patient meets criterion	Patient only meets this criterion within this category (might also meet criteria in other categories)	Exclusivity measure: patient only meets this criterion in this category, and does not meet criteria in other categories
Diagnosis code category criteria			
ICD-9-CM 250.xx codes	18 893	16 456	3441
ICD-9-CM expanded codes*	2524	87	25
Any ICD-9-CM code (250.xx and/or expanded; (ie, codeset for CMS CCW phenotype definition)†	18 980	–	70
Laboratory category criteria			
Abnormal HbA1c (>=6.5%)	12 182	3246	578
Abnormal random glucose (>=200 mg/dl)	15 320	6532	3332
Abnormal fasting glucose (>=126 mg/dl)	507	98	52
Any abnormal laboratory test result‡,§	18 833	–	241
Medication category criteria			
Medication classification: alphaglucoisidase inhibitors	37	5	2
Medication classification: insulin secretagogues	4651	826	131
Medication classification: incretin hormone based	1130	37	7
Medication classification: insulin	4890	2019	123
Medication classification: metformin	7900	3209	638
Medication classification: thiazolidinediones	1596	99	13
Any diabetes-related medication‡	11 800	–	215
Patients meeting criteria in any of the three categories	24 520		

*Expanded codes=249.xx, 357.2, 362.01–7, 366.41; the code 362.07 is not included in CMS CCW definition.

†OGTT results are omitted here for simplicity; the numbers were very low and OGTT tends to be used in combination with pregnancy exclusion criteria.

‡Because patients can have more than one instance of these codes, the numbers for ‘any’ code are not the sum of the code subsets identified above.

§CMS CCW, Centers for Medicare and Medicaid Services Chronic Condition Data Warehouse; DUHS, Duke University Health System; HbA1c, hemoglobin A1c; ICD-9-CM, International Classification of Disease, revision 9, clinical modification; OGTT, oral glucose tolerance test.

multiple abnormal tests are required for a diagnosis of diabetes and subsequent medical management. The presence of so few fasting glucose tests was surprising. Although random glucose is not used for diabetes screening, it is more common because patients are often not fasting, and test orders often do not specify the test as a fasting even if patients have fasted. More importantly, the populations returned by applying these definitions in our sample probably represent prevalent populations that are managed with A1c tests. If we restricted our base sample to an incident population, we would expect to find many more fasting glucose tests used for initial diagnosis.

However, as the use of A1c for diabetes screening becomes more popular, fasting glucose tests will decrease.

The use of ICD-9-CM diagnostic codes (alone or in combination with other criteria) also identified significant numbers of patients. Diabetes-related medication criteria also identified a significant minority of patients, although most of these met laboratory and diagnostic criteria as well. These data suggest that phenotype definitions with multiple components, such as SUPREME-DM and DDC, are required for a comprehensive definition for diabetes. These phenotype definitions did indeed identify the most patients and more than ICD-9-CM codes alone (SUPREME-DM only slightly more and the DDC significantly more) from our reference population.

Given that diabetes is a chronic condition with many comorbidities, we expect these patients to have repeated interaction with healthcare providers. The average number of encounters (a crude estimate of interaction with our healthcare system) is much larger for each phenotype cohort than for the DUHS reference population, as is the mean number of days between first and last DUHS visit (a crude estimate of length of observation) within the 5-year observation period. The cohorts from the NYC A1c Registry and the diabetes-associated medications definitions (which use only HbA1c or medication criteria, respectively, without coded diagnoses of diabetes) had the greatest mean number and time span of encounters, which might imply that those patients are sicker or uncontrolled. A possible explanation is that patients with a HbA1c test are likely to have had primary or endocrine care with DUHS (and hence more visits), and that encounters for patients without HbA1c tests represent emergent or non-diabetes-related care. The DDC definition cohort showed lower values for these health utilization metrics, perhaps suggesting fewer complications and fewer visits to DUHS, both of which are consistent with the possibility that broader and less stringent

