ORIGINAL RESEARCH

An algorithm for identification and classification of individuals with type 1 and type 2 diabetes mellitus in a large primary care database

Manuj Sharma[1]
Irene Petersen[1,2]
Irwin Nazareth[1]
Sonia J Coton[1]

[1]Department of Primary Care and Population Health, University College London, London, UK; [2]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark

Background: Research into diabetes mellitus (DM) often requires a reproducible method for identifying and distinguishing individuals with type 1 DM (T1DM) and type 2 DM (T2DM).

Objectives: To develop a method to identify individuals with T1DM and T2DM using UK primary care electronic health records.

Methods: Using data from The Health Improvement Network primary care database, we developed a two-step algorithm. The first algorithm step identified individuals with potential T1DM or T2DM based on diagnostic records, treatment, and clinical test results. We excluded individuals with records for rarer DM subtypes only. For individuals to be considered diabetic, they needed to have at least two records indicative of DM; one of which was required to be a diagnostic record. We then classified individuals with T1DM and T2DM using the second algorithm step. A combination of diagnostic codes, medication prescribed, age at diagnosis, and whether the case was incident or prevalent were used in this process. We internally validated this classification algorithm through comparison against an independent clinical examination of The Health Improvement Network electronic health records for a random sample of 500 DM individuals.

Results: Out of 9,161,866 individuals aged 0–99 years from 2000 to 2014, we classified 37,693 individuals with T1DM and 418,433 with T2DM, while 1,792 individuals remained unclassified. A small proportion were classified with some uncertainty (1,155 [3.1%] of all individuals with T1DM and 6,139 [1.5%] with T2DM) due to unclear health records. During validation, manual assignment of DM type based on clinical assessment of the entire electronic record and algorithmic assignment led to equivalent classification in all instances.

Conclusion: The majority of individuals with T1DM and T2DM can be readily identified from UK primary care electronic health records. Our approach can be adapted for use in other health care settings.

Keywords: diabetes and endocrinology, epidemiology, public health, databases, algorithm

Introduction

Diabetes mellitus (DM) is a disease characterized by chronic hyperglycemia that occurs due to a deficiency of or resistance to the hormone insulin. It is a major cause of morbidity with estimated 347 million cases worldwide and is expected to become the seventh leading cause of death in the world by 2030.[1] Several subtypes of DM exist, with type 1 DM (T1DM) and type 2 DM (T2DM) being the most widely occurring forms and accounting for over 95% of cases.[2,3] T1DM is an autoimmune disease that peaks in incidence at puberty, though it can manifest at any age and accounts for 5%–10% of all cases of DM.[3] T2DM is an acquired form of DM that is strongly associated with being overweight and accounts for ~90% of all cases of DM.[4] The prevalence

Correspondence: Manuj Sharma
Department of Primary Care and Population Health, University College London, Rowland Hill St, London NW3 2PF, UK
Tel +44 20 7508 3702 40
Fax +44 20 7472 6871
Email manuj.sharma.11@ucl.ac.uk

and incidence of T2DM has been increasing worldwide,[3,5] particularly among older age groups and certain ethnic groups such as people of African, Caribbean, and Southeast Asian origins.[6] Despite an overlap in symptoms, both T1DM and T2DM have different prognoses and are managed differently pharmacologically.[7,8] Individuals with T1DM require insulin for survival due to the lack of insulin production, whereas those with T2DM do not stop producing insulin but develop a resistance to its effects.[3] Management of T2DM is initially through the use of various other antidiabetic agents though they do often progress to needing insulin as well.[7] Other DM subtypes such as gestational diabetes, maturity-onset diabetes of the young, latent autoimmune diabetes in adults, drug-induced diabetes, and even less well-defined idiopathic DM cases account for <5% of all DM cases.[3]

Epidemiological research conducted using electronic health records into DM can provide essential and valuable insight into prevalence, incidence, management, and prognosis of the disease but requires careful and correct identification of DM type to ensure clinical questions are accurately answered. Miscoding, misclassification, and even misdiagnosis are well-established problems with identifying DM type in health records and hence identification and classification of cases can be challenging.[9] This study aims to provide a transparent, reproducible method for classifying diabetics as T1DM and T2DM in UK electronic general practice clinical records that is readily replicable and modifiable for other epidemiological settings.

