Journal of Clinical Epidemiology 153 (2023) 34−44

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Identifying type 1 and 2 diabetes in research datasets where classification biomarkers are unavailable: assessing the accuracy of published approaches

Nicholas J. Thomas[a,b], Andrew McGovern[a,b], Katherine G. Young[a], Seth A. Sharp[a], Michael N. Weedon[a], Andrew T. Hattersley[a,b], John Dennis[a], Angus G. Jones[a,b,*]

[a]*Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK*
[b]*Department of Diabetes and Endocrinology, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK*

Accepted 31 October 2022; Published online 9 November 2022

Abstract

Objectives: We aimed to compare the performance of approaches for classifying insulin-treated diabetes within research datasets without measured classification biomarkers, evaluated against two independent biological definitions of diabetes type.

Study Design and Setting: We compared accuracy of ten reported approaches for classifying insulin-treated diabetes into type 1 (T1D) and type 2 (T2D) diabetes in two cohorts: UK Biobank (UKBB) $n = 26{,}399$ and Diabetes Alliance for Research in England (DARE) $n = 1{,}296$. The overall performance for classifying T1D and T2D was assessed using: a T1D genetic risk score and genetic stratification method (UKBB); C-peptide measured at >3 years diabetes duration (DARE).

Results: Approaches' accuracy ranged from 71% to 88% (UKBB) and 68% to 88% (DARE). When classifying all participants, combining early insulin requirement with a T1D probability model (incorporating diagnosis age and body image issue [BMI]), and interview-reported diabetes type (UKBB available in only 15%) consistently achieved high accuracy (UKBB 87% and 87% and DARE 85% and 88%, respectively). For identifying T1D with minimal misclassification, models with high thresholds or young diagnosis age (<20 years) had highest performance. Findings were incorporated into an online tool identifying optimum approaches based on variable availability.

Conclusion: Models combining continuous features with early insulin requirement are the most accurate methods for classifying insulin-treated diabetes in research datasets without measured classification biomarkers.

Keywords: Diabetes classification; Population studies; Cohort stratification; Type 1 diabetes; Type 2 diabetes; Diabetes epidemiology

Funding: The DARE study was funded by the Welcome Trust and supported by the National Institute of Health and Care Research (NIHR) Exeter Clinical Research Facility. NJT is funded by a Welcome Trust funded GW4 PhD. AM is supported by a NIHR Academic Clinical Fellowship. M.N.W. is supported by the Welcome Trust Institutional Support Fund (WT097835MF). SAS is supported by a Diabetes UK PhD studentship (17/0005757). JMD is supported by an Independent Fellowship funded by Research England's Expanding Excellence in England (E3) fund. KGY is supported by Research England's Expanding Excellence in England (E3) fund. ATH is supported by the NIHR Exeter Clinical Research Facility and a Welcome Senior Investigator award and an NIHR Senior Investigator award. AGJ was supported by an NIHR Clinician Scientist award (CS-2015-15-018). The views given in this article do not necessarily represent those of the NIHR, the National Health Service, or the Department of Health.

Conflict of interest: AGJ contributed to the development of the two classification models assessed in this work. Other authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Ethics Approval: Ethics for the DARE study was granted by the Devon & Torbay Research Ethics Committee, ref: 2002/7/118.

Availability of data and materials: UKBB data are available through a procedure described at https://www.ukBiobank.ac.uk/using-the-resource/. DARE data are available through application to the Peninsula Research Bank https://exetercrfnihr.org/about/exeter-10000-prb/

Authors Contributions: NJT, AM, and AGJ designed the study. SAS, KGY, and MNW acquired the data and SAS and MNW generated the T1DGRS. NJT, JD, AM, and AGJ analyzed the data. NJT wrote the first draft of the report. All authors reviewed the draft, contributed to the revision of the report and gave final approval for publication. AGJ and NJT are the guarantors of this work.

* Corresponding author. Institute of Biomedical and Clinical Science, University of Exeter, College of Medicine and Health, Exeter, UK. Tel.: 01392408538.

E-mail address: angus.jones@exeter.ac.uk (A.G. Jones).

https://doi.org/10.1016/j.jclinepi.2022.10.022

What is new?

Key findings

- Across two different datasets classification models incorporating continuous clinical features combined with early insulin requirement, or (where available) interview- reported diabetes type consistently achieved high accuracy (≥85%).
- When identifying a type 1 diabetes (T1D) cohort with minimal misclassification, young age at diagnosis (<20 years) or models with high thresholds had very high predictive value but modest sensitivity.

What this adds to what is known?

- The best approaches for classifying diabetes type in research datasets without measured classification biomarkers were previously unclear. This work allows researchers to identify the optimum classification approach for their dataset and research question.

Implications

- The optimal method for identifying diabetes subtypes in observational data will depend on available data and research question. Researchers can select the optimum approach using an online tool devised using the study findings (Classifying Diabetes for Research: Method Selector (newcastlerse.github.io)).

1. Introduction

1.1. Robustly classifying diabetes type in research datasets without measured classification biomarkers is challenging

Large population-level research datasets are widely used for clinical studies of people with diabetes; however, for results to be robust, accurate diabetes classification is fundamental. Together, type 1 diabetes (T1D) and type 2 diabetes (T2D) account for ≥98% of all diabetes cases [1], but these two subtypes have marked differences in aetiology, pathophysiology, and management [2]. While absence of insulin treatment in longstanding diabetes is highly specific for T2D [2,3], classifying currently insulin- treated diabetes cases is challenging [3−6]. Clinical diagnosis is frequently unavailable in research datasets and if available will include substantial misclassification and miscoding (≈15%) [7−12]. In research datasets, biomarkers that can help improve classification, such as C-peptide or islet autoantibodies [4,13], are rarely available. The rarity of T2D in children makes young age of diabetes onset specific for T1D, but the over half of T1D cases occurring in adults will be missed [3−5,14].

