BMJ Open Diabetes Research & Care

Comparison of several survey-based algorithms to ascertain type 1 diabetes among US adults with self-reported diabetes

Sarah S Casagrande ,[1] Sarah E Lessem,[2] Trevor J Orchard,[3] Kai McKeever Bullard,[4] Linda S Geiss,[4] Sharon H Saydah,[4] Andy Menke ,[1] Giuseppina Imperatore,[4] Keith F Rust,[5] Catherine C Cowie[6]

To cite: Casagrande SS, Lessem SE, Orchard TJ, *et al*. Comparison of several survey-based algorithms to ascertain type 1 diabetes among US adults with self-reported diabetes. *BMJ Open Diab Res Care* 2020;**8**:e001917. doi:10.1136/bmjdrc-2020-001917

► Supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjdrc-2020-001917).

Received 23 September 2020
Revised 4 November 2020
Accepted 15 November 2020

For numbered affiliations see end of article.

Correspondence to
Dr Sarah S Casagrande; scasagrande@s-3.com

ABSTRACT

Introduction Defining type of diabetes using survey data is challenging, although important, for determining national estimates of diabetes. The purpose of this study was to compare the percentage and characteristics of US adults classified as having type 1 diabetes as defined by several algorithms.

Research design and methods This study included 6331 respondents aged ≥18 years who reported a physician diagnosis of diabetes in the 2016–2017 National Health Interview Survey. Seven algorithms classified type 1 diabetes using various combinations of self-reported diabetes type, age of diagnosis, current and continuous insulin use, and use of oral hypoglycemics.

Results The percentage of type 1 diabetes among those with diabetes ranged from 3.4% for those defined by age of diagnosis <30 years and continuous insulin use (algorithm 2) to 10.2% for those defined only by continuous insulin use (algorithm 1) and 10.4% for those defined as self-report of type 1 (supplementary algorithm 6). Among those defined by age of diagnosis <30 years and continuous insulin use (algorithm 2), by self-reported type 1 diabetes and continuous insulin use (algorithm 4), and by self-reported type 1 diabetes and current insulin use (algorithm 5), mean body mass index (BMI) (28.6, 27.4, and 28.5 kg/m^2, respectively) and percentage using oral hypoglycemics (16.1%, 11.1%, and 19.0%, respectively) were lower than for all other algorithms assessed. Among those defined by continuous insulin use alone (algorithm 1), the estimates for mean age and age of diagnosis (54.3 and 30.9 years, respectively) and BMI (30.9 kg/m^2) were higher than for other algorithms.

Conclusions Estimates of type 1 diabetes using commonly used algorithms in survey data result in varying degrees of prevalence, characteristic distributions, and potential misclassification.

Significance of this study

What is already known about this subject?

- Defining type of diabetes using survey data is challenging, although important, for determining national estimates of diabetes.
- Various algorithms for type 1 diabetes using survey data have been published, but no gold standard exists.

What are the new findings?

- A wide range of prevalence estimates and characteristic distributions were found according to algorithms used to define type 1 diabetes as well as potential misclassification.

How might these results change the focus of research or clinical practice?

- The findings from this study should inform future diabetes research that uses survey data, and algorithm choice may depend on the specific research question.
- Future research may link survey questions to medical records, providing additional information for choosing an algorithm for type 1 diabetes in survey data.
- These results demonstrate the need for more focus on how diabetes type is defined and used in survey research given the demonstrated variability in algorithms.

INTRODUCTION

Defining type of diabetes using survey data is challenging, although important, for determining national estimates of diabetes. The most common types are type 1 diabetes, an autoimmune disease in which the pancreatic cells virtually cease to make insulin, and thought to account for 5%–10% of diabetes cases; and type 2 diabetes, characterized by insulin resistance and insufficient beta cell insulin secretion.[1] Given the difference in etiology, natural history, and disease trajectories for type 1 and type 2 diabetes, it is important to distinguish type of diabetes for surveillance purposes and for studying risk factors and health outcomes that are associated with each type of diabetes. Many people do not know their type of diabetes, which may be due to poor patient education

or poor physician–patient communication; in addition, physicians may be unsure of patient type of diabetes and simply focus on treatments to achieve glycemic and other comorbidity control.[2 3] Validating self-report of type 1 diabetes is difficult without autoimmune testing, measured C-peptide, or medical records.[2] Historically, diabetes diagnosed before age 30 years with continuous insulin use soon after diagnosis was classified as type 1 diabetes; however, accumulating research finds that many type 1 diabetes cases are diagnosed at older ages.[4] Type 2 diabetes is also increasingly diagnosed in adolescence and young adulthood.[5]

The 2016–2017 National Health Interview Survey (NHIS) contained a supplement including questions about self-reported diabetes type, time between diabetes diagnoses and initiation of insulin use, if the respondent ever stopped insulin for more than 6 months, and if this was only during the first year after diagnosis. The 2016–2017 NHIS also included core questions about doctor-diagnosed diabetes, age of diagnosis, and use of oral hypoglycemic agents and insulin. Together, the NHIS core and supplement questions provide an opportunity to use nationally representative data to look at several algorithms for classifying diabetes type.[4 6]

