



Published in final edited form as:

Ann Intern Med. 2009 December 1; 151(11): 775–783. doi:
10.1059/0003-4819-151-11-200912010-00005.

A patient self-assessment diabetes screening score: development, validation, and comparison to other diabetes risk assessment scores

Heejung Bang, PhD¹, Alison M. Edwards, MStat¹, Andrew S. Bomback, MD, MPH², Christie M. Ballantyne, MD³, David Brillon, MD⁴, Mark A. Callahan, MD⁵, Steven M. Teutsch, MD, MPH⁶, Alvin I. Mushlin, MD, ScM^{4,7}, and Lisa M. Kern, MD, MPH^{4,7}

¹Division of Biostatistics and Epidemiology, Department of Public Health, Weill Medical College of Cornell University, New York, NY

² Division of Nephrology, Department of Medicine, Columbia University College of Physicians and Surgeons, New York, NY

³Section of Atherosclerosis and Lipoprotein Research, Department of Medicine, Baylor College of Medicine, and Center for Cardiovascular Disease Prevention, Methodist DeBakey Heart Center, Houston, TX

⁴Department of Medicine, Weill Medical College of Cornell University, New York, NY

⁵FOJP Service Corporation, New York, NY

⁶Los Angeles County Department of Public Health, Los Angeles, CA

⁷Department of Public Health, Weill Medical College of Cornell University, New York, NY

Abstract

Background—National guidelines disagree on who should be screened for undiagnosed diabetes. No existing diabetes risk score is highly generalizable or widely followed.

Corresponding author: Heejung Bang PhD, 402 East 67th Street, New York, NY 10065. heb2013@med.cornell.edu Phone: (646) 692-8015 Fax: (646) 692-0281.

Requests for Reprints: Heejung Bang PhD, 402 East 67th Street, New York, NY 10065; heb2013@med.cornell.edu

Current Author Addresses: Drs. Bang, Mushlin and Kern and Ms. Edwards: Department of Public Health, Weill Cornell Medical College, 402 East 67th Street, New York, NY 10065.

Dr. Bomback

Division of Nephrology, Department of Medicine, Columbia University Medical Center, 622 West 168th Street, PH 4-124, New York, NY 10032.

Dr. Ballantyne:

Section of Atherosclerosis and Lipoprotein Research, Department of Medicine, Baylor College of Medicine, 6565 Fannin, M.S. A-60, Houston, TX 77030.

Dr. Brillon:

Department of Medicine, Weill Cornell Medical College, Baker Pavilion, 20th Floor, 525 East 68th Street, New York, NY 10065.

Dr. Callahan:

FOJP Service Corporation, 28 East 28th Street, New York, NY 10016.

Dr. Teutsch:

Los Angeles County Public Health, 313 N Figueroa St, Rm 708, Los Angeles, CA 90012.

Potential Conflicts of Interest: Dr. Teutsch is a former employee and an option holder in Merck and Co. Inc.

Disclaimer: The NHANES were supported by the Centers for Disease Control and Prevention. The ARIC and CHS studies were supported by the National Heart, Lung and Blood Institute. This manuscript was prepared using public use datasets (for NHANES) and limited access datasets (for ARIC and CHS) and does not necessarily reflect the opinions or views of these studies or agencies.

Reproducible Research Statement: Study protocol: Not available. Statistical code (secondary analysis): Available from Dr. Bang (heb2013@med.cornell.edu) upon request. Data set: NHANES are publicly available (at <http://www.cdc.gov/nchs/nhanes.htm>) and ARIC and CHS are available through a limited-access distribution agreement (http://www.nhlbi.nih.gov/resources/deca/datasets_obv.htm).

Objectives—To develop a new diabetes screening score and compare it to other available screening instruments (Centers for Disease Control and Prevention, American Diabetes Association (ADA) and U.S. Preventive Services Task Force guidelines; two ADA risk questionnaires; and Rotterdam model)

Design—Cross-sectional data.