1.2. The comparative performance of approaches to classify insulin-treated diabetes in epidemiological studies is unknown

The optimum approach for classifying T1D and T2D in research datasets remains unclear. Previously published approaches vary and include the following: clinician or interview-reported diabetes type, diabetes treatment, billing codes, or using specific cut offs of diabetes-related features, for example body mass index (BMI) or age at diabetes diagnosis [15−25]. Where the performance of these approaches has been assessed, it has normally been against a clinical-based assessment of T1D or T2D diagnosis [10,15−18,22−25]. These assessments will not only suffer from the inaccuracies of clinical diagnosis and coding but also a circularity bias where features favored by clinicians for determining diabetes type will appear most discriminatory. While prediction models for classification have been developed and tested against C-peptide and histology defined diabetes types, these have not been compared to other approaches [19,20,26]. To date, there has not been an evaluation of the comparative performance of existing classification approaches against a robust independent biomarker.

1.3. Aim

To help researchers choose the optimum diabetes classification approach for research datasets without measured classification biomarkers, we aimed to compare the performance of a number of published approaches for classifying insulin-treated diabetes in two population-level research datasets. Classification approaches were evaluated against two independent biological definitions of diabetes type based on T1D genetic risk scores (T1DGRS) and measured C-peptide.

2. Method

Within two population research datasets we assessed the performance of different published approaches for classifying insulin-treated diabetes into T1D and T2D against biomarker-defined diabetes subtypes. In UKBB, we used a T1DGRS within a previously published genetic stratification method [3,27] to compare the proportion of T1D and T2D cases correctly and incorrectly classified by each approach. We also assessed the performance of these approaches in a large unselected research dataset with diabetes (the DARE cohort) against diabetes type defined by C-peptide level (a measure of endogenous insulin secretion), measured after a median 14 years duration [13].

2.1. Study design and participants

2.1.1. UK Biobank

We evaluated a subset of 26,399 unrelated individuals self-reporting diabetes from the UKBB [28]. To allow direct comparison of classification approaches in the same cohort, individuals were excluded based on missing BMI measurement (n = 237) or age at diabetes diagnosis (n = 1,675). A further 1,389 participants were excluded where it was not possible to generate a T1DGRS. Overall, 23,098 participants met the study eligibility criteria, and a study flowchart is shown in Electronic Supplementary Materials (ESM) Figure 1A. A subset of 45% (10,491/23,098) of participants had linkage to their primary care record.

The main analysis was restricted to the 72% (16,619/23,098) of participants of White European descent, as the T1DGRS used to define diabetes type has not been validated in nonWhite ethnicities [29,30]. People of White European descent were those who self-identified as White European and were confirmed as ancestrally White by the use of principal components analyses of genome-wide genetic information [31]. A secondary exploratory analysis was undertaken including all 23,098 participants of all ethnicities. The clinical history was interview-reported diabetes type via an interactive questionnaire and nurse-led interview, and further details of clinical features and lipid assessment are given in ESM.

2.1.2. DARE cohort

The DARE study recruited, predominantly though primary care in the South West of England, an unselected population of adults with diabetes (regardless of age of onset or diabetes type; gestational diabetes excluded) [8]. We evaluated 1,296 participants (22% [1,296/5,991] of the DARE cohort) receiving insulin treatment. C-peptide was measured on stored nonfasting Ethylenediaminetetraacetic acid at DARE recruitment after January 2010 as previously described (see ESM) [8]. Participants were excluded when BMI measurement was missing (n = 6) or if diabetes duration at recruitment was $\leq$3 years (n = 49) due to the limitations of C-peptide assessment in short-duration diabetes [13]. A study flow chart is shown in ESM Figure 1B. Although all ethnicities were recruited to DARE, 99% were White (1,224/1,241). In DARE, all clinical history was self-reported by participants in an interview with a research nurse as reported previously [8].

2.2. Assessment of population-level approaches for classifying diabetes type in insulin-treated individual

Overall we compared ten different approaches for the classification of insulin-treated diabetes selected based on those commonly used in the literature [10,15−20,22,32]. The variables required for each approach are listed in Table 1. For all approaches using continuous variables, cut offs to classify either T1D or T2D were selected based on previously proposed values were available [10,15,16,19,20]. Different cut offs were used where the aim was to classify all insulin-treated participants or select a T1D or T2D cohort with minimal misclassification (Table 1). For identifying 'pure' type 1 and 2 diabetes using prediction models, no previous cut off has been recommended, therefore cut offs were chosen prior to analysis based on probability thresholds that gave high positive predictive value (PPV) for type 1 or 2 diabetes in previous literature [19]: T1D $\geq$ 80% probability and T2D $<$5% probability, for defining T1D a further cut-off of 20% probability, were evaluated to give a high PPV while aiming to capture a high percentage of all T1D cases. Insulin within a year of diagnosis and oral hypoglycaemic agents (OHA) treatment are well-reported to associate with T1D and T2D, respectively [10]. Therefore, as an additional analysis, performance of approaches was further evaluated with the addition of knowledge of insulin within a year of diagnosis, defined as insulin treatment within a year of diagnosis, or also by current treatment with any OHA. Full details for each approach are given in ESM methods.

2.3. Biological definitions of diabetes type approaches evaluated against

2.3.1. UK Biobank

We have recently shown that measuring the average polygenic susceptibility to T1D (captured by a [T1DGRS]) of a cohort with diabetes can allow the proportion of T1D in that cohort to be estimated based on enrichment for genetic susceptibility to T1D over and above population susceptibility, as described in statistical analysis below [3,27]. Importantly, at an individual level a high genetic susceptibility for T1D does not prevent a person having T2D and those developing T1D can do so without T1D genetic risk [33]. Therefore, this analysis is evaluated within a cohort, as on average, those with T1D will have a significantly higher genetic predisposition to T1D than those without [29,30]. Calculations of proportions with and without T1D using this method are estimates but have been previously shown to be robust with the accuracy and precision of these estimates discussed in detail elsewhere [27]. Full details of T1DGRS generation used are given in ESM methods.

2.3.2. Diabetes Alliance for Research in England

T1D was defined as severe insulin deficiency: measured non-fasting C-peptide $<$200 pmol/L. T2D was defined as participants currently insulin-treated with a C-peptide $\geq$200 pmol/L. All analyzed participants had a duration of diabetes at C-peptide measurement of over 3 years [13].