Various algorithms for type 1 diabetes using survey data have been published. One early study used 1976–1980 National Health and Nutrition Examination Survey (NHANES) data to differentiate diabetes type among adults. Adult-onset insulin-dependent diabetes was defined by age of diabetes diagnosis ≥30 years, continuous or nearly continuous insulin use since diagnosis, and less than 125% 'desirable body weight' (ie, normal weight for height). Using this algorithm, the prevalence of type 1 diabetes in the general population was 0.30% among adults aged 30–74 years.[7] Another study using 1999–2010 NHANES data classified type 1 diabetes as age of diagnosis <30 years, beginning insulin within 1 year of diagnosis, and currently taking insulin; it found a 0.26% prevalence of type 1 diabetes among children and adults.[8] Several studies examining type 2 diabetes using NHIS or NHANES data exclude those diagnosed ≤30 years, who take insulin, and do not take oral hypoglycemic agents as likely having type 1 diabetes.[9–11]

Using age of diagnosis to classify type 1 diabetes is problematic as approximately 40% of people with type 1 diabetes are diagnosed between ages 31 and 60.[4] Recent analysis of the 2016 NHIS estimated the prevalence of type 1 diabetes in the general population of adults aged ≥18 to be 0.55%, defining type 1 diabetes as current insulin use and self-reported type 1 diabetes.[12] However, current insulin use is not exclusive to those with type 1 diabetes; many adults with type 2 diabetes take insulin.[13–15] Therefore, this algorithm may result in misclassification. Except for an initial period of generally several months after diagnosis, continuous use of insulin is essential in people with type 1 diabetes. Adults with type 1 diabetes are sometimes misdiagnosed during the initial period as having type 2 diabetes.[16] Additionally, respondents do not always interpret these survey questions as they are intended.[6 17] Also, previous research found that people with type 1 diabetes are more often normal weight and less likely to use oral hypoglycemics than those with type 2 diabetes.[4 13]

Given the various definitions used to classify type 1 diabetes using survey data and the lack of a gold standard, we examined the percentages of adults with diagnosed diabetes classified as having type 1 diabetes in the NHIS, using five primary and two supplementary algorithms, included for comparison purposes. We describe the percentage of type 1 diabetes and the sociodemographic and diabetes-specific characteristics of adults classified by each algorithm. Finally, we discuss the shortfalls and adequacy of each algorithm.

METHODS

The NHIS is a nationally representative, cross-sectional household survey of the civilian non-institutionalized US population that has been conducted since 1957 by the National Center for Health Statistics (NCHS). Interviews are conducted at the respondent's home; follow-ups to complete interviews may be conducted by telephone. All information is self-reported. All participants gave oral consent, and all personally identifiable information was removed before the data were released.[18] The risk of statistical disclosure was reduced by not releasing many geographic variables, and the data were top coded and/or responses collapsed when there was a risk of statistical disclosure. Participants were not paid for their participation in NHIS. Detailed information about the survey methods has been described elsewhere.[19]

Study respondents

The sample consisted of sample adult respondents (aged ≥18) from the 2016–2017 NHIS who answered yes to the question '[(If female), Other than during pregnancy] Have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?' (n=6331). Respondents with diagnosed diabetes were asked to report their age of diagnosis, use of oral hypoglycemic agents and/or insulin, and diabetes type (which was not verified by a doctor or other health professional). Respondents reporting current insulin use were asked how long after diagnosis they began using insulin, if they ever stopped taking insulin for more than 6 months, and if so, whether this was only during the first year of diagnosis. Women who reported ever being pregnant were asked if they had gestational diabetes during pregnancy. Demographic factors (age, sex, race/ethnicity) and height and weight (for calculating body mass index (BMI); normal weight BMI <25 kg/m^2; overweight BMI 25 to <30 kg/m^2; obese BMI ≥30 kg/m^2) were also used in this analysis. Self-reported height and weight have been shown to correlate well with direct measures; however, some studies have reported less accurate correlation among those with greater BMI.[20–22]

Table 1 Algorithm definitions for type 1 diabetes, NHIS 2016–2017

	Algorithm							
	1	2	3	4	5	S6	S7	Questionnaire item
Current insulin use	X	X	X	X	X		X	Insulin can be taken by shot or pump. Are you NOW taking insulin?
Continuous insulin use beginning within 1 year of diagnosis	X	X	X	X			X	Thinking back to when you were first diagnosed with diabetes, how long was it before you started taking insulin? Since you started taking insulin, have you ever stopped taking it for more than 6 months? Was this only during the first year after you were diagnosed with diabetes?
Diagnosed at age <30 years		X					X	How old were you when a doctor FIRST told you that you had diabetes or sugar diabetes?
Diagnosed at age <40 years			X					How old were you when a doctor FIRST told you that you had diabetes or sugar diabetes?
Self-reported type 1 diabetes				X	X	X		What type of diabetes do you have?
No use of oral hypoglycemic medication							X	Are you now taking diabetic pills to lower your blood sugar? These are sometimes called oral agents or oral hypoglycemic agents.

NHIS, National Health Interview Survey.