Materials and methods
Data source
The Health Improvement Network (THIN) primary care database contains anonymized longitudinal electronic health records from 587 primary care practices throughout the UK with over 12 million individuals contributing data. Information available in THIN is collected during routine consultations with general practitioners (family physicians) and health care staff from when an individual registers at a general practice to when they leave the practice or die. THIN is broadly representative of the UK population in terms of patient characteristics, disease burden, and mortality.[10] Data stored in THIN include information on demographics, diagnoses, symptoms of disease, specialist referrals, laboratory testing, disease monitoring, prescribing, secondary care discharge information, and death. Symptoms, diagnosis, and disease monitoring are recorded using Read codes and AHD (Additional Health Data) codes, hierarchical coding systems within medical records, and additional health record files.[11] Using

Read code dictionaries, lists can be created to identify individuals with different symptoms and disease.[12] Each unique medication type and strength is given a drug code which can be used for creating drug code lists of medications prescribed.

Study population
All data included in this study were from practices that met quality assurance criteria in THIN, as determined by the acceptable mortality reporting and computer usage standards.[13,14] We included all individuals aged 0–99 years who were registered with a general practice contributing data between 2000 and 2014 and had at least 1 year of quality-assured data following registration.

Algorithm generation
Our method for identifying and then classifying individuals with T1DM and T2DM involved the use of a two-step algorithm. In the first step, we identified all individuals with potential T1DM or T2DM while excluding those coded as having only rarer subtypes of the disease. With the second step, we distinguished cases as having T1DM and T2DM. This two-step algorithm was devised following several discussions within a multidisciplinary clinical research team.

Algorithm step 1 – Identification of individuals with potential T1DM or T2DM
A list of Read codes, drug codes, and AHD codes indicative of DM was prepared. All individuals with any such code indicative of DM in their health record were then identified. We then removed individuals who had no DM records except for metformin prescriptions (probable polycystic ovary syndrome and metabolic disease cases), individuals with only a single record of DM, and individuals who had no diagnostic record (Read code or AHD code) for DM.

Sensitivity analysis on individuals remaining revealed that one particular AHD code being used entitled, "HbA1c diabetic control", was misclassifying cases as DM. Though this code was designed for use in monitoring of DM individuals, exploration revealed that general practitioners were also using this code among nondiabetic and prediabetic individuals as well (potentially for screening purposes). To overcome this problem, individuals who had been assigned as having DM due only to the presence of this code were examined. If they had a HbA1c result above the World Health Organization recommended threshold value of 48 mmol/mol (6.5%), these individuals were classified as having DM; otherwise, they were excluded.[15]

Finally, we excluded individuals with diagnostic codes for other DM subtypes only, for example, gestational diabetes to

obtain the final cohort. The earliest date on which any DM code was recorded was defined as the index date for the start of DM.

Algorithm step 2 – Classification of individuals with T1DM and T2DM

Within the cohort of individuals identified with potential T1DM or T2DM, we generated five variables to help distinguish the DM type. These are listed in a descending level of importance as follows:

- Diagnostic code type assigned
- Cumulative days of noninsulin prescriptions
- Number of insulin prescriptions
- Incident or prevalent case
- Age at first record of DM

Diagnostic code type assigned

We categorized individuals as those who only had T1DM-specific diagnostic codes used in their health record, T2DM-specific codes used in their health record, T1DM- and T2DM-specific codes used in their health record due to diagnostic or coding errors, and finally those with only non-specific DM diagnostic codes. Examples of Read codes are detailed in Table 1 and in full in the Supplementary material.

Cumulative days of other antidiabetic prescriptions

The number of days an individual was prescribed other antidiabetic (noninsulin) treatment was determined by dividing the quantity of medication issued by the daily dose the individuals were prescribed. In instances where either of these variables was missing, we used a deterministic method of imputing quantity or daily dose based on examination of what was common for that medication quantity or daily dose for individuals whose value were recorded. Where information was completely missing for quantity and daily dose, we assumed prescription was for 28 days as the majority of DM treatments were issued for this duration.

Number of insulin prescriptions issued

The total number of insulin prescriptions issued per individual was also determined. Insulin is needed by individuals with T1DM for survival once the disease has fully set in. However, it is needed less commonly among T2DM individuals, usually for more advanced stages of the disease.[9]

Incident or prevalent case

Mamtani et al showed that if the first record of DM appears for an individual, ≥9 months after registering with a general practice, then that individual is likely to be an incident case of DM.[16]

However, if the first record of DM appears before 9 months in their electronic health record then this is most probably due to the recording of a DM case for someone who already had the disease before practice registration (prevalent).[16] This application was useful as it allowed us to assess whether we potentially had a complete DM record for an individual or whether there was historical DM data for an individual from before practice registration that we may not have access to.