2.4. Statistical analysis

When classifying all insulin-treated cases, approaches were ranked by the overall accuracy of each definition,

defined as the proportion of all T1D and T2D cases correctly classified relative to the total number of all cases classified. For each approach, the PPV of cases classified as T1D and T2D (percent of those identified who have the condition as defined by the biological standard) and sensitivity for detecting T1D and T2D (percentage of cases with the condition identified) were also calculated. Where aiming to classify just a T1D or T2D cohort, approaches were ranked firstly based on PPV and then secondly by sensitivity.

2.4.1. UK Biobank

For each classification approach, the mean T1DGRS for cases classified as T1D ($Approach_{Called\ T1D}$) and T2D ($Approach_{Called\ T2D}$) were separately evaluated against mean T1DGRS for reference T1D cases ($Reference_{T1D}$) (n = 6,483 mean T1DGRS = 14.50) and reference T2D equivalent cohort ($Reference_{T2D}$) (n = 9,246 mean T1DGRS = 10.37), both taken from the T1D genetics consortium [34]. Reference T1D cases were White European, who were clinically diagnosed and aged <17 years at diagnosis. The higher the proportion of diabetes cases correctly defined by a classification approach, the more the T1DGRS of the groups classified as T1D or T2D will, respectively, genetically resemble true T1D and T2D reference populations (method shown in ESM Figure 2). The proportion of T1D within groups, defined by each classification approach, is then estimated according to the normalized difference of each clinical definitions mean T1DGRS ($Approach_{Called\ (T1D/T2D)}$) and the mean T1DGRS of the two reference populations ($Reference_{T1D}$ and $Reference_{T2D}$) in the equations below and as described previously [27,35]. For cases defined as having T1D by each classification approach, PPV for T1D is equivalent to $Proportion_{T1D}$. For cases defined as having T2D by each classification approach, PPV for T2D is calculated as 1- $Proportion_{T1D}$.

$$Proportion(PPV)_{T1D} = \left| \frac{ApproachT1DGRS_{CalledT1D} - ReferenceT1DGRS_{T2D}}{ReferenceT1DGRS_{T1D} - ReferenceT1DGRS_{T2D}} \right|$$

$$Proportion(PPV)_{T2D} = 1 - \left| \frac{ApproachT1DGRS_{CalledT2D} - ReferenceT1DGRS_{T2D}}{ReferenceT1DGRS_{T1D} - ReferenceT1DGRS_{T2D}} \right|$$

Sensitivity was estimated as

$$Sensitivity\ T1D = \left| \frac{(PPV_{T1D} \times n_{T1D})}{(PPV_{T1D} \times n_{T1D}) + ((1 - PPV_{T2D}) \times n_{T2D})} \right|$$

Where n_{T1D} is the number of cases called as having T1D and n_{T2D} is the number of cases called as having T2D by each approach.

2.5. Determining accuracy in UK Biobank and DARE

Where all insulin-treated participants were classified as having either T1D or T2D, accuracy was calculated as:

$$Accuracy = \left| \frac{(PPV_{T1D} \times n_{T1D}) + (PPV_{T2D} \times n_{T2D})}{(n_{T1D} \times n_{T2D})} \right|$$

All analyses were performed using Stata 16 (StataCorp LP, College Station, TX).

3. Results

3.1. Performance of approaches to classify all insulin-treated White European participants with diabetes in UK Biobank

Within the UKBB, of the White European participants meeting eligibility criteria, 21% (3,534/16,619) were insulin treated. The clinical characteristics of all participants split by insulin treatment status are shown in ESM Table 1. In the 13,085 participants with diabetes not currently insulin treated, the mean T1DGRS (10.32, SD 2.38) was consistent with a classical nonT1D reference population [34] mean T1DGRS (10.37 SD 2.26), suggesting little to no T1D in this group. The genetically assessed estimated performance of classification approaches to classify all insulin-treated diabetes cases as either T1D or T2D ranked by accuracy are shown in Table 2.

The median classification accuracy was 85% and varied substantially by approach (range 71% to 88%). The highest accuracy overall was insulin within a year of diagnosis combined with the *clinical model* overall correctly classifying 87% and *lipid model* overall correctly classifying

Table 1. Diabetes specific factors required for each approach and the different cut offs required for classifying all cases, or defining T1D or T2D. Where available cut offs were taken from existing literature [10,15-19,22]

		Cut offs used and reference code		
		Whole cohort		
Reference name (approach number)	**Clinical information required**	**For defining T1D remainder T2D**	**For defining T1D only**	**For defining T2D only**
Age (1)	Age at diagnosis	<35 yr [10]	≤20 yr [16]	≥40 yr [10]
BMI (2)	Current BMI	≤25 kg/m^2 [14]	≤23 kg/m^2 [10]	≥28 kg/m^2 [10]
Clinical model (3)	Current BMI, age at diagnosis	Model probability ≥ 12% [19]	Model probability ≥ 80%[a]	Model probability < 5%[a]
Lipid mod (4)	Current BMI, age at diagnosis, Sex, HDL, triglyceride, and total cholesterol	Model probability ≥ 12% [19,20]	Model probability ≥ 80%[a]	Model probability < 5%[a]
ICD codes (5)	(ICD 10 or 9 code), OHA, age at diagnosis, and DKA episode history	Algorithm T1D [18]	N/A	N/A
UKBB algorithm (6)	Age at diagnosis, time to Insulin, nonmetformin OHA, and interviewreport of T1D, ethnicity	Possible and probable T1D [15]	Probable T1D[13]	Probable T2D [13]
Interview reported (7)	Interview-reported diabetes type	Interview-reported diabetes T1D [32]	N/A	N/A
Diagnosis codes algorithm (8)	Diabetes diagnosis codes, non metformin OHA, prescription for glucagon, and prescription for urine acetone strip	Ratio of T1D to T2D diagnosis codes >0.5 with either glucagon, non metformin OHA prescription, or prescription of urine acetone strip alone [16]	N/A	N/A
Diagnosis code + age (9)	Diabetes diagnosis codes and age at diagnosis.	Any diagnosis code of T1D or age at diagnosis < 22 yr [17]	N/A	N/A
Majority diagnosis codes (10)	Diabetes diagnosis codes	Ratio of TID to T2D diagnosis codes >0.5 [22]	N/A	N/A

Abbreviations: T1D, type 1 diabetes; T2D, type 2 diabetes; BMI, body mass index; CI, confidence interval; UKBB, UK Biobank; DKA, diabetic ketoacidosis.