Algorithms for type 1 diabetes

We selected five primary and two supplementary algorithms for type 1 diabetes using various combinations of self-reported diabetes type, age of diagnosis, current and continuous insulin use, and use of oral hypoglycemics. Continuous insulin use was defined as reporting all three of the following: (1) current insulin use, (2) starting insulin within 1 year of diagnosis, and (3) not stopping insulin for more than 6 months except during the first year after diagnosis. The third criterion was added because of the honeymoon period, in which some individuals with type 1 diabetes go into apparent remission after starting insulin but later need insulin for the rest of their lives.[23 24]

Table 1 illustrates the criteria used in each survey-based algorithm. Algorithms were chosen based on their use in previous publications. Algorithm 1 includes continuous insulin use only. Algorithm 2 starts with algorithm 1 and additionally requires that diabetes was diagnosed at age <30. Algorithm 2 has been used in epidemiologic and some clinical research[8] because historically it had been thought that type 1 diabetes only occurs during youth. Algorithm 3 is a variant of algorithm 2 that increases the age of diagnosis to <40 years, which intends to include more people with type 1 diabetes but may also incorrectly capture more with type 2 diabetes.[8] Algorithms 4 and 5 use self-reported diabetes type and insulin use. Algorithm 4, using continuous insulin use, is more specific to type 1 diabetes.[9 10] Algorithm 5, using current insulin use, has been used in recent publications.[11] Algorithms 6 and 7 were included in online supplemental material for comparison purposes. Supplementary algorithm 6 includes self-reported type 1 diabetes as the only criterion. However, as insulin is an essential treatment for type 1 diabetes, excluding insulin as a criterion will miss persons with type 1 diabetes. Supplementary algorithm 7 defines type 1 diabetes as diagnosis at age <30, current insulin use, and not using oral hypoglycemics. The latter criterion has been used in analyses of NHIS and NHANES data[9–11] because oral hypoglycemics were historically only taken by persons with type 2 diabetes; it is included in online supplemental material because oral hypoglycemics are increasingly used with insulin to treat type 1 diabetes.

In presenting the percentages of type 1 diabetes among those with diagnosed diabetes, and the distribution of characteristics among those classified as having type 1 diabetes, we generally focus the Results section on comparisons between algorithms that are similar—except for changing one or two criteria—to more clearly show changes in percentages.

Statistical analysis

Percentages, means, and SEs were used to show: (1) the percentage of type 1 diabetes among those with diagnosed diabetes defined by each algorithm and by selected characteristics among US adults with diagnosed diabetes and (2) the distribution of population characteristics (demographic and diabetes-related factors) among US adults with type 1 diabetes defined by each algorithm. The latter was assessed to determine whether the distribution of certain characteristics aligned with expectations from previous research. Significance testing (two-sided t-tests, $p<0.05$) was performed for some highlighted comparisons of overall proportions in the text. Kappa statistics were determined to assess the magnitude of overlap between participants assigned to each algorithm. All statistical analyses used sample weights and accounted

for the complex survey design using STATA, SAS Version 9.4 and SUDAAN.[25 26] The NCHS Data Presentation Standards for Proportions were used to determine the reliability of proportions. Estimates were suppressed when the data were found to be unreliable.[27]

RESULTS

Percentage of type 1 diabetes among US adults with diagnosed diabetes according to algorithms

Algorithm 1: continuous insulin use

The overall percentage of type 1 diabetes among people with diabetes based on algorithm 1 was 10.2% and was two to three times higher for adults aged 18–44 versus older ages (p<0.001 for both older age groups) (table 2). It was higher for non-Hispanic white and black adults (vs Mexican Americans) (p<0.05 for both), for those who self-reported type 1 diabetes (vs type 2 diabetes (p<0.001)), for those with normal weight (vs overweight or obesity, p<0.001 for both), and for those not reporting oral hypoglycemic use (p<0.001).

Algorithm 2: continuous insulin use and diagnosis at age <30

With the addition of the age criterion of diabetes diagnosis to continuous insulin use, the overall percentage of type 1 diabetes using algorithm 2 decreased from 10.2% (algorithm 1) to 3.4% (p<0.001). When compared with algorithm 1, the percentage with type 1 diabetes was greater for young versus older ages, and greater in non-Hispanic white and black adults versus non-Hispanic Asian and Hispanic adults (p<0.05 for both).

Algorithm 3: continuous insulin use and diagnosis at age <40

When the age criterion was increased to <40 years, the overall percentage of type 1 diabetes increased to 4.9% compared with 3.4% with algorithm 2 (p<0.001) and was significantly less than the percentage of type 1 diabetes in algorithm 1 (10.3%, p<0.001).

Algorithm 4: self-reported type 1 diabetes and continuous insulin use

When self-reported type 1 diabetes was added to the criterion of continuous insulin use (algorithm 1), the overall percentage of type 1 diabetes was 3.7%. Compared with algorithm 1, the percentage of type 1 diabetes using algorithm 4 was greater for young versus older ages, and greater for non-Hispanic white adults versus other race/ethnic groups (p<0.001 for all).

Algorithm 5: self-reported type 1 diabetes and current insulin use

Using self-reported type 1 diabetes and current (algorithm 5) rather than continuous insulin use (algorithm 4), the overall percentage of type 1 diabetes increased to 5.6% (vs 3.7% using algorithm 4, p<0.001). Among those who self-reported type 1 diabetes, the percentage with type 1 diabetes increased to 53.8% using current insulin use as a criterion (algorithm 5), compared with using 35.2% using continuous insulin use (algorithm 4). As compared with algorithm 4, the percentage of type 1 diabetes at youngest versus older current ages decreased and the percentage of type 1 diabetes diagnosed at age ≥30 increased (p<0.001 for both).

Supplementary algorithm 6: self-reported type 1 diabetes

The overall percentage of type 1 diabetes using self-report alone (supplementary algorithm 6) was 10.4% (online supplemental table S1). The percentage diagnosed at age <20 increased for algorithm 6 compared with algorithm 5 (self-report and current insulin use) or algorithm 4 (self-report and continuous insulin use) (39.2% vs 33.1% and 26.9%, respectively). Those classified as having type 1 diabetes in supplementary algorithm 6 were more likely to have obesity than those in algorithm 4 or 5.