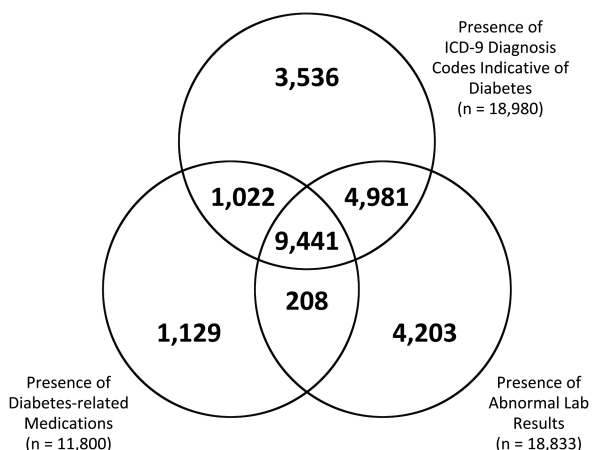


Figure 1 Overlap of diabetes cohorts identified from different categories of phenotype eligibility criteria; n=24 520 patients identified by criteria from any of the three categories.

Table 4 Overlap of different phenotype definitions, presented as proportion of patients returned by one definition to unique patients returned by both definitions

Phenotype definition	Phenotype definition						
	ICD-9-CM250.xx phenotype cohort (n=18 893)	CMS CCW phenotype cohort (n=16 320)	NYC A1c Registry phenotype cohort (n=12 182)	Diabetes-associated medication phenotype cohort (n=11 800)	DDC phenotype cohort (n=22 050)	SUPREME-DM phenotype cohort (n=18 958)	eMERGE T2DM phenotype cohort (n=11 620)
ICD-9-CM250.xx							
CMS CCW phenotype	16 251/18 962 (86%)	16 251/18 962 (86%)	11 278/19 797 (57%)	10 451/20 242 (52%)	18 831/22 112 (85%)	16 731/21 120 (79%)	11 468/19 045 (60%)
NYC A1c Registry phenotype	11 278/19 797 (57%)	10 936/17 566 (62%)	10 936/17 566 (62%)	10 002/18 118 (55%)	16 301/22 069 (74%)	16 308/18 970 (86%)	10 547/17 393 (61%)
Diabetes-associated medication phenotype	10 451/20 242 (52%)	10 002/18 118 (55%)	8504/15 478 (55%)	8504/15 478 (55%)	11 650/22 582 (52%)	11 513/19 627 (59%)	7940/15 862 (50%)
DDC phenotype	18 831/22 112 (85%)	16 301/22 069 (74%)	11 650/22 582 (52%)	11 800/22 050 (54%)	11 800/22 050 (54%)	18 866/22 142 (85%)	11 620/22 050 (53%)
SUPREME-DM phenotype	16 731/21 120 (79%)	16 308/18 970 (86%)	11 513/19 627 (59%)	10 891/19 867 (55%)	18 866/22 142 (85%)	11 030/19 548 (56%)	11 030/19 548 (56%)
eMERGE phenotype	11 468/19 045 (60%)	10 547/17 393 (61%)	7940/15 862 (50%)	7159/16 261 (44%)	11 620/22 050 (53%)	11 030/19 548 (56%)	11 030/19 548 (56%)

The denominator for each cell is number of unique patients retrieved by either phenotype. Shaded cells indicate overlap of >=75%. CMS CCW, Centers for Medicare and Medicaid Services Chronic Condition Data Warehouse; DDC, Durham Diabetes Coalition; eMERGE, electronic medical records and genomics; ICD-9-CM, International Classification of Disease, revision 9, clinical modification; NYC, New York City; SUPREME-DM, Surveillance, Prevention, and Management of Diabetes Mellitus; T2DM, type 2 diabetes mellitus.

inclusion criteria might capture patients with well-controlled diabetes or those without medications or laboratory tests present, due to shorter duration of disease or the use of providers outside of DUHS.

Previous research has demonstrated reduced efficiency of diabetes detection algorithms due to chart fragmentation in patients receiving care from multiple providers.¹⁸ Presumably, more longitudinal medical history data in a patient's EHR will improve the accuracy of an algorithm, especially for phenotypes (eg, DDC, SUPREME-DM, and eMERGE) that require multiple observations or time points. Our data suggest that the amount of follow-up (ie, EHR data available) did not differ across phenotype cohorts, reducing concerns that an obvious observation bias is present. A recently published study demonstrated a remarkable reduction in the accuracy of a T2DM algorithm when data were limited to shorter observation periods. Authors concluded that the absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying T2DM.²³