[a] For the previously published models, cut offs were not available for selecting pure T1D and T2D cohorts so pragmatic values were chosen from published data aiming for 100% and >90% PPV for T1D classification and 100% PPV for T2D classification [19].

88%. Interview-reported diabetes type, with or without insulin, within a year of diagnosis had an accuracy of 87% but was available in just 15% (519/3,534) of all cases. For the majority of approaches, adding insulin within a year of diagnosis to define T1D substantially improved accuracy with the absence of OHA treatment only slightly less accurate, ESM Table 2. The lowest accuracy was seen in approaches using simple cut-offs for individual variables, such as *age of diagnosis (<35 years) 82%* and *BMI (≤25 kg/m^2) 71%*. In the 47% (1,644/3,534) of the insulin-treated cohort with linked primary care data, diabetes *diagnosis codes algorithm* alongside insulin within a year of diagnosis gave the highest accuracy of approaches that incorporate electronic health care record data and diagnosis codes at 85%. For direct comparison ESM Table 3 gives the performance of other classification approaches in this reduced subset of the dataset with linked primary care records, with results broadly similar in this subset.

3.2. *Performance of approaches to classifying all insulin-treated participants with diabetes in DARE*

In the DARE cohort, we identified 1,241 people with diabetes who met our inclusion criteria, 63% (784/1,241) were insulin treated with 42% (333/784) having a C-peptide <200 pmol/L consistent with T1D, at a median duration of 18 years. Table 3 gives the performance of classification approaches to classify all insulin-treated diabetes cases as either T1D or T2D against a C-peptide definition of diabetes type. Accuracy values and overall ranking of approaches were similar to when the diabetes type was defined genetically in UKBB, with a median

Table 2. Comparative performance of approaches classifying all insulin treated White European participants with diabetes in UKBB

	Called T1D			Called T2D			
Approach	**(*n*)**	**PPV**	**Sensitivity**	**(*n*)**	**PPV**	**Sensitivity**	**Accuracy**
Lipid model probability ≥12% and insulin within year of diagnosis	1,169	87% (84-90)	79% (77-81)	2,365	88% (86-91)	93% (92-94)	88%
Clinical model probability ≥12% and insulin a within year of diagnosis	1,047	89% (86-92)	72% (70-75)	2,487	86% (83-88)	95% (94-96)	87%
Interview-reported diabetes type (n = 519 available) and insulin within a year of diagnosis	224	85% (77-92)	86% (81-91)	295	89% (82-97)	89% (85-92)	87%
Interview-reported diabetes type (n = 519 available)	253	80% (73-87)	92% (88-95)	266	93% (86-101)	83% (79-87)	87%
UKBB probable & possible T1D and insulin within a year of diagnosis	988	90% (87-93)	69% (66-71)	2,546	84% (82-87)	96% (95-96)	86%
ICD algorithm and insulin within a year of diagnosis	1,025	89% (86-92)	71% (68-73)	2,509	85% (82-87)	95% (94-96)	86%
UKBB probable & possible T1D and insulin within a year of diagnosis (no interview report)	918	93% (89-96)	66% (63-68)	2,616	83% (81-85)	97% (96-98)	85%
ICD algorithm	1,184	82% (79-85)	75% (73-78)	2,350	86% (84-89)	91% (90-92)	85%
Age diabetes diagnosed <35 yr and insulin within a year of diagnosis	867	93% (89-96)	62% (59-65)	2,667	82% (79-84)	97% (96-98)	84%
Lipid model probability ≥12%	1,501	74% (71-77)	86% (84-88)	2,033	91% (89-94)	83% (81-84)	84%
UKBB probable & possible type 1 diabetes (no interview report)	1,142	80% (77-83)	70% (68-73)	2,392	84% (81-87)	90% (88-91)	83%
UKBB probable & possible T1D	1,231	78% (75-81)	74% (72-77)	2,303	86% (83-88)	88% (87-89)	83%
Clinical model probability ≥12%	1,325	76% (73-79)	78% (76-80)	2,209	87% (84-90)	86% (84-87)	83%
Age diabetes diagnosed <35 yr	1,065	80% (77-84)	66% (64-69)	2,469	82% (80-85)	91% (89-92)	82%
BMI ≤25 (kg/m^2) and insulin within a year of diagnosis	511	80% (75-85)	32% (29-34)	3,023	71% (68-73)	95% (95-96)	72%
BMI ≤25 (kg/m^2)	658	70% (65-74)	35% (33-38)	2,876	71% (68-73)	91% (90-92)	71%

Abbreviations: T1D, type 1 diabetes; T2D, type 2 diabetes; PPV, positive predictive value; BMI, body mass index; CI, confidence interval; UKBB, UK Biobank.

Cases are classified as T1D if they meet the stated criteria, and are otherwise classified as T2D. Results ranked by accuracy (total correctly classified) then T1D PPV. Brackets signify 95% CI, positive predictive value (PPV).

accuracy of 83% (range 68–88%). In DARE, the clinical model combined with insulin within a year of diagnosis had an accuracy of 85%. Interview-reported diabetes type alone gave the highest accuracy of 88%. The Biobank algorithm (incorporating interview-reported diabetes type) with insulin within a year of diagnosis had an accuracy of 87%. This reduced to 84% when interview-reported diabetes type was not included within the algorithm. Again all methods were improved by adding insulin within a year of diagnosis. The 451 noninsulin-treated participants with C-peptide measured at 99.6% (449/451), had a C-peptide ≥200 consistent with T2D.

3.3. Performance of approaches to optimally identify type 1 and type 2 diabetes among insulin-treated participants with diabetes

The performance of methods to optimally identify T1D, ranked by PPV in UKBB (percent of those identified as T1D who have the condition genetically) are shown in Table 4. A pure T1D cohort was generated when insulin within a year of diagnosis was combined with either age at diagnosis ≤20 years (PPV 100%) or a clinical model probability ≥80% (PPV 99%). However, these approaches had low sensitivity respectively only identifying 33% and 37% of all T1D cases. Using probable T1D in the Biobank algorithm combined with insulin within a year of diagnosis identified 69% of all T1D cases, with a PPV of 90%. This was similar to using a lower clinical model probability of ≥20% identifying 67% of all T1D cases with a PPV of 91%. Comparable results for the majority of approaches for both PPV and sensitivity of T1D identified were achieved in DARE, using C-peptide-defined diabetes type, Table 4.