Supplementary algorithm 7: diagnosis at age <30, current insulin use, and not using oral hypoglycemics

The overall percentage of type 1 diabetes by supplementary algorithm 7 was 4.1%. The percentage of type 1 diabetes among those diagnosed at age <20 was higher in supplementary algorithm 7 than all other algorithms except supplementary algorithm 6. Otherwise, the percentages of type 1 diabetes by characteristics were most similar to those observed for algorithms 2 and 4.

Characteristics of adults with diagnosed diabetes classified as having type 1 diabetes

Algorithm 1: continuous insulin use

Among those classified as having type 1 diabetes by algorithm 1, mean age was 54.3 years, there were more men than women, and the majority were non-Hispanic white adults (61.9%) (table 3). Mean age of diagnosis was 30.9 years and 37.1% self-reported type 1 diabetes. Almost half had obesity (47.3%) and 43.3% reported taking oral hypoglycemics.

Algorithm 2: continuous insulin use and diagnosis at age <30

With the addition of the age criterion of diabetes diagnosis, the mean age (40.9 years) and mean age of diagnosis (15.6 years) decreased among those with type 1 diabetes using algorithm 2 compared with algorithm 1 (p<0.001 for both). The percentage of non-Hispanic white adults increased. The percentage who self-reported type 1 diabetes was twice as high as in algorithm 1 (74.7% vs 37.1%, p<0.001). A lower percentage had obesity and used oral hypoglycemics (p<0.001 for both).

Algorithm 3: continuous insulin use and diagnosis at age <40

When the age criterion was changed to <40, mean BMI was higher (30.1 vs 28.6 kg/m^2), as was the percentage of those using oral hypoglycemics (29.7% vs 16.1%) among those with type 1 diabetes using algorithm 3 as compared with algorithm 2 (p<0.001 for all).

Algorithm 4: self-reported type 1 diabetes and continuous insulin use

When self-reported type 1 diabetes was added to the criterion of continuous insulin use alone (algorithm 1), mean age was 10 years younger in algorithm 4 than algorithm 1

Table 2 Percentage (SE) of adults classified as having type 1 diabetes according to five algorithms among US adults with diagnosed diabetes, overall and among subgroups, USA 2016–2017

	Algorithm 1 (n=645)	Algorithm 2 (n=189)	Algorithm 3 (n=286)	Algorithm 4 (n=226)	Algorithm 5 (n=355)
	Continuous insulin use* beginning within 1 year of diabetes diagnosis	Diabetes diagnosed at age <30 years and continuous insulin use beginning within 1 year of diagnosis	Diabetes diagnosed at age <40 years and continuous insulin use beginning within 1 year of diagnosis	Self-report of type 1 diabetes and continuous insulin use beginning within 1 year of diagnosis	Self-report of type 1 diabetes and current insulin use
Weighted population size (millions)	2.38	0.77	1.13	0.85	1.30
	Percent (SE)				
Overall	10.2 (0.48)	3.4 (0.29)	4.9 (0.34)	3.7 (0.29)	5.6 (0.35)
Age (years)					
18–44	21.3 (1.93)	15.3 (1.62)	20.4 (1.83)	14.0 (1.57)	16.9 (1.83)
45–64	9.7 (0.76)	2.2 (0.35)	3.7 (0.45)	2.7 (0.38)	4.4 (0.47)
≥65	7.2 (0.59)	0.8 (0.22)	1.4 (0.26)	1.4 (0.24)	3.2 (0.37)
Sex					
Men	11.4 (0.71)	3.2 (0.41)	5.2 (0.54)	4.0 (0.43)	5.9 (0.49)
Women	8.9 (0.63)	3.5 (0.43)	4.7 (0.47)	3.3 (0.43)	5.2 (0.54)
Race/ethnicity					
Non-Hispanic white	10.4 (0.61)	4.1 (0.39)	5.4 (0.44)	4.6 (0.41)	7.0 (0.47)
Non-Hispanic black	12.9 (1.50)	3.1 (0.79)	5.6 (1.00)	1.8 (0.57)	3.3 (0.72)
Non-Hispanic Asian	†	0.1 (0.06)	0.6 (0.55)	1.6 (0.82)	†
Hispanic	8.7 (1.21)	2.2 (0.64)	4.1 (0.88)	2.3 (0.69)	3.5 (0.83)
Mexican American	6.4 (1.28)	1.2 (0.51)	2.6 (0.71)	1.3 (0.49)	1.6 (0.52)
Age of diagnosis (years)					
<20	32.4 (3.07)	32.4 (3.07)	32.4 (3.07)	26.9 (2.99)	33.1 (3.31)
20–29	17.1 (2.44)	17.1 (2.44)	17.1 (2.44)	9.9 (1.67)	13.1 (1.90)
30–39	11.7 (1.39)	‡	11.7 (1.39)	3.3 (0.68)	6.9 (9.8)
40–49	9.1 (1.00)	‡	‡	1.5 (0.36)	2.9 (0.51)
50–59	7.1 (0.80)	‡	‡	0.9 (0.28)	1.8 (0.36)
≥60	5.7 (0.73)	‡	‡	0.7 (0.24)	1.5 (0.40)
Self-reported type					
Type 1 diabetes	35.0 (2.21)	23.6 (2.1)	27.9 (2.22)	35.0 (2.21)	53.4 (2.56)
Type 2 diabetes	7.5 (0.48)	1.0 (0.18)	2.2 (0.27)	‡	‡
Body mass index (kg/m^2)					
<25.0	16.8 (1.57)	8.8 (1.32)	11.1 (1.45)	10.2 (1.35)	13.3 (1.54)
25.0–29.9	9.7 (0.94)	3.4 (0.53)	4.5 (0.60)	4.6 (0.64)	6.6 (0.73)
≥30.0	8.9 (0.63)	2.1 (0.31)	3.7 (0.43)	1.5 (0.24)	3.1 (0.35)
Gestational diabetes among women who have been pregnant					
No	8.0 (0.73)	2.5 (0.43)	3.3 (0.47)	2.6 (0.42)	4.1 (0.51)
Yes	7.1 (1.24)	4.0 (0.98)	5.7 (1.09)	1.5 (0.58)	3.8 (0.94
Continuous insulin use within 1 year of diagnosis*					
No	‡	‡	‡	‡	2.0 (0.21)
Yes	100	32.8 (2.27)	48.6 (2.46)	35.8 (2.38)	35.8 (2.38)
Oral hypoglycemic medication					
No	21.1 (1.27)	10.3 (0.92)	12.7 (1.05)	11.9 (1.00)	16.5 (1.13)
Yes	6.1 (0.48)	0.7 (0.18)	2.0 (0.28)	0.6 (0.15)	1.5 (0.21)