Age of diagnosis of DM

Age of diagnosis of DM was calculated for individuals who were classified as incident cases (first record of DM appearing ≥9 months after practice registration) and for those who had a record of DM that predated their practice registration (entered retrospectively into their health record after practice registration). The first date for a record of DM when pre-registration records available were included helped inform when the disease was first diagnosed for that individual. There was a subset of individuals whose first record of DM appeared between 0 and 9 months after practice registration for whom the age of diagnosis could not be confirmed. We used, when necessary, guidance from the Royal College of General Physicians that recommends an age threshold of 35 years for distinguishing individuals with T1DM and T2DM.[9]

Validation

In order to internally validate our classification algorithm, a practically feasible sample of 500 individuals identified with DM was chosen at random from THIN. This sample included both cases classified by the algorithm as T1DM and T2DM.

Table 1 Example of diabetes mellitus Read codes

Read code	Description	Code type
C10E611	Type 1 diabetes mellitus with gangrene	T1DM
C108011	Type 1 diabetes mellitus with renal complications	T1DM
C108411	Unstable type 1 diabetes mellitus	T1DM
C10EA00	Type 1 diabetes mellitus without complication	T1DM
C109D11	Type 2 diabetes mellitus with hypoglycemic coma	T2DM
C10F700	Type 2 diabetes mellitus - poor control	T2DM
C10FJ11	Insulin-treated type 2 diabetes mellitus	T2DM
C10F000	Type 2 diabetes mellitus with renal complications	T2DM
C107y00	Other specified diabetes mellitus with peripheral circulatory complications	Nonspecific
2G5I.00	Left diabetic foot at low risk	Nonspecific
ZC2C800	Dietary advice for diabetes mellitus	Nonspecific
F372.11	Diabetic polyneuropathy	Nonspecific

Abbreviations: T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Each case was then examined and classified into DM type by a clinician independently based on assessment of each individual's full electronic THIN health record consisting of medical, prescription and additional health records. This assessment served as our reference standard. The classification assigned to these 500 individuals by the clinician was then compared with our classification by algorithmic methods to ascertain diagnostic accuracy of the algorithm.

Ethics

THIN has been used for scientific research since approval from the NHS South-East Multi-Centre Research Ethics Committee in 2003. Scientific approval to undertake this study was obtained from CMD Medical Research's Scientific Review Committee in February 2015. (SRC Reference Number: 15-011).

Results

Algorithm step 1 – Identification of individuals with potential T1DM or T2DM

We identified 9,161,866 individuals aged 0–99 years between 2000 and 2014. From this cohort, we identified 457,918 individuals with potential T1DM or T2DM. The number of individuals removed at each step during the application of the algorithm is illustrated in Figure 1.

Algorithm step 2 – Classification of individuals with T1DM and T2DM

Of the cohort of 457,918 individuals identified through use of algorithm 1, we classified 37,693 (8.2%) individuals as T1DM; 418,433 (91.4%) as T2DM; and 1,792 (0.4%) remained unclassified (Figure 2). Only 1,155 (3.1%) individuals with T1DM and 6,139 (1.5%) with T2DM were classified with some degree of uncertainty. Thus, the vast majority of individuals were classified with confidence (36,538 [96.9%] individuals with T1DM and 412,294 [98.5%] with T2DM).

The full criteria for classification of individuals into T1DM and T2DM are detailed in Table 2 and summarized below. Unspecific diagnostic codes refer to when both T1DM and T2DM codes were used in the same individual record or when no type-specific code was used to record an individual's DM diagnosis. The individuals classified with uncertainty are highlighted with an asterisk in the following paragraphs and in Table 2.