The performance of methods to optimally identify T2D, ranked by PPV in UKBB (percent of those identified as T2D who have the condition genetically) are shown in ESM Table 4. A pure T2D cohort was generated using

Table 3. Comparative performance of approaches classifying all insulin-treated participants with diabetes in DARE

	Called T1D			Called T2D			
Approach	**(*n*)**	**PPV**	**Sensitivity**	**(*n*)**	**PPV**	**Sensitivity**	**Accuracy**
Interview-reported diabetes type and insulin within a year of diagnosis	310	89% (85-92)	82% (78-86)	474	88% (85-91)	92% (90-95)	88%
Interview-reported diabetes type	335	86% (83-90)	87% (83-90)	449	90% (87-93)	90% (87-93)	88%
UKBB probable & possible T1D and insulin within a year of diagnosis (including interview report)	325	86% (82-90)	84% (80-88)	459	88% (85-91)	90% (87-93)	87%
Clinical model probability ≥12% and insulin a within year of diagnosis	278	90% (86-93)	75% (70-79)	506	83% (80-86)	94% (91-96)	85%
UKBB probable & possible T1D and insulin within a year of diagnosis (no interview report)	257	90% (87-94)	69% (65-74)	527	81% (77-84)	94% (92-97)	84%
UKBB probable & possible T1D (including interview report)	392	76% (72-80)	89% (86-92)	392	91% (88-93)	79% (75-83)	83%
Age diabetes diagnosed <35 yr and insulin within a year of diagnosis	242	90% (86-94)	65% (60-70)	542	79% (75-82)	95% (93-97)	82%
Clinical model probability ≥12%	346	78% (74-82)	81% (77-85)	438	85% (82-89)	83% (80-87)	82%
UKBB probable & possible T1D (no interview report)	300	81% (77-85)	73% (68-78)	484	81% (78-85)	87% (84-90)	81%
Age diabetes diagnosed <35 yr	280	80% (76-85)	67% (62-72)	504	78% (75-82)	88% (85-91)	79%
BMI ≤25 (kg/m^2) and insulin within a year of diagnosis	140	87% (82-93)	37% (31-42)	644	67% (63-71)	96% (94-98)	71%
BMI ≤25 (kg/m^2)	187	72% (66-79)	40% (35-46)	597	67% (63-70)	88% (85-91)	68%

Abbreviations: T1D, type 1 diabetes; T2D, type 2 diabetes; PPV, positive predictive value; BMI, body mass index; CI, confidence interval; UKBB, UK Biobank.

Cases are classified as T1D if they meet the stated criteria and are otherwise classified as T2D. Results ranked by accuracy (total correctly classified) then T1D PPV. Brackets signify 95% CI, positive predictive value (PPV). Lipid model and Diagnosis codes not evaluated as unavailable in DARE.

probable T2D with the Biobank algorithm and PPV of 100% but this had low sensitivity capturing at just 17% of all insulin-treated T2D cases. A clinical model probability <5% gave a T2D PPV of 94% and captured 67% of all T2D cases. Adding absence of insulin within a year of diagnosis to all definitions of T2D increased T2D PPV in all approaches but resulted in a lower proportion of all T2D cases being captured. Comparable results for both PPV and sensitivity for each approach were achieved in DARE using C-peptide-defined diabetes type, ESM Table 4.

3.4. Performance of approaches to classifying all insulin-treated participants with diabetes in UK Biobank

As an exploratory analysis, we evaluated the performance of approaches to classify all participants with diabetes in UKBB regardless of ethnicity. Within the 4,845 insulin-treated participants, the overall performance of approaches was similar when the analysis was undertaken in just White Europeans, with a median accuracy of 85% (range 75% to 91%) with the best accuracy achieved using probability models combined with insulin within a year of diagnosis: lipid model 91% and clinical features-only model 90%, ESM Table 5.

3.5. Development of algorithm for optimal approach selection

We developed a pragmatic online tool for researchers to select the optimum approach of those evaluated for classifying insulin-treated diabetes cases in research datasets, based on the findings in UKBB: Classifying Diabetes for Research: Method Selector (newcastlerse.github.io). The optimal approach varies based on the research question being asked and the diabetes outcomes available in the dataset being used. ESM Appendix 2 provides researchers the R code to implement all methods, which is also provided within the online tool.

4. Discussion

We evaluated the performance of approaches for classifying the diabetes type in two different population-level research datasets: UKBB and DARE. Results were consistent across datasets despite using two different biological definitions of diabetes type. The impact of classification approach selection on study results and conclusions is highlighted by the marked variation in accuracy observed in our study. Across the two different datasets combining insulin within

Table 4. Comparative performance of approaches classifying T1D with minimal misclassification in UKBB and DARE in insulin- treated participants

	UK Biobank		DARE	
Approach	**PPV of cases called T1D**	**Sensitivity for identifying T1D**	**PPV of cases called T1D**	**Sensitivity for identifying T1D**
Age diabetes diagnosed ≤20 yr and insulin within a year of diagnosis	100% (99-100)	33% (30-35)	96% (93-100)	40% (32-49)
Clinical model probability ≥80% and insulin within a year of diagnosis	99% (98-100)	37% (34-39)	96% (93-99)	47% (39-54)
Lipid model probability ≥80% and insulin within a year of diagnosis	97% (95-98)	40% (38-43)	n/a	n/a
Lipid model probability ≥80%	92% (90-94)	42% (39-45)	n/a	n/a
Age diabetes diagnosed ≤20 yr	92% (90-95)	34% (31-36)	96% (92-99)	40% (32-49)
Clinical model probability ≥20% and insulin within a year of diagnosis	91% (89-92)	67% (65-70)	91% (88-95)	70% (65-76)
Clinical model probability ≥80%	91% (88-93)	37% (35-40)	93% (90-97)	47% (39-55)
UKBB probable T1D and insulin within a year of diagnosis	90% (88-92)	69% (66-71)	86% (82-90)	84% (80-88)
Lipid model probability ≥20% and insulin within a year of diagnosis	89% (87-91)	75% (73-77)	n/a	n/a
UKBB probable T1D	89% (87-91)	70% (67-72)	84% (80-88)	88% (84-91)
BMI ≤23 (kg/m^2) and insulin within a year of diagnosis	82% (78-87)	16% (14-18)	90% (83-97)	19% (10-28)
Interview-reported T1D	80% (75-85)	92% (88-95)	86% (83-90)	87% (83-90)
Lipid model probability ≥20%	80% (78-82)	81% (79-84)	n/a	n/a
Clinical model probability ≥20%	80% (78-83)	71% (69-74)	84% (80-88)	74% (69-79)
BMI ≤23 (kg/m^2)	75% (70-80)	17% (15-19)	79% (70-87)	20% (11-28)