The eMERGE definition returned a smaller number of patients than other phenotype definitions, and the proportion (7%) is lower than the 8.9% expected prevalence of T2DM for Durham County, North Carolina (calculated by authors from CDC estimates),²⁴ suggesting that the eMERGE definition is potentially missing cases of T2DM. This was fully expected given the design of the eMERGE phenotype, which is appropriate for research applications that require high specificity (for T2DM) with the understanding that this will be at the expense of lower sensitivity. This restrictive algorithm was the only definition designed specifically to identify and exclude T1DM, which was critical for the genetic studies that it was designed to support. As such, the eMERGE algorithms use all three data categories (diagnoses, laboratory results, and medications) and excluded more patients than the other phenotype definitions that we examined. Patients excluded by the eMERGE definition might include T2DM patients who were treated with insulin initially or currently, or who may have had T1DM diagnosis codes incorrectly applied with one or more encounters. Populations derived from ICD-9-CM, DDC and SUPREME-DM phenotype definitions all returned greater prevalence estimates than the 8.9% estimate for our county, which was not surprising given the broad inclusion criteria and limited exclusion criteria for each of these phenotypes, plus the nature of our sample, which includes people with visits to our health system. Furthermore, the county prevalence estimates are based on patient-reported data, and as nearly a third of patients with diabetes do not know they have diabetes, the 8.9% estimate probably underestimates the true prevalence of diabetes in our county. It is possible that one of the phenotype definitions explored in this study will support future county, state, and national surveillance statistics.

Our results are consistent with the design of each definition. Those with more inclusion criteria and fewer exclusion criteria (eg, SUPREME-DM and the DDC) returned larger populations than did single criteria definitions (ICD-9-CM, CMS CCW, HbA1c, or medications) or the eMERGE definition, which had the most explicit exclusion criteria designed to identify and remove T1DM. Presumably, the larger populations returned by the broader definitions might include cases of T1DM, pre-diabetes, or other false positives (eg, polycystic ovary syndrome, non-alcoholic fatty liver disease, impaired glucose tolerance for patients with severe acute illness). However, these definitions are designed to identify the largest possible populations for screening and intervention, and require greater sensitivity (and less specificity) than do research definitions such as eMERGE. The relative sizes of the populations returned from different phenotype

definitions illustrate that the selection of an 'ideal' phenotype definition will depend on its intended use, associated targets for specificity and sensitivity, and tolerance for false positives and negatives.

Our results also show that the populations identified by the seven phenotype definitions are different. For example, although the eMERGE and NYC A1c Registry definitions return approximately the same number of patients (11 620 and 12 182, respectively), the overlap is only 50%. As eMERGE only uses HbA1c criteria in two of five case selections (and in those two cases it must be present in combination with diagnosis code criteria or medication criteria), the low concordance between the eMERGE and NYC Registry cohorts is not surprising given the patterns of criteria combinations illustrated in figure 1. The phenotypes that we examined were designed to target different populations for different purposes, including patients with T2DM specifically (for research) and diabetes of multiple types (for community surveillance and intervention). The variation among diabetes cohorts raises the question of whether heterogeneous definitions are identifying populations with different clinical disease profiles (ie, different phenotypes) or if they are failing to identify the 'same' clinical phenotype consistently. Based on our research, we see justification for at least two different clinical phenotypes: one that is strictly T2DM and another that is more inclusive for all types of diabetes. In either case, further exploration of the clinical and demographic features of different diabetes cohorts and in different settings is warranted. Part of this exploration will include comparisons of basic demographics (eg, gender, race, age) on the population returned by a definition with the population originally published by that definition. For example, HMORN populations returned using the SUPREME-DM definition were approximately half women (48–52% range across HMORN sites), but the phenotype definition applied to our sample yielded 56% women. It would not be expected that the population characteristics of different populations will match perfectly, but large differences could indicate that the selected phenotype definition was not properly implemented, or was not appropriate for the population or application.

The lack of clinical validation of any of the seven phenotype definitions in our DUHS population is a major limitation of our study. In addition, the medication data we report were collected only from the outpatient reconciliation process, and did not include medications ordered, filled, or administered in emergency department or inpatient contexts. Therefore, the results presented here should be considered descriptive and preliminary. The formal validation of the performance and results of the DDC phenotype definition in our institution are underway.