*Continuous insulin use implies current insulin use as well. Only those who report current insulin use are asked when they started insulin treatment.
†Data do not meet the National Center for Health Statistics Data Presentation Standards for Proportions.
‡Values are not applicable due to inclusion/exclusion criteria of the algorithm.

Table 3 Characteristics (means and percent distributions) among US adults classified as having type 1 diabetes according to five algorithms, USA 2016–2017

	Algorithm 1 (n=645)	Algorithm 2 (n=189)	Algorithm 3 (n=286)	Algorithm 4 (n=226)	Algorithm 5 (n=355)
	Continuous insulin use* beginning within 1 year of diabetes diagnosis	Diabetes diagnosed at age <30 years and continuous insulin use beginning within 1 year of diagnosis	Diabetes diagnosed at age <40 years and continuous insulin use beginning within 1 year of diagnosis	Self-report of type 1 diabetes and continuous insulin use beginning within 1 year of diagnosis	Self-report of type 1 diabetes and current insulin use
Age					
Mean	54.3 (0.85)	40.9 (1.48)	44.0 (1.17)	44.4 (1.49)	49.2 (1.31)
18–44 years	27.9 (2.41)	60.9 (4.45)	55.0 (3.61)	51.0 (4.24)	40.3 (3.62)
45–64 years	42.4 (2.64)	28.8 (3.98)	33.3 (3.41)	33.5 (3.83)	35.7 (3.13)
≥65 years	29.7 (2.19)	10.3 (2.68)	11.7 (2.16)	15.5 (2.63)	24.0 (2.63)
Sex					
Men	57.6 (2.34)	48.7 (4.59)	54.2 (3.79)	55.9 (4.32)	54.8 (3.41)
Women	42.4 (2.34)	51.3 (4.59)	45.8 (3.79)	44.1 (4.32)	45.2 (3.41)
Race/ethnicity					
Non-Hispanic white	61.9 (2.75)	74.0 (3.92)	66.8 (3.40)	77.1 (3.76)	76.0 (3.00)
Non-Hispanic black	19.2 (2.75)	14.2 (3.29)	17.4 (2.80)	7.6 (2.33)	8.9 (1.89)
Non-Hispanic Asian	2.4 (0.89)	0.1 (0.10)	0.6 (0.55)	2.2 (1.11)	2.3 (1.02)
Hispanic	14.4 (1.99)	11.0 (3.02)	14.0 (2.82)	10.9 (2.88)	10.7 (2.29)
Mexican American	†	†	†	†	†
Age of diagnosis					
Mean (years)	30.9 (0.38)	15.6 (0.81)	21.6 (0.88)	23.9 (1.37)	27.7 (1.22)
<20 years	21.0 (2.17)	63.0 (4.69)	42.8 (3.85)	48.2 (4.37)	39.3 (3.39)
20–29 years	12.3 (1.75)	37.1 (4.69)	25.2 (3.38)	19.8 (3.26)	17.3 (2.48)
30–39 years	15.7 (1.83)	‡	32.1 (3.30)	12.4 (2.41)	17.0 (2.21)
40–49 years	19.8 (1.99)	‡	‡	8.8 (2.08)	11.5 (1.97)
50–59 years	18.7 (1.94)	‡	‡	6.5 (2.01)	8.7 (1.70)
≥60 years	12.5 (1.58)	‡	‡	4.4 (1.44)	6.2 (1.63)
Self-reported type					
Type 1 diabetes	37.1 (2.44)	74.7 (3.80)	61.3 (3.74)	100	100
Type 2 diabetes	62.9 (2.44)	25.3 (3.80)	38.7 (3.74)	‡	‡
Body mass index (kg/m^2)					
Mean	30.9 (0.38)	28.6 (0.57)	30.1 (0.60)	27.4 (0.48)	28.5 (0.44)
<25.0	23.6 (2.13)	37.0 (4.33)	32.1 (3.59)	39.6 (4.10)	34.2 (3.24)
25.0–29.9	29.1 (2.47)	30.1 (4.24)	27.8 (3.36)	38.3 (4.27)	36.2 (3.48)
≥30.0	47.3 (2.59)	32.9 (4.08)	40.1 (3.68)	22.1 (3.17)	29.7 (3.07)
Gestational diabetes among women who have been pregnant					
No	80.5 (3.06)	69.5 (6.46)	67.7 (5.13)	†	79.9 (4.49)
Yes	19.5 (3.06)	30.5 (6.46)	32.3 (5.13)	†	20.1 (4.49)
Continuous insulin use within 1 year of diagnosis*					
No	‡	‡	‡	‡	32.5 (2.85)
Yes	100	100	100	100	67.5 (2.85)
Oral hypoglycemic medication					
No	56.7 (2.77)	83.9 (3.60)	70.3 (3.65)	88.9 (2.92)	80.9 (2.69)
Yes	43.3 (2.77)	16.1 (3.60)	29.7 (3.65)	11.1 (2.92)	19.0 (2.69)
Only oral hypoglycemic medication					
No	100	100	100	100	100
Yes	‡	‡	‡	‡	‡