Abbreviations: T1D, type 1 diabetes; T2D, type 2 diabetes; PPV, positive predictive value ; BMI, body mass index; CI, confidence interval; UKBB, UK Biobank; DARE, Diabetes Alliance for Research in England.

Cases are classified as T1D if they meet the stated criteria, and are otherwise classified as T2D. Results ranked in UKBB by T1D PPV then sensitivity for identifying T1D. Analysis in UKBB restricted to White Europeans. Cut offs chosen to give high T1D PPV as per Table 1. Brackets signify 95% CI, Positive predictive value (PPV).

a year of diagnosis with T1D models incorporating BMI and age at diagnosis (*clinical model*), and these features with lipids (*lipid model*) consistently achieved the highest accuracy for classifying all insulin-treated participants (≥85%). Interview-reported diabetes type showed similar accuracy in both UKBB and the DARE cohort but was only recorded in the minority (15%) of UKBB participants, limiting its utility.

Our results suggest that probability models combined with insulin within a year of diagnosis provide a highly accurate approach to classifying research cohorts with insulin- treated diabetes. As a simple alternative, interview-reported diabetes type can be used although this was only available in 15% of UKBB participants. Why a low percentage of participants-reported diabetes type in UKBB is unclear. To explore this, we compared those with and without interview-reported diabetes type which suggested a slight trend towards more T1D in those reporting a diabetes type at interview but no major differences in clinical features (ESM Table 6). Furthermore, with recent changes in guidance for biomarker testing in national and international guidance, [36–39] it is possible, clinical diagnosis and therefore interview-reported diagnosis may become more accurate over time. This study also highlights the limitations of using single cutoffs, particularly age of diagnosis, likely to reflect the finding that nearly half of all T1D cases occur after 30 years of age [3–5,14]. All approaches are improved by adding variables capturing either insulin within a year of diagnosis or current OHA treatment. It was possible to identify pure T1D cohorts in both datasets through the use of a combination of early insulin treatment and either high-model probability or very young age at diagnosis.

A key strength of our study was that performance was evaluated against biological definitions of diabetes type. This reduces the potential for inaccuracies and bias if testing against clinical definitions, which are subject to both error and circularity (with features accurate for clinical classification reflecting features clinicians consider important) [7,8,10]. The main analysis in UKBB was restricted to White European participants, where the T1DGRS has been validated. As an exploratory analysis, we evaluated all participants to show that the ranking of approaches remained similar (meaning the optimum approach remains valid) even if the absolute accuracy of approaches in all nonWhite European ethnicities should be interpreted with caution. While all ethnicities were included in DARE, 99% of participants were White European.

Few studies have compared different classification methods to robust biomarker-defined diabetes types. In a cohort with insulin-treated diabetes, Hope et al. evaluated the performance of age of diagnosis <35 to classify diabetes cases with T1D defined by C-peptide deficiency and cases with preserved C-peptide defined as T2D [10]. Age at diagnosis correctly classified 85% of all cases in their study comparable within our study: 82% in UKBB and 79% in DARE. This remained comparable when age of diagnosis was combined with insulin within a year of diagnosis: Hope et al. study's accuracy of 87% vs. 84% UKBB and 82% DARE. Model performance was also high when previously assessed against diabetes subtypes defined by pancreatic histology [26]. The importance of insulin treatment in helping initially determine diabetes type in research datasets is emphasized by the genetic susceptibility of all participants not currently insulin-treated being consistent with little to no T1D in this group. In DARE, the absence of insulin treatment was also almost never associated with C-peptide deficiency mirroring previous studies defining diabetes type using C-peptide [40].

Limitations to our study include the fact that both the Biobank algorithm (developed in UKBB) and the T1D clinical model (developed in a cohort that included DARE) were evaluated in the same cohorts they were developed in. Reassuringly both methods performed comparatively well in the alternative data set they were not developed in, suggesting any bias was minimal. Despite using both T1D probability models in all participants even though they were developed in adults aged 18-50 years, they were consistently high-performing approaches in both datasets [19,20]. It is possible that accuracy could have been further improved by varying cutoffs in older adults; however this would have risked being over fitted. Lipids in UKBB were also non-fasted, in contrast to the model development dataset, and it is therefore possible that performance would increase where fasted lipids are available [20]. Using genetic predisposition to T1D can be helpful in diabetes classification, in the original development of the clinical model adding T1DGRS improved performance, [19] and we would recommend using this when genetic data is available, however as T1DGRS was our outcome we were unable to evaluate this approach. Islet autoantibodies used in combination with clinical models also improve performance [19] but are rarely available in research datasets as is the case in UKBB. Classifying diabetes as only being T1D or T2D will miss other types of diabetes. Reassuringly, in DARE just 2% (29/1,241) of the cohort had a clinician diagnosis which was not T1D or T2D. T1DGRS is known to modestly reduce with increasing age of T1D diagnosis, [41−43] and our T1D reference cohort was diagnosed at <17 years of age. In previous studies, the mean T1DGRS of those with confirmed T1D diagnosed over 18 years of age was 2.5% lower than those diagnosed at <18 years of age [44]. Given that over half of T1D develops in adults, this means our genetically estimated T1D prevalence will be a slight underestimate. However, the comparative performance results, as in the same datasets remains robust, and reassuringly, in DARE defining diabetes type by C-peptide similar results were found. It is possible that interview-reported diabetes type could be influenced by the research staff conducting the interviews and there appears a subtle suggestion of bias towards T1D in those reporting vs. not reporting diabetes type in UKBB. While other methods of collecting self-report may potentially have lower accuracy, recent research has found similar PPV of 83% and sensitivity of 92% for T1D when assessing self-reported diabetes type via a telephone survey [32]. It has also previously been reported that UKBB is not truly representative of the UK population due to participants being from less deprived areas, and more predominantly of White ethnicity than the general population [45]. These issues with recruited population level research datasets are not unique to UKBB but caution should be used while applying these findings to nonWhite or low income populations [46].