The logic embedded within different phenotype definitions (eg, inpatient/outpatient setting, the number of events, time period or time between events) makes it difficult to compare definitions and identify which data components have the most impact. For example, the CMS CCW cohort (using 68 different ICD-9-CM codes) identified fewer patients than the 250.xx cohort (using only 40 codes), because the CMS CCW distinguishes between inpatient and outpatient codes and requires more than one instance of outpatient codes on separate dates. Furthermore, because our base population was defined to include patients with one or more DUHS encounters, and some definitions require two outpatient diagnoses, admittedly some number of patients in the base population with only one encounter cannot possibly meet the two diagnoses criterion. Given the mean numbers of encounters (range 20–54) in table 2, this proportion should be quite small. The complexity of

phenotype definitions is an important consideration for potential implementers. The translation of concepts from phenotype definitions into operational queries of phenotype definitions to local applications can lead to variation and the risk of errors, and complex phenotypes might amplify these risks.

Customizations of any phenotype definition will be inevitable given the heterogeneous nature of EHR systems in the USA, and will have to address variation in ICD-9-CM coding practices in different organizations. For example, investigators from eMERGE observed a consistent billing pattern of using ICD-9-CM code 250.00 for both T1DM and T2DM patients at one site, necessitating the 'adaptation' of standard phenotyping algorithms within a single study.⁸ Because ICD-9-CM codes are an important component of most phenotype definitions, the implementation of ICD-10-CM in 2014 will create the need to revise and re-evaluate the majority of current phenotype definitions as well as programming and artifacts that support local customizations and application.

This research underscores the outstanding and important need for clearly defined phenotype definitions and consistent application in multisite projects. Currently, there are no standard EHR phenotype definitions for most chronic conditions, including diabetes. A recent study examined the definitions in five Department of Health and Human Services data systems (National Health Interview Survey, National Ambulatory Medical Care Survey, Medical Expenditure Panel Survey, Nationwide Inpatient Sample of the Healthcare Cost and Utilization Project, and CMS claims administrative data) for 20 chronic conditions, and showed differences in data types and ICD-9-CM codes used that prevented comparison of study results and disease prevalence.²⁵ Our early experience comparing existing phenotype definitions should inform the selection and implementation of phenotypes in diabetes and other diseases, and motivate future phenotype users to explore the implications of different phenotypes within their organization. This exploration should include detailed examination of important and disease-relevant features of diabetes cohorts, as well as known and unknown features of health information systems that might introduce bias and affect phenotype performance.²⁶ Further research on the clinical characteristics of standard diabetes cohorts will be important to identify appropriate phenotype definitions for health, policy, and research.

CONCLUSION

Identifying patients with diabetes is critical to monitoring the prevalence, morbidity, mortality, and impact of the disease, and standard phenotype definitions can support these activities. Cohorts of patients who are accurately and reliably classified with diabetes can be used to support research, quality, and care endeavors. The use of different phenotype definitions for diabetes can return different patient cohorts and impact the implementation of multisite projects and the evaluation of results from different studies. The intended purpose and available data drive the phenotype to be chosen. Methods for comparing existing phenotype definitions should inform the selection and implementation of phenotypes for diabetes and other diseases.

Acknowledgements The authors wish to thank Deepthi Krishnamaneni (Duke Health Technology Solutions) for her review of the phenotype definition queries; Lesley Curtis, Melissa Greiner, and Laura Qualls (Duke Clinical Research Institute) for their support in the identification of relevant phenotype definitions; Tanya Saloom (Duke Health Technology Solutions) for her work in calculating SD and significance; and Claire Osgood (Nicholas School of the Environment) for her perspective and insight into this dataset and patient population. They are also grateful to Dr Ashley Dunham (Duke Translational Medicine Institute) for clarification of DDC activities and

procedures, and to Andrew Lanza (CDC) and Caitlin Barnett (CMS) for their review and helpful suggestions on this manuscript.

Contributors The concept and design of this manuscript was developed by SAR, RLR and SES. The data analysis was conducted by SAR. All authors made substantial contributions to the interpretation of the results and to the discussion section of this paper. In addition, all authors were involved in drafting the article or revising it critically for important intellectual content. All authors gave final approval of the version to be published.

Funding This publication was supported in part by grant number 1C1CMS331018-01-00 from the Department of Health and Human Services, Centers for Medicare & Medicaid Services, and in part by the Bristol Myers Squibb Foundation Together on Diabetes program.