Setting—National Health and Nutrition Examination Survey (NHANES) 1999–2004 for model development, and NHANES 2005–2006 plus a combined cohort of two community studies, Atherosclerosis Risk in Communities (ARIC) and Cardiovascular Health Study (CHS), for validation.

Participants—U.S. adults 20 years old.

Measurements—A risk scoring algorithm for undiagnosed diabetes, defined as fasting plasma glucose ≥ 7.0 mmol/L (126 mg/dL) without known diabetes, was developed in the development dataset. Logistic regression was used to determine participant characteristics that were independently associated with undiagnosed diabetes. The new algorithm and other methods were evaluated by standard diagnostic and feasibility measures.

Results—Age, sex, family history of diabetes, history of hypertension, obesity, and physical activity were associated with undiagnosed diabetes. In NHANES (in ARIC/CHS), the cutpoint of 5 selected 30(40)% of persons for diabetes screening and yielded sensitivity of 79(72)%, specificity of 67(62)%, positive predictive value of 10(10)% and likelihood ratio-positive of 2.39(1.89). In contrast, the comparison scores yielded sensitivity of 44–100%, specificity of 10–73%, positive predictive value of 5–8%, and likelihood ratio-positive of 1.11–1.98.

Limitations—Data during pregnancy were not available.

Conclusions—This new diabetes screening score, simple and easily implemented, seems to demonstrate improvements upon the existing methods. Future studies are needed to evaluate it in diverse populations in real world settings.

Primary Funding Source—Clinical and Translational Science Center at Cornell Medical College.

INTRODUCTION

Diabetes and its complications are major causes of morbidity and mortality worldwide. Over 60 million U.S. adults are estimated to have either diagnosed diabetes, undiagnosed diabetes, or pre-diabetes, with approximately 30% of diabetes cases estimated to be undiagnosed. With the steadily rising prevalence of diabetes, prevention of diabetes has become a major health priority (1–5). Recent clinical trials demonstrate that lifestyle(6–8) and pharmaceutical(6, 9, 10) interventions in individuals with impaired glucose tolerance can prevent or delay the development of diabetes, providing a rationale for the identification of high-risk subjects who may benefit from early lifestyle interventions.

National guidelines for diabetes screening are available to help detect undiagnosed disease, and various risk assessment tools for prevalent or incident diabetes have been developed to identify individuals most in need of screening. Yet many of these risk assessment tools were developed from specific cohorts, often with restrictive age range or race/ethnic groups, limiting generalizability to the entire population. There are three national guidelines for diabetes screening in the U.S.: the Centers for Disease Control and Prevention (CDC)(11), the American Diabetes Association (ADA)(12), and the U.S. Preventive Services Task Force (USPSTF)(13). Additionally, two risk scoring algorithms for undiagnosed diabetes have been derived from nationally representative samples: Herman et al.'s model from the

National Health and Nutrition Examination Survey (NHANES) II (conducted in 1976–1980) (14), and Heikes et al.'s model from the NHANES III (conducted in 1988–1994)(4). These two algorithms are also known as the ADA diabetes questionnaires.

In this paper, we have developed a new screening score for undiagnosed diabetes in multi-ethnic U.S. adults using readily available health information. Our aim was to improve existing diabetic risk scoring algorithms by using a more contemporary NHANES (conducted in 1999–2006) and formulating an easy, systematic scoring system that enables lay persons to assess their own risk of undiagnosed diabetes.

METHODS

Study Design and Participants

The NHANES is a cross-sectional study conducted by the National Center for Health Statistics in the CDC. In order to represent the US population, NHANES utilized complex, multistage probability sampling of the civilian, non-institutionalized population. To produce reliable statistics, NHANES over-sampled elderly persons and some racial/ethnic minorities. We utilized de-identified data from multiple waves of NHANES (i.e., 1999–2006) that are publicly available.