Our results are important for all researchers studying type 1 or 2 diabetes. The considerable differences in pathophysiology, treatment, and associated risks of T1D and T2D means of inadvertently studying mixed cohorts could lead to misleading study findings [47]. Our results allow determination of the optimal approach for classifying insulin- treated diabetes cases while also confirming that noninsulin-treated cases of over 3 years of duration can confidently be labelled as having T2D. Approaches can be selected based on which diabetes specific outcomes are available and the research question being asked. An added advantage of our study is that researchers can understand the accuracy of the approach used and how this might impact their results and their relatability to other studies where different approaches may have been used. For ease, our findings have been translated into an online tool allowing researchers to determine and then implement the optimal approach for their research question and dataset.

5. Conclusion

With two separate research datasets and using two different biological definitions of diabetes, we show the performance of approaches for classifying insulin-treated diabetes type for research studies and translate this into an online tool for optimal approach selection for researchers. Interview-reported diabetes type diagnosis and models combining continuous features are the most accurate methods of classifying insulin-treated diabetes in research datasets without measured classification biomarkers.

Acknowledgments

The authors thank participants who took part in the study and the research teams who undertook cohort

recruitment. This research has in part been conducted using UK Biobank Resource. Biobank application 9055. The authors are grateful to Mike Simpson of Newcastle University for developing the online tool.

Appendix A

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2022.10.022.

References

[1] Group SDD. Scottish Diabetes Survey 2019. Scottish Diabetes Survey; 2019. Available at https://www.diabetesinscotland.org.uk/wp-content/uploads/2020/10/Diabetes-Scottish-Diabetes-Survey-2019.pdf. Accessed June 1, 2022: In press.

[2] American Diabetes Association. 2. Classification and diagnosis of diabetes. Diabetes Care 2017;40:S11−24.

[3] Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. Lancet Diabetes Endocrinol 2018;6(2):122−9.

[4] Diaz-Valencia PA, Bougneres P, Valleron AJ. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. BMC Public Health 2015;15:255.

[5] Bruno G, Gruden G, Songini M. Incidence of type 1 diabetes in age groups above 15 years: facts, hypothesis and prospects for future epidemiologic research. Acta Diabetol 2016;53(3):339−47.

[6] Leslie RD, Evans-Molina C, Freund-Brown J, Buzzetti R, Dabelea D, Gillespie KM, et al. Adult-onset type 1 diabetes: current understanding and challenges. Diabetes Care 2021;44:2449−56.

[7] Foteinopoulou E, Clarke CAL, Pattenden RJ, Ritchie SA, McMurray EM, Reynolds RM, et al. Impact of routine clinic measurement of serum C-peptide in people with a clinician-diagnosis of type 1 diabetes. Diabetic Med 2020;38:e14449.

[8] Thomas NJ, Lynam AL, Hill AV, Weedon MN, Shields BM, Oram RA, et al. Type 1 diabetes defined by severe insulin deficiency occurs after 30 years of age and is commonly treated as type 2 diabetes. Diabetologia 2019;62:1167−72.

[9] Munoz C, Floreen A, Garey C, Karlya T, Jelley D, Alonso GT, et al. Misdiagnosis and diabetic ketoacidosis at diagnosis of type 1 diabetes: patient and caregiver perspectives. Clin Diabetes 2019;37(3):276−81.

[10] Hope SV, Wienand-Barnett S, Shepherd M, King SM, Fox C, Khunti K, et al. Practical Classification Guidelines for Diabetes in patients treated with insulin: a cross-sectional study of the accuracy of diabetes diagnosis. Br J Gen Pract 2016;66(646):E315−22.

[11] Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K. Incorrect and incomplete coding and classification of diabetes: a systematic review. Diabetic Med 2010;27(5):491−7.

[12] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting diabetes mellitus with machine learning techniques. Front Genet 2018;9:515.

[13] Jones AG, Hattersley AT. The clinical utility of C-peptide measurement in the care of patients with diabetes. Diabetic Med 2013;30(7):803−17.

[14] Harding JL, Wander PL, Zhang X, Li X, Karuranga S, Chen H, et al. The incidence of adult-onset type 1 diabetes: a systematic review from 32 countries and regions. Diabetes Care 2022;45:994−1006.

[15] Eastwood SV, Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK biobank. PLoS One 2016;11:e0162388.

[16] Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. Diabetes Care 2013;36:914−21.

[17] Lethebe BC, Williamson T, Garies S, McBrien K, Leduc C, Butalia S, et al. Developing a case definition for type 1 diabetes mellitus in a primary care electronic medical record database: an exploratory study. CMAJ Open 2019;7(2):E246−51.

[18] Lo-Ciganic W, Zgibor JC, Ruppert K, Arena VC, Stone RA. Identifying type 1 and type 2 diabetic cases using administrative data: a tree-structured model. J Diabetes Sci Technol 2011;5(3):486−93.

[19] Lynam A, McDonald T, Hill A, Dennis J, Oram R, Pearson E, et al. Development and validation of multivariable clinical diagnostic models to identify type 1 diabetes requiring rapid insulin therapy in adults aged 18-50 years. BMJ Open 2019;9(9):e031586.

[20] Lynam AL, Dennis JM, Owen KR, Oram RA, Jones AG, Shields BM, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. Diagn Progn Res 2020;4:6.