Competing interests None.

Ethics approval The study received approval from Duke University institutional review board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data reported in this study are aggregate data from our health system. The raw data cannot be shared, but the manuscript will be placed on PubMed. The details of phenotype definition and code will be made available via JAMIA online appendices.

REFERENCES

- 1 Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. *JAMA* 2002;288:1909–14.
- 2 Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. *JAMA* 2002;288:1775–9.
- 3 American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2013;36(Suppl. 1):S67–74.
- 4 eMERGE. What is the Phenotype KnowledgeBase?: Vanderbilt.; 2012. <http://www.phekb.org/> [cited 2013 Mar 25].
- 5 Hebert PL, Geiss LS, Tierney EF, et al. Identifying persons with diabetes using Medicare claims data. *Am J Med Qual* 1999;14:270–7.
- 6 Weiner JP, Parente ST, Garnick DW, et al. Variation in office-based quality. A claims-based profile of care provided to Medicare patients with diabetes. *JAMA* 1995;273:1503–8.
- 7 Fowles JB, Lawthers AG, Weiner JP, et al. Agreement between physicians' office records and Medicare Part B Claims data. *Health Care Financ Rev* 1995;16:189.
- 8 Kho AN, Hayes MG, Rasmussen-Torvik L, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc* 2012;19:212–18.
- 9 Lin MC, Vreeman DJ, McDonald CJ, et al. A characterization of local LOINC mapping for laboratory tests in three large institutions. *Methods Inf Med* 2011;50:105–14.
- 10 Denny J, Basford M. Type 2 Diabetes—Demonstration Project. (Created Mar. 20, 2012). PheKM: eMERGE; 2012 [cited 2013 April 15]. <http://www.phekb.org/phenotype/type-2-diabetes-demonstration-project>.
- 11 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.
- 12 McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* 2011;4:13.
- 13 Moulton G. HMO research network to focus on cancer prevention and control. *J Natl Cancer Inst* 1999;91:1363.
- 14 Vogt TM, Elston-Lafata J, Tolsma D, et al. The role of research in integrated healthcare systems: the HMO Research Network. *Am J Manag Care* 2004;10:643–8.
- 15 Leibson CL, O'Brien PC, Atkinson E, et al. Relative contributions of incidence and survival to increasing prevalence of adult-onset diabetes mellitus: a population-based study. *Am J Epidemiol* 1997;146:12–22.
- 16 Pacheco JA, Thompson W. Type 2 Diabetes Mellitus. Type 2 Diabetes Mellitus Electronic Medical Record Case and Control Selection Algorithms (from Northwestern University) PheKB website. Vanderbilt University, 2011 [cited 2013 March 25]. <http://phenotype.mc.vanderbilt.edu/sites/phenotype/files/T2DM-algorithm.pdf>.
- 17 Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med* 2011;3:79re1.
- 18 Wei WQ, Leibson CL, Ransom JE, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *J Am Med Inform Assoc* 2012;19:219–24.
- 19 Centers for Medicare and Medicaid Services. Chronic condition data warehouse: Centers for Medicare and Medicaid Services 2013 [cited 2013 May 27]. <http://www.ccwdata.org/index.htm>.
- 20 Gorina Y, Kramarow EA. Identifying chronic conditions in Medicare claims data: evaluating the Chronic Condition Data Warehouse algorithm. *Health Serv Res* 2011;46:1610–27.
- 21 Chamany S, Silver LD, Bassett MT, et al. Tracking diabetes: New York City's A1C Registry. *Milbank Q* 2009;87:547–70.
- 22 Miranda ML, Edwards SE. Use of spatial analysis to support environmental health research and practice. *N C Med J* 2011;72:132–5.
- 23 Wei WQ, Leibson CL, Ransom JE, et al. The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. *Int J Med Inform* 2013;82:239–47.
- 24 CDC. National Diabetes Surveillance System. <http://apps.nccd.cdc.gov/DDTSTRS/default.aspx>. Retrieved 4/16/2013.: Centers for Disease Control and Prevention; 2013.
- 25 Goodman RA, Posner SF, Huang ES, et al. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. *Prev Chronic Dis* 2013;10:E66.
- 26 Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* 2013;20:117–21.