Individuals with T1DM met one of the following criteria:

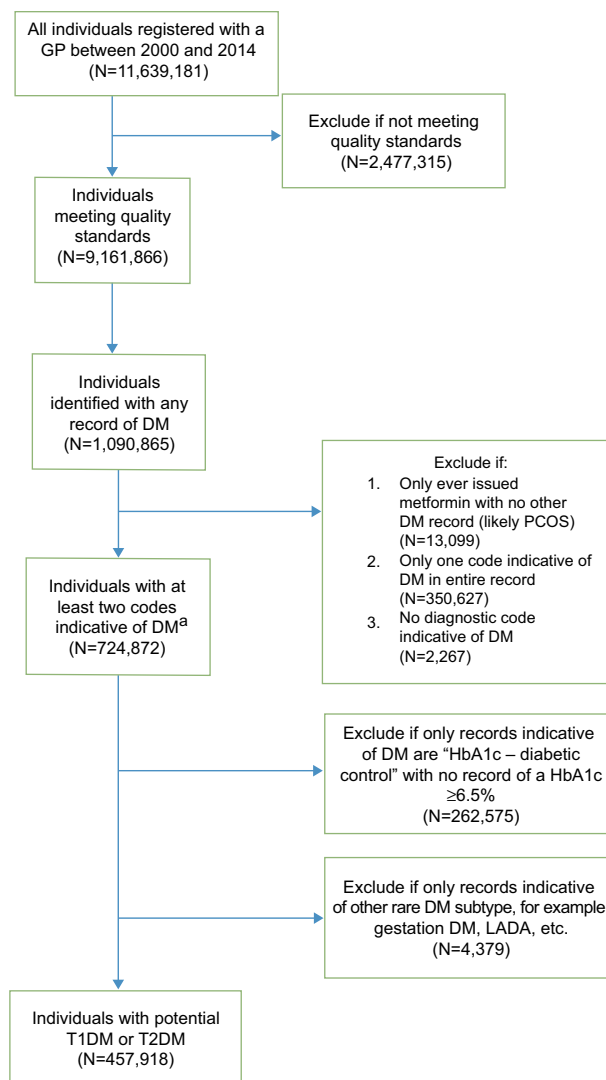1. A diagnostic code of T1DM only and prescription for insulin only.

Figure 1 Flowchart for algorithm step 1: Identification of individuals with potential T1DM or T2DM.
Note: aTwo codes must include at least one diagnostic Read code or AHD code.
Abbreviations: AHD code, Additional Health Data; DM, diabetes mellitus; GP, general practitioner; LADA, latent autoimmune diabetes in adults; PCOS, polycystic ovary syndrome; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; THIN, The Health Improvement Network.

2. A diagnostic code of T1DM only, a prescription for insulin, and <6 months cumulatively of other antidiabetic agents.

3. A T2DM code only or unspecific diagnostic codes, a prescription for insulin only, and an incident case of DM or diagnosed with DM at <35 years of age.

4. Unspecific diagnostic codes, a prescription for insulin and <6 months cumulatively of other antidiabetic agents, and an incident case of DM or diagnosed with DM at <35 years of age.*

Individuals with T2DM met one of the following criteria:

1. A diagnostic code for T2DM only and any quantity of prescription for other antidiabetic agents with or without insulin.
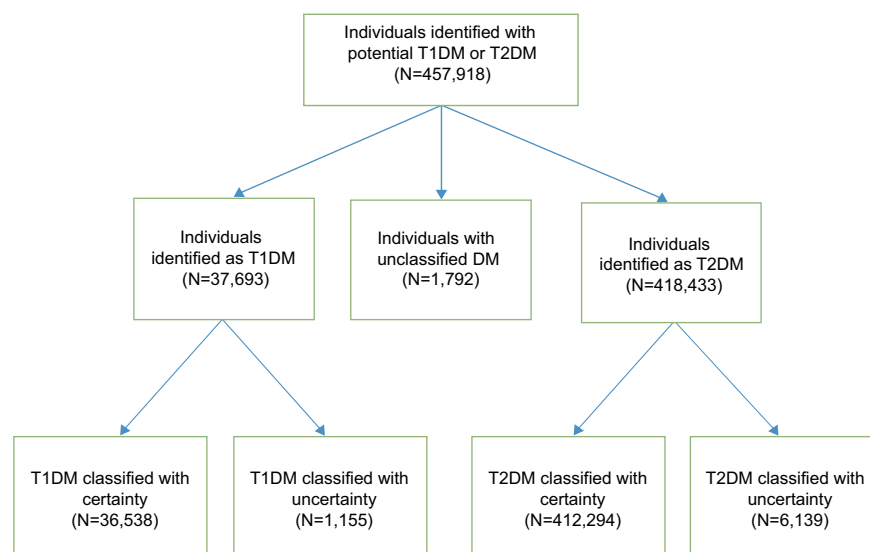
Figure 2 Flowchart for algorithm step 2: Classification of individuals with T1DM and T2DM.
Abbreviations: DM, diabetes mellitus; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

2. A diagnostic code for DM of any type and prescriptions for ≥6 months cumulatively of other antidiabetic agents with or without insulin.