[21] Practitioner RCoG. NHS Diabetes Coding, classification and diagnosis of diabetes A review of the coding, classification and diagnosis of diabetes in primary care in England with recommendations for improvement. 2011. Available at https://orchid.phc.ox.ac.uk/wp-content/uploads/2017/02/nhs_diabetes_and_rcgp_cod_final_report.pdf. Accessed June 1, 2022.

[22] Schroeder EB, Donahoo WT, Goodrich GK, Raebel MA. Validation of an algorithm for identifying type 1 diabetes in adults based on electronic health record data. Pharmacoepidemiol Drug Saf 2018;27(10):1053−9.

[23] Sharma M, Petersen I, Nazareth I, Coton SJ. An algorithm for identification and classification of individuals with type 1 and type 2 diabetes mellitus in a large primary care database. Clin Epidemiol 2016;8:373−80.

[24] Weisman A, Tu K, Young J, Kumar M, Austin PC, Jaakkimainen L, et al. Validation of a type 1 diabetes algorithm using electronic medical records and administrative healthcare data to study the population incidence and prevalence of type 1 diabetes in Ontario, Canada. BMJ Open Diabetes Res Care 2020;8(1):e001224.

[25] Zhong VW, Pfaff ER, Beavers DP, Thomas J, Jaacks LM, Bowlby DA, et al. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study. Pediatr Diabetes 2014;15(8):573−84.

[26] Carr ALJ, Perry DJ, Lynam AL, Chamala S, Flaxman CS, Sharp SA, et al. Histological validation of a type 1 diabetes clinical diagnostic model for classification of diabetes. Diabet Med 2020;37(12):2160−8.

[27] Evans BD, Słowiński P, Hattersley AT, Jones SE, Sharp S, Kimmitt RA, et al. Estimating disease prevalence in large datasets using genetic risk scores. Nat Commun 2021;12(1):6441.

[28] Allen NE, Sudlow C, Peakman T, Collins R, Biobank UK. UK biobank data: come and get it. Sci Transl Med 2014;6(224):224ed4.

[29] Oram RA, Patel K, Hill A, Shields B, McDonald TJ, Jones A, et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. Diabetes Care 2015;39:337−44.

[30] Patel KA, Oram RA, Flanagan SE, De Franco E, Colclough K, Shepherd M, et al. Type 1 diabetes genetic risk score: a novel tool to discriminate monogenic and type 1 diabetes. Diabetes 2016;65:2094−9.

[31] Tyrrell J, Jones SE, Beaumont R, Astley CM, Lovell R, Yaghootkar H, et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. BMJ 2016;352:i582.

[32] Nooney JG, Kirkman MS, Bullard KM, White Z, Meadows K, Campione JR, et al. Identifying optimal survey-based algorithms to distinguish diabetes type among adults with diabetes. J Clin Transl Endocrinol 2020;21:100231.

[33] Mishra R, Chesi A, Cousminer DL, Hawa MI, Bradfield JP, Hodge KM, et al. Relative contribution of type 1 and type 2 diabetes loci to the genetic etiology of adult-onset, non-insulin-requiring autoimmune diabetes. BMC Med 2017;15(1):88.

[34] Rich SS, Akolkar B, Concannon P, Erlich H, Hilner JE, Julier C, et al. Overview of the type I diabetes genetics consortium. Genes Immun 2009;10 Suppl 1:S1−4.

[35] Sukcharoen K, Sharp SA, Thomas NJ, Kimmitt RA, Harrison J, Bingham C, et al. IgA nephropathy genetic risk score to estimate the prevalence of IgA nephropathy in UK biobank. Kidney Int Rep 2020;5(10):1643−50.

[36] Holt RIG, DeVries JH, Hess-Fischl A, Hirsch IB, Kirkman MS, Klupa T, et al. The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 2021;64:2609−52.

[37] Guideline N. Type 1 diabetes in adults: diagnosis and management 2022. Available at https://www.nice.org.uk/guidance/ng17/chapter/rationale-and-impact#diagnosis. Accessed June 2, 2022.

[38] Tatovic D, Jones AG, Evans C, Long AE, Gillespie K, Besser REJ, et al. Diagnosing type 1 diabetes in adults: guidance from the UK T1D immunotherapy consortium. Diabet Med 2022;39:e14862.

[39] Diabetologists AoBC. Standards of care for management of adults with type 1 diabetes 2017 2017. Available at https://abcd.care/sites/abcd.care/files/resources/Standards_of_Care_T1DM_ABCD_FINAL.pdf. Accessed April 14, 2020.

[40] Shields BM, Peters JL, Cooper C, Lowe J, Knight BA, Powell RJ, et al. Can clinical features be used to differentiate type 1 from type 2 diabetes? A systematic review of the literature. BMJ Open 2015;5(11):e009088.

[41] Graham J, Hagopian WA, Kockum I, Li LS, Sanjeevi CB, Lowe RM, et al. Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. Diabetes 2002;51:1346−55.

[42] Howson JM, Rosinger S, Smyth DJ, Boehm BO, Group A-ES, Todd JA. Genetic analysis of adult-onset autoimmune diabetes. Diabetes 2011;60:2645−53.

[43] Perry DJ, Wasserfall CH, Oram RA, Williams MD, Posgai A, Muir AB, et al. Application of a genetic risk score to racially diverse type 1 diabetes populations demonstrates the need for diversity in risk-modeling. Sci Rep 2018;8(1):4529.

[44] Thomas NJ, Walkey HC, Kaur A, Misra S, Oliver NS, Colclough K, et al. The relationship between islet autoantibody status and the genetic risk of type 1 diabetes in adult-onset type 1 diabetes. Diabetologia 2022. https://doi.org/10.1007/s00125-022-05823-1. In press.

[45] Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am J Epidemiol 2017;186:1026−34.

[46] Manolio TA, Weis BK, Cowie CC, Hoover RN, Hudson K, Kramer BS, et al. New models for large prospective studies: is there a better way? Am J Epidemiol 2012;175:859−66.

[47] Jones AG, McDonald TJ, Shields BM, Hagopian W, Hattersley AT. Latent Autoimmune Diabetes of Adults (LADA) is likely to represent a mixed population of autoimmune (Type 1) and nonautoimmune (Type 2) diabetes. Diabetes Care 2021;44:1243−51.