We included participants who were aged ≥ 20 years and had fasting plasma glucose (FPG) results. We excluded pregnant women. We used NHANES 1999–2004 for prediction modeling and screening score development, and NHANES 2005–6 for validation. Additionally, we conducted further validation combining the baseline data from two biracial cohort studies: the Atherosclerosis Risk in Communities (ARIC) study and the Cardiovascular Health Study (CHS). Detailed descriptions of these two studies have been published previously(15, 16). Briefly, ARIC enrolled 15,732 participants aged 45 to 64 years between 1987 and 1989 from four communities, and CHS recruited 5,201 participants 65 years and older between 1989 and 1990 from four communities. Between 1992 and 1993, CHS enrolled an additional 687 blacks to increase minority participation.

Data

For each participant, we retrieved data that were collected through interviews, physical examinations, and laboratory tests. Specifically, we utilized data regarding participants' demographic and socioeconomic characteristics, health care utilization, personal and family medical histories, health habits, physical examinations including anthropometric findings, and laboratory tests. For the obesity measure, we combined body mass index and waist circumference. For variable categorization, we used conventional cutoffs or well accepted clinical guidelines, if available. If information was missing or unknown in categorical variables, we defined the condition as absent, a convention commonly adopted in the risk questionnaire setting. We planned to instruct users, "Enter your score (But if you don't know the answer, enter 0)," in our questionnaire.

We stratified the participants into four groups by diabetes status: known diabetes, normal glucose metabolism (fasting plasma glucose (FPG) < 5.5 mmol/L (100 mg/dL)), impaired fasting glucose or pre-diabetes (FPG 5.6–6.9 mmol/L (100–125 mg/dL)), and undiagnosed diabetes (FPG ≥ 7.0 mmol/L (126 mg/dL))(4, 17, 18). Specifically, we defined known diabetes as participants who answered "yes" to the question "Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?" or reported using insulin or other diabetic medications.

For medical history variables, we considered data from multiple sources. For example, we classified participants as having hypertension if they reported a history of hypertension,

reported using prescribed medication for hypertension, had a systolic blood pressure ≥ 140 mmHg, or had a diastolic blood pressure ≥ 90 mmHg. We defined hyperlipidemia if total cholesterol ≥ 5.17 mmol/L (200 mg/dL) or triglycerides ≥ 1.69 mmol/L (150 mg/dL)(19), and high cholesterol if a person reported a history of high cholesterol, used cholesterol-lowering medication, or had a fasting low-density lipoprotein cholesterol ≥ 2.59 mmol/L (100 mg/dL) with a history of cardiovascular disease(20). Cardiovascular disease was defined as myocardial infarction or stroke.

When definitions of variables were not identical across the different studies (e.g., physical activity), we tried to use the best available variables to achieve reasonable consistency across databases. For example, in NHANES, we defined ‘physically active’ if a person answered “more active” to the question, “Compare your activity with others of the same age”. Otherwise, we classified subjects as ‘not physically active’. In ARIC, physical activity was assessed in a yes vs. no question, while in CHS, we dichotomized the physical activity question into “no” or “low” vs. “moderate” or “high”. None of the databases we used collected data during pregnancy. Finally, some ARIC participants (N=521) did not fast. For those, we used a random blood glucose of 11.1 mmol/L (200 mg/dL) cutpoint to define diabetes(18, 21, 22). Non-fasting participants were not included in the model development (using NHANES) but they were included in the external validation to reflect a realistic scenario.

Statistical Analysis

We used descriptive statistics to characterize the 4 groups according to diabetes status: mean and standard error were used for continuous variables and percentage was used for categorical variables. For model fitting, multiple logistic regression was adopted with undiagnosed diabetes cases as the endpoint, excluding diagnosed diabetes cases. Due to small proportions of missing data, we used all non-missing observations available in the relevant analyses. The only exception is that we imputed ‘family history of diabetes’ using a statistical technique (missing data imputation procedure, Proc MI, in SAS) for handling missing data in CHS, one of the external validation datasets, as this information was not collected in CHS(23, 24).