3. A diagnostic code for DM of any type and any quantity of prescription for other antidiabetic agents with no insulin prescription.

4. A diagnostic code for T2DM or unspecific diagnostic codes and no prescribed treatment.

5. A diagnostic code for T1DM only and no prescribed treatment.*

6. A diagnosis of T2DM only or unspecific diagnostic codes, prescribed insulin only, but were a prevalent case and diagnosed with DM at ≥35 years of age.*

7. Unspecific diagnostic codes, prescribed insulin with <6 months cumulatively of other antidiabetic agents, a prevalent case, and diagnosed with DM at ≥35 years of age.*

Uncertainty in classification

T1DM cases classified with uncertainty were those with T2DM or unspecific codes only and up to 6 months of other antidiabetics prescribed in addition to insulin. Though individuals with T1DM do ultimately require insulin for survival, a small proportion of them have a slower onset of disease and may erroneously have other antidiabetics agents prescribed while some residual pancreatic insulin production remains and diagnosis is unclear.[9] Furthermore, it is unusual for T2DM individuals to progress to needing insulin rapidly after diagnosis. For these uncertain cases, we determined

if they were incident DM cases and thus whether we had a full history of treatment for that individual. In addition, we also examined the age of diagnosis in cases where there was uncertainty. This is because individuals diagnosed with diabetes at <35 years of age and prescribed insulin were more likely to have T1DM.[9]

T2DM cases classified with uncertainty included individuals with T1DM codes only but not prescribed treatment, individuals with unspecific diagnostic codes and prescribed insulin (and none or <6 months of other antidiabetics), and ≥35 years of age at diagnosis.[9] Though it is rare for T2DM individuals to be managed on insulin alone or progress to needing insulin rapidly after treatment initiation,[7,9] given that they were diagnosed at age ≥35 years and these were prevalent cases that had a history of DM prior to registration that we had incomplete data on, we classified these cases as T2DM but with uncertainty. These uncertain cases represented 1.5% of our total classified T2DM cohort.

Validation

In our internal validation of the classification algorithm using 500 random individuals with DM, the manual assignment of DM type based on clinical assessment of each individual's health record in THIN (reference standard) and algorithmic assignment led to equivalent classification in all instances. Though our sample size was small for feasibility purposes, we observed complete agreement for both T1DM and T2DM classification, hence sensitivity, specificity, positive and negative predictive values were all 100%.

Table 2 Algorithm step 2: classification of individuals with T1DM and T2DM

Type assigned	Code type used	Treatment	Case type	Age at diagnosis	Number
Type 1	T1DM only	Insulin only	–	–	27,942
		Insulin + OAD <6 m	–	–	1,922
	T2DM only	Insulin only	Incident	<35	150
				≥35	1,427
			Prevalent	<35	487
	Unspecific[‡]	Insulin only	Incident	<35	890
				≥35	1,364
			Prevalent	<35	2,356
		Insulin + OAD <6 m	Incident	<35	238*
				≥35	675*
			Prevalent	<35	242*
Type 2	T1DM only	Insulin + OAD ≥6 m	–	–	3,745
		OAD <6 m	Incident	<35	7
				≥35	13
			Prevalent	<35	8
				≥35	17
		OAD ≥6 m	–	–	107
		No treatment	–	–	611*
	T2DM only	Insulin	Prevalent	≥35	2,975*
		Insulin + OAD <6 m	–	–	2,993
		Insulin + OAD ≥6 m	–	–	45,896
		OAD <6 m	–	–	22,968
		OAD ≥6 m	–	–	202,865
		No treatment	–	–	70,266
	Unspecific[‡]	Insulin only	Prevalent	≥35	2,043*
		Insulin + OAD <6 m	Prevalent	≥35	510*
		Insulin + OAD ≥6 m	–	–	11,197
		OAD <6 m	–	–	5,775
		OAD ≥6 m	–	–	11,319
		No treatment	–	–	35,118
Unclassified	T1DM only	OAD <6 m	Prevalent	§	17
	T2DM only	Insulin only	Prevalent	§	448
	Unspecific[‡]	Insulin only	Prevalent	§	1,059
		Insulin + OAD <6 m	Prevalent	§	268

Notes: [‡]T1DM and T2DM codes or nonspecific codes; *individuals classified with a degree of uncertainty; [§]age of diagnosis could not be confirmed.
Abbreviations: OAD, other antidiabetics; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Discussion

In this study, we described a two-step algorithm to identify and classify individuals with T1DM and T2DM in a large UK primary care database and demonstrated that the vast majority of individuals can be classified with confidence: 36,538 (96.9%) individuals with T1DM and 412,294 (98.5%) with T2DM.