*Continuous insulin use implies current insulin use as well. Only those who report current insulin use are asked when they started insulin treatment.
†Data do not meet the National Center for Health Statistics Data Presentation Standards for Proportions.
‡Values are not applicable due to inclusion/exclusion criteria of the algorithm.

(p<0.001). Among those with type 1 diabetes using algorithm 4 compared with algorithm 1, a larger percentage were non-Hispanic white adults (77.1% vs 61.9%) and a smaller percentage were non-Hispanic black adults (7.6% vs 19.2%); mean age of diabetes diagnosis was younger (23.9 vs 30.9 years); BMI was lower (27.4 vs 30.9 kg/m^2); and a smaller percentage used oral hypoglycemics (11.1% vs 43.3%) (p<0.001 for all).

Algorithm 5: self-reported type 1 diabetes and current insulin use

Among those with type 1 diabetes using algorithm 5, mean age was older relative to that for continuous insulin use and self-reported type 1 diabetes (algorithm 4) (49.2 vs 44.4 years). Mean age of diagnosis (27.7 vs 23.9 years), BMI (28.5 vs 27.4 kg/m^2), and use of oral hypoglycemics (19.0% vs 11.1%) were higher for algorithm 5 than algorithm 4 (p<0.001 for all).

Supplementary algorithm 6: self-reported type 1 diabetes

Among those defined by self-reported type 1 diabetes alone (supplementary algorithm 6), mean age was 53.7 years compared with 44.4 years for algorithm 4 (self-reported type 1 diabetes and continuous insulin use), and a smaller percentage were non-Hispanic white adults (59.8% vs 77.1%) and a larger percentage were non-Hispanic black adults (14.3% vs 7.6%) (p<0.001 for all) (online supplemental table S2). Mean age of diagnosis was 50% higher (36.0 vs 23.9 years) and mean BMI was also higher (29.6 vs 27.4 kg/m^2) compared with algorithm 4. Use of oral hypoglycemics (46.7% vs 11.1%) was four times higher for those defined by supplementary algorithm 6 versus algorithm 4 (p<0.001). Only 35.5% reported continuous insulin use.

Supplementary algorithm 7: diabetes diagnosed at age <30 years, current insulin use, and not using oral hypoglycemics

Among those with type 1 diabetes defined by supplementary algorithm 7, the distributions of age and race/ethnicity were similar to algorithms 2 and 4. The percentage of women who had gestational diabetes among women who had been pregnant was highest among those using supplementary algorithm 7. Just over two-thirds (70.6%) reported continuous insulin use.

Kappa statistics

Kappa statistics ranged in value from 0.28 for algorithms 1 and 6 to 0.79 for algorithms 2 and 3. Results for all algorithm comparisons can be found in online supplemental table S3.

DISCUSSION

The results from this analysis demonstrate the variability in the percentage of adults classified as having type 1 diabetes, as well as the distribution of sociodemographic and diabetes-specific characteristics by algorithms used in survey research. The lack of biomarkers or medical records to confirm diabetes type limits our ability to compare the algorithms to a gold standard. The information presented here illustrates the pros and cons of each algorithm allowing researchers to select an algorithm suited to their research.

Using algorithm 1, continuous insulin use alone, two-thirds of people with type 1 diabetes were diagnosed at age ≥30 and 54% diagnosed at ages 30–59. This may include some people with type 2 diabetes who are generally diagnosed later in life. However, previous work suggests that 40% of people with type 1 diabetes are diagnosed between ages 31 and 60.[4]

Over time, type 1 diabetes has been diagnosed later in life and type 2 diabetes at younger ages,[4 28] indicating that age of diagnosis as a criterion (algorithms 2 and 3) may not be an essential part of algorithms.[4] Nonetheless, our analyses showed that larger percentages of individuals classified as having type 1 diabetes using algorithms with age of diagnosis cut-points and continuous insulin use self-reported type 1 diabetes, were of normal weight, and were not using oral hypoglycemics as compared with those classified by continuous insulin use alone. The age restriction likely removes some people with type 2 diabetes, while excluding later onset type 1 diabetes. For researchers studying younger adults, the use of age cut-points and continuous insulin use may be appropriate.

It is important to highlight that this national study defined continuous insulin use as using insulin within 1 year of diagnosis and not stopping thereafter (with the exception of stopping for >6 months within the first year of diagnosis) as compared with a historical definition of continuous insulin use from the time of diagnosis. Since adults with type 1 diabetes may experience a honeymoon period after diagnosis and initial treatment of near normal glucose levels without taking insulin,[23 24] the historical definition of continuous insulin use would have excluded some adults with type 1 diabetes.