Development of a new screening score—Using the development dataset (NHANES 1999–2004), we included a comprehensive list of predictors known to be potentially associated with undiagnosed diabetes in an initial model. Specifically, the main effects of all variables listed in Table 1 and their interaction effects with age were included. Due to a large number of covariates, we started with continuous variables and later categorized them in the final model. We employed backward elimination (deleting the covariate with the largest p-value, one at a time) from the initial model until we reached a final model with statistically significant covariates. We were guided by statistical significance for the model building but also used scientific and qualitative judgments as well. For example, although income and health insurance status were statistically significant, we decided not to keep these variables in the final model as they were deemed less appropriate or less user-friendly in risk assessment questionnaires. Physical activity showed borderline significance, p-value=0.06, in the development dataset but was kept as this is an underlying protective factor that often fails to reach statistical significance for various reasons (e.g., difficult to quantify, misclassification, or insufficient statistical power). Moreover, physical activity is highly modifiable, in contrast to demographic and health history variables which are not(25).

In the final model, we double-checked that any important covariates were not erroneously omitted in this sequential process. We intentionally used only categorized variables that

captured easy but relevant and validated health information in the prediction model, aiming to develop a user-friendly and educational screening score. We created a weighted scoring system by rounding up all regression coefficients in the final model to the nearest integer (when strong monotonicity was observed, we broke the tie accordingly).

Validation and comparison to other methods—We evaluated our scoring system in NHANES 2005–6. We computed standard validation measures: the proportion of high risk individuals, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Youden index ($=1-\text{false positive rate}-\text{false negative rate}=\text{sensitivity}+\text{specificity}-1$), likelihood ratios for a positive test ($=\text{sensitivity}/(1-\text{specificity})$) and for a negative test ($=(1-\text{sensitivity})/\text{specificity}$), and the area under the receiver-operating-characteristic curve (AUC) as a discrimination statistic(26–28). In addition, our final model was re-fitted to ARIC/CHS. We estimated the prevalence of undiagnosed diabetes per individual score in NHANES and ARIC/CHS.

In validation samples using the aforementioned evaluation measures, we compared our new classification rule with the national screening guidelines and other assessment algorithms for undiagnosed diabetes: CDC(11), ADA(22), USPSTF(13), two ADA diabetes risk questionnaires(4, 14), and the Rotterdam model that was derived from a European sample(29). The last model was included to evaluate the generalizability and transferability of a validated non-U.S. model to the U.S. population.

Ancillary analyses—We performed three ancillary analyses for checking sensitivity/robustness and testing the utility of the new screening score in broadened practical contexts. We repeated the NHANES analyses 1) using ‘undiagnosed diabetes and pre-diabetes’ as an expanded endpoint (with ‘normal glucose’ as the reference group); 2) separately for those <45 years old vs. those ≥45 years old, as 45 years is the age threshold proposed by the ADA for universal screening; and 3) using an alternative definition of the endpoint based on Hemoglobin A1c ≥6.5%(12, 30). The first analysis may have particular importance for predicting pre-diabetes, the condition for which prevention has been shown to be more beneficial compared to undiagnosed yet manifest disease.

For statistical analyses, SAS version 9.1 software (Cary, NC) was used. For NHANES analyses, survey procedures with options of strata, cluster and weight were used to account for the complex survey design(24, 26). Two-sided hypotheses and tests were adopted for all statistical inferences.

Role of the Funding Source

The Clinical and Translational Science Center at Weill Cornell Medical College (UL1-RR024996) provided partial support for data analyses. This funding source did not have a role in the design of our analysis, its interpretation, or the decision to submit the manuscript for publication.