Other algorithms have been previously developed in clinical studies to identify individuals with T2DM specifically,[17] and advise on how to diagnostically distinguish T1DM from T2DM.[9] There was, however, an absence of a clear approach for distinguishing between T1DM and T2DM in a general practice database such as THIN.

The main strengths of this two-step algorithm are that it identifies and classifies the majority of individuals with T1DM and T2DM with confidence and clearly outlines

individuals for whom classification is challenging and where it is not possible. This means that depending on the clinical question of interest, the DM cohort chosen for the study can be modified; for example, by excluding individuals classified with uncertainty, one can ensure greater confidence in classification in the cohort. Additionally, code lists were generated by two researchers independently and reviewed by a clinician, and our internal validation showed high diagnostic accuracy for the algorithm. The values of this algorithm has also been demonstrated in published studies where incidence, prevalence, and prescribing patterns for T2DM were shown to compare favorably with data collected by other UK and international bodies.[5,18]

Though this algorithm is mostly suited for use in the UK general practice databases such as THIN and Clinical Practice Research Datalink, they can be adapted for use in

epidemiological research for other settings. International Classification of Diseases 10 codes or other hierarchical coding systems indicative of DM could be used instead of Read codes, whereas pharmacological therapy and thresholds for the age at diagnosis could be modified as necessary according to local treatment and monitoring guidelines.

The quality and outcomes framework introduced as part of the GP contract for the UK in 2004 brought in several indicators for DM to help improve disease management.[19] However, as financial incentives were introduced for the use of certain T1DM- and T2DM-specific codes, overzealous recording may have led to erroneous diagnoses.[9] Our algorithm considers medications prescribed, HbA1c results, age of diagnosis, and whether a case is incident or prevalent, which will reduce such errors.

There are, however, some limitations to acknowledge. In this study, we did not seek validation by comparison of our classification systems based on the algorithm to complete patient case notes. This would further strengthen the case for use of this algorithm. The sample of 500 records for internal validation was chosen for feasibility purposes however given the significant size of the cohort, a larger sample size may have been preferable to ensure more accurate validation. Markers such as body mass index and ethnicity can potentially be used to additionally support DM type classification. Body mass index is generally higher among individuals with T2DM rather than T1DM,[20] whereas T2DM is known to be more prevalent among certain ethnic groups.[21] However, given the variables we included already facilitated confident classification for 98.0% of our cohort, we did not investigate further.

We excluded cases with only diagnostic codes related to rarer subtypes of DM such as maturity-onset diabetes of the young, latent autoimmune diabetes in adults, drug-induced diabetes, and gestational diabetes. This, of course, cannot guarantee that some miscoded and misdiagnosed cases did not enter our cohort. In other epidemiological settings, where complete data for secondary care are also available, women with gestational diabetes having their first and final record of DM while pregnant could also be excluded.

Electronic health records in THIN are dynamic, that is, individuals register and leave the general practices at different points in time and some individuals have been registered for much longer than others. Individuals with only a short duration of registration may not have a DM diagnosis entered in their records or a sufficient time to be issued treatment for DM. Therefore, varying record lengths can risk introducing bias. When this algorithm is applied to other datasets, it is worth noting that the longer the record lengths following the first record of DM, the lower the risk of any such bias will be. Finally, with recent recommendations by bodies such as the National Institute for Health and Care Excellence in 2015 to consider prescribing metformin for T1DM individuals with higher body mass index, this treatment combination is likely to become increasingly common. Thus, the algorithm will need to be adapted for use in future years. This could be achieved by further scrutinizing the records of individuals on metformin and insulins only, for indicators that may help distinguish them as T1DM or T2DM such as diagnostic codes and age of diagnosis.[8]

Conclusion

We have provided a transparent and reproducible method with which the vast majority of individuals with T1DM and T2DM can be identified with confidence in primary care databases such as THIN and the Clinical Practice Research Database. With some modifications accounting for dataset type and hierarchical coding systems employed, the two-step algorithm we provide can also be applied to other electronic health record databases both in the UK and worldwide. The algorithm is flexible and can be modified as needed to vary the level of confidence in classification needed to help identify individuals with DM of interest for different epidemiological studies.