Among persons who self-reported type 1 diabetes alone (supplementary algorithm 6), about one-third reported taking oral hypoglycemics and no insulin and approximately one-third reported continuous insulin use, the hallmark of type 1 diabetes. This demonstrates that self-report alone is an unreliable method to classify diabetes type. However, adding the criterion of continuous insulin use to self-report of type 1 may reduce misclassification. Among those classified using algorithm 5 (self-reported type 1 diabetes and current insulin use), only two-thirds reported continuous insulin use indicating this algorithm may include a high percentage of individuals with type 2 diabetes.

More recently, oral hypoglycemics are used along with insulin to improve control in type 1 diabetes, but they are used much less frequently than insulin alone.[13 14] Using supplementary algorithm 7 (diagnosis at age <30, current insulin use, and not using oral hypoglycemics),[9–11] nearly a third of adults did not report continuous insulin use suggesting that it includes many people with type 2 diabetes. For algorithms 2 and 4, the percentage using oral hypoglycemics was 16.1% and 11.1%, respectively. Data from the Type 1 Diabetes Exchange Network

found that 6% age ≥26 were taking metformin.[13] Use of non-insulin hypoglycemic medication was 3.1% in the Epidemiology of Diabetes Interventions and Complications study and 2.4% in the Pittsburgh Epidemiology of Diabetes Complications Study (C Cowie and T Orchard, personal communication). Thus, using oral hypoglycemics as an exclusion criterion for type 1 diabetes may result in misclassification.

The American Diabetes Association (ADA) estimates that the percentage of type 1 diabetes among those with diabetes ranges from 5% to 10%.[1] The percentages estimated in this analysis generally fell close to the ADA estimates, ranging from 3.3% to 10.5%. When comparing algorithm 2 (diagnosed <30 and continuous insulin use) and algorithm 4 (self-reported type 1 diabetes and continuous insulin use), the percentage with type 1 diabetes was similar overall, and similar for normal weight, and for those taking oral hypoglycemics. Since smaller, non-representative type 1 diabetes populations vary depending on age, duration of type 1 diabetes, and glucose control there is no benchmark for BMI and oral hypoglycemic use in type 1 diabetes; however, these attributes align with the expectation that those with type 1 diabetes have lower levels of obesity and lower frequency of using oral hypoglycemics compared with those with type 2 diabetes.[4 29] While self-reported type 1 diabetes and current insulin use (algorithm 5) have been used previously[9] and percentages for mean BMI and oral hypoglycemic use were similar to algorithms 2 and 4, information on continuous rather than current insulin use is preferred since persons with type 1 diabetes require insulin for the duration of their disease.

While survey research presents unique challenges, defining type 1 diabetes can be challenging even with rigorous laboratory testing. A study from the SEARCH for Diabetes in Youth study found that most youth fell into the traditional categories for type 1 and type 2 diabetes (70.4%), but the etiology of the remaining participants was unknown or would warrant further testing.[28]

A strength of this study was the use of a nationally representative sample, which allowed for generalization to the US civilian adult non-institutionalized population. In addition, there were numerous questionnaire items related to diabetes, which permitted us to analyze several different algorithms for type 1 diabetes. Limitations of this study were that all data were self-reported and type of diabetes could not be validated by medical records or laboratory measures. However, medical records may be limited in information on type of diabetes, and laboratory measures (eg, autoantibodies, C-peptide) fluctuate and may not capture occurrence during the natural history of the disease.[30] Cognitive testing showed that respondents often had limited knowledge with respect to type of diabetes and type of glycemic medication treatment.[6 17] For example, some respondents based their response to their diabetes type on sometimes inaccurate ideas they had about the different types, such as type 1 only occurring at younger ages, or that diabetes progresses in stages in a similar way to cancer (with type 2 being a more severe type of diabetes than type 1). Respondents diagnosed at young ages or those very recently diagnosed could give an exact age of diagnosis, while most of those diagnosed as adults estimated their age of diagnosis. Generally, respondents knew if they were taking diabetes medication, but some thought pills were a type of insulin. As a result, the prelude 'Insulin can be taken by shot or pump' was added to the question about insulin use. As shown throughout our results, self-reported type of diabetes alone should not be used to define type of diabetes.

The findings from this study may inform future diabetes research that uses survey data. A wide range of prevalence estimates and characteristic distributions were found according to algorithms used to define type 1 diabetes as well as potential misclassification. Using the results of this study as a foundation, future research may link survey questions to medical records, providing additional information on which algorithms perform best for classifying type 1 diabetes using survey data.

Author affiliations
[1]Social & Scientific Systems, Public Health Research, Silver Spring, Maryland, USA
[2]Division of Health Interview Statistics, National Center for Health Statistics, Hyattsville, Maryland, USA
[3]Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[4]Division of Diabetes Translation, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[5]Westat, Rockville, Maryland, USA
[6]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

Contributors SSC contributed to the research design, analyzed the data, and wrote, reviewed and edited the manuscript. SEL contributed to the research design and analyzed the data. TJO, KMB, LSG, SHS, AM and GI contributed to the discussion. KFR provided statistical support. CCC contributed to the research design. SEL, TJO, KMB, LSG, SHS, AM, GI, KFR and CCC reviewed and edited the manuscript. SSC is the guarantor and takes full responsibility for the contents of the article.

Funding This study was funded by the National Institute of Diabetes and Digestive and Kidney Diseases. SSC and KFR were supported by this NIDDK contract (GS-10F-0381L).