Acknowledgments

Author contributions

MS, IP, IN, and SJC collectively planned the study and writing of the manuscript. MS performed the analysis and with IP, IN and SJC interpreted the results. MS drafted the manuscript, IP, IN, and SJC revised it critically for content. MS, IP, IN, and SJC agreed the final version to be published.

Disclosure

The authors report no conflicts of interest in this work.

References

1. World Health Organisation. *WHO: 10 Facts about Diabetes*; 2014. Available from: http://www.who.int/features/factfiles/diabetes/en/. Accessed January 25, 2016.

2. World Health Organisation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Report of a WHO/IDF consultation; 2006. Available from: http://www.who.int/diabetes/publications/diagnosis_diabetes2006/en/. Accessed May 4, 2015.

3. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 2014;37(Suppl 1):S81–S90.

4. Public Health England. *Adult Obesity and Type 2 Diabetes*; 2014. Available from: https://www.gov.uk/government/publications/adult-obesity-and-type-2-diabetes. Accessed December 10, 2014.

5. Sharma M, Nazareth I, Petersen I. Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study. *BMJ Open.* 2016;6(1):e010210.

6. Brooks AP, Chong JSW. Changes in age at diagnosis and prevalence of positive family history in patients with Type 1 diabetes over five decades. *Diabet Med.* 2014;31:181.

7. National Institute for Clinical Excellence. NICE CG28: Type 2 diabetes in adults: management; 2015. Available from: https://www.nice.org.uk/guidance/ng28/resources/type-2-diabetes-in-adults-management-1837338615493. Accessed January 21, 2016.

8. National Institute for Clinical Excellence. NICE CG17: Type 1 diabetes in adults: diagnosis and management; 2015. Available from: https://www.nice.org.uk/guidance/ng17/resources/type-1-diabetes-in-adults-diagnosis-and-management-1837276469701. Accessed January 21, 2016.

9. Royal College of General Practitioners. *Coding, Classification and Diagnosis of Diabetes*; 2011. Available from: http://www.sdrn.org.uk/sites/sdrn.org.uk/files/nhs%20diagnosis%20classification%20report.pdf. Accessed January 6, 2016.

10. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care.* 2011;19(4):251–255.

11. Chisholm J. Read clinical classification. *BMJ.* 1990;300(6737):1467.

12. Dave S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf.* 2009;18(8):704–707.

13. Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf.* 2013;22(1):64–69.

14. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf.* 2009;18(1):76–83.

15. World Health Organisation. *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia*; 2006. Available from: http://www.who.int/diabetes/publications/diagnosis_diabetes2006/en/. Accessed January 6, 2016.

16. Mamtani R, Haynes K, Finkelman BS, Scott FI, Lewis JD. Distinguishing incident and prevalent diabetes in an electronic medical records database. *Pharmacoepidemiol Drug Saf.* 2014;23(2):111–118.

17. Holden SH, Barnett AH, Peters JR, et al. The incidence of type 2 diabetes in the United Kingdom from 1991 to 2010. *Diabetes Obes Metab.* 2013;15(9):844–852.

18. Coton SJ, Nazareth I, Petersen I. A cohort study of trends in the prevalence of pregestational diabetes in pregnancy recorded in UK general practice between 1995 and 2012. *BMJ Open.* 2016;6(1):e009494.

19. Health and Social Care Information Centre. *Quality and Outcomes Framework*; 2004. Available from: http://www.hscic.gov.uk/qof. Accessed March 3, 2016.

20. Eckel RH, Kahn SE, Ferrannini E, et al. Obesity and type 2 diabetes: what can be unified and what needs to be individualized? *Diabetes Care.* 2011;34(6):1424–1430.

21. Riste L, Khan F, Cruickshank K. High prevalence of type 2 diabetes in all ethnic groups, including Europeans, in a British inner city: relative poverty, history, inactivity, or 21st century Europe? *Diabetes Care.* 2001;24(8):1377–1383.