Disclaimer The findings and conclusions are those of the authors and do not necessarily represent the official position of the National Institute of Diabetes and Digestive and Kidney Diseases, the National Center for Health Statistics, or the Centers for Disease Control and Prevention Diabetes-Division of Diabetes Translation.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Centers for Disease Control's IRB approves the NHIS annual survey.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are publicly available at https://www.cdc.gov/nchs/nhis/nhis_questionnaires.htm

ORCID iDs
Sarah S Casagrande http://orcid.org/0000-0002-6293-383X
Andy Menke http://orcid.org/0000-0003-0729-7161

REFERENCES

1 American Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes-2019*. *Diabetes Care* 2019;42:S13–28.
2 Ng E, Vanderloo SE, Geiss L, *et al*. Concordance between self-report and a survey-based algorithm for classification of type 1 and type 2 diabetes using the 2011 population-based Survey on Living with Chronic Diseases in Canada (SLCDC)-Diabetes component. *Can J Diabetes* 2013;37:249–53.
3 Kruse RL, Olsberg JE, Oliver DP, *et al*. Patient-Provider communication about diabetes self-care activities. *Fam Med* 2013;45:319–22.
4 Thomas NJ, Jones SE, Weedon MN, *et al*. 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;6:122–9.
5 Reinehr T. Type 2 diabetes mellitus in children and adolescents. *World J Diabetes* 2013;4:270–81.
6 Lessem S. *Distinguishing diabetes type*. Madison, WI, 2015.
7 Harris MI, Robbins DC. Prevalence of adult-onset IDDM in the U.S. population. *Diabetes Care* 1994;17:1337–40.
8 Menke A, Orchard TJ, Imperatore G, *et al*. The prevalence of type 1 diabetes in the United States. *Epidemiology* 2013;24:773–4.
9 Cheung BMY, Ong KL, Cherny SS, *et al*. Diabetes prevalence and therapeutic target achievement in the United States, 1999 to 2006. *Am J Med* 2009;122:443–53.
10 Dodd AH, Colby MS, Boye KS, *et al*. Treatment approach and HbA1c control among US adults with type 2 diabetes: NHANES 1999-2004. *Curr Med Res Opin* 2009;25:1605–13.
11 Saydah SH, Imperatore G, Beckles GL. Socioeconomic status and mortality: contribution of health care access and psychological distress among U.S. adults with diagnosed diabetes. *Diabetes Care* 2013;36:49–55.
12 Bullard KM, Cowie CC, Lessem SE, *et al*. Prevalence of Diagnosed Diabetes in Adults by Diabetes Type - United States, 2016. *MMWR Morb Mortal Wkly Rep* 2018;67:359–61.
13 Miller KM, Foster NC, Beck RW, *et al*. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D exchange clinic registry. *Diabetes Care* 2015;38:971–8.
14 Saydah SH. Medication Use and Self-care Practices in Persons with Diabetes. In: Cowie CC, Menke A, Cissell MA, *et al*, eds. *Diabetes in America*. 3rd ed. Bethesda, MD: National Institutes of Health, 2014.
15 Ahlqvist E, Storm P, Käräjämäki A, *et al*. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6:361–9.
16 Leslie RD, Pozzilli P. Type I diabetes masquerading as type II diabetes. Possible implications for prevention and treatment. *Diabetes Care* 1994;17:1214–9.
17 Dunston SK, Willson S, Lessem S, *et al*. Cognitive Testing of the 2016 NHIS Diabetes Primary Prevention Questions. In: *Center for questionnaire design and evaluation research*. Hyattsville, MD: National Center for Health Statistics, 2017.
18 Description NS. *Hyattsville, Maryland: centers for disease control and prevention, us. Department of health and human services*, 2017.
19 Centers for Disease Control and Prevention. NHIS Survey Description. In: *U.S. department of health and human services*. Hyattsville, Maryland, 2017.
20 Davies A, Wellard-Cole L, Rangan A, *et al*. Validity of self-reported weight and height for BMI classification: a cross-sectional study among young adults. *Nutrition* 2020;71:110622.
21 Hodge JM, Shah R, McCullough ML, *et al*. Validation of self-reported height and weight in a large, nationwide cohort of U.S. adults. *PLoS One* 2020;15:e0231229.
22 Lin CJ, DeRoo LA, Jacobs SR, *et al*. Accuracy and reliability of self-reported weight and height in the sister study. *Public Health Nutr* 2012;15:989–99.
23 Juvenile Diabetes Research Foundation. *Honeymoon phase*. New York, 2019.
24 Sokołowska M, Chobot A, Jarosz-Chobot P. The honeymoon phase - what we know today about the factors that can modulate the remission period in type 1 diabetes. *Pediatr Endocrinol Diabetes Metab* 2016;22:66–70.
25 International RTI. *SUDAAN® statistical software for analyzing correlated data 11.0.1*. RTI International, 2017.
26 SAS Institute Inc. *Base SAS® 9.4 procedures guide. in*. NC: Cary, 2017.
27 Parker JD, Talih M, Malec DJ, *et al*. National center for health statistics data presentation standards for proportions. *Vital Health Stat* 2017;2:1–22.
28 SEARCH Study Group. Search for diabetes in youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. *Control Clin Trials* 2004;25:458–71.
29 Beaser RS. *Joslin's Diabetes Deskbook: A guide for primary care providers*. Boston, MA: Joslin Diabetes Center, 2014.
30 Ginsberg BH. Factors affecting blood glucose monitoring: sources of errors in measurement. *J Diabetes Sci Technol* 2009;3:903–13.