RESULTS

Our sample was comprised of 5,258 participants in the development dataset. Characteristics of participants according to diabetes status are summarized in Table 1. The crude weighted prevalence of undiagnosed diabetes based on fasting glucose in adults aged 20 years or older was 2.8%(2). Diabetic (diagnosed or undiagnosed) participants tended to be older and have lower education and household income than their non-diabetic counterparts. Diabetics also tended to be hypertensive, do less exercise, and have family history of diabetes and personal history of cardiovascular disease. Interestingly, participants with undiagnosed diabetes were more likely to have higher blood pressure, body mass index, waist circumference, total

cholesterol, and dyslipidemia than participants with diagnosed diabetes (differences were not formally tested).

Table 2 presents the final regression model derived from the development dataset. Age, sex, family history of diabetes, personal history of hypertension, obesity, and physical activity were statistically significant predictors of undiagnosed diabetes. Age and obesity status needed multiple categories (with scores 0–3 assigned) to capture the risk gradient, while other risk factors were binary (with score 0 or 1 assigned). The six risk factors jointly yielded an AUC of 0.79.

The diagnostic characteristics of different cutpoints for total score were assessed in the development as well as validation NHANES. We selected the cutpoint 5 to designate an individual as having a high risk for undiagnosed diabetes. This cutpoint defined approximately 35% of the adult population as high risk for undiagnosed diabetes and yielded a sensitivity of 79%, specificity of 67%, PPV of 10%, and NPV of 99%, with an AUC of 0.83 in the validation NHANES dataset (see Table 3). Based on these results, if we assume 1,000 new persons will go through the risk assessment and use the cutpoint of 5, then we can estimate 350 persons (35%) would undergo diagnostic testing, 31 new cases of diabetes would be identified, and 6–7 persons with diabetes would remain untested and undetected(31). If a lower cutpoint of 4 is used, then approximately 510 persons (51%) would undergo diagnostic testing, and we can expect 41 cases of diabetes newly identified and fewer than 3 diabetes cases untested and undetected.

When our final prediction model was re-fitted to ARIC/CHS, consistent results were obtained: all of the risk factors were significant (with p values < 0.001) and the magnitude of the associations was comparable, with an AUC of 0.74 (see Appendix Table 1, available at www.annals.org).

Figure 1 depicts the prevalence of undiagnosed diabetes for individual total scores in NHANES and ARIC/CHS. A monotonic (quadratic) relationship was clearly observed. ARIC/CHS showed higher disease prevalence than NHANES, probably due to older ages in the ARIC/CHS populations (> 45 years old). Table 4 summarizes the performance characteristics of the existing guidelines or scores and our method. Our screening score (cutpoint 5) tended to identify smaller proportions of people being at high risk but resulted in higher overall test accuracy (reflected in Youden index), PPV, and likelihood ratio for a positive test, compared to other methods. NPV was high (> 0.96) for all methods. Among existing methods, the Rotterdam model (developed from a European sample) and the new ADA questionnaire seemed to perform best.

We performed three ancillary analyses, detailed in the Methods. We again used cutpoint scores of 5 and 4; results below are for a cutpoint of 5 with values in parentheses reflecting a cutpoint of 4. In the first ancillary analysis, discrimination ability was somewhat reduced as anticipated when the endpoint combined undiagnosed diabetes and pre-diabetes together (AUC=0.72), yielding a sensitivity of 57 (vs. 73)%, specificity of 74 (57)%, PPV of 56 (50)%, and NPV of 74 (78)%. In the second ancillary analysis using only participants > 45 years old, we had a sensitivity of 88 (97)%, specificity of 40 (20)%, PPV of 9 (8)%, and NPV of 98 (99)% with an AUC=0.73. Using only participants < 45 years old, we had a sensitivity of 35 (76)%, specificity of 93 (80)%, PPV of 6 (5)%, and NPV of 99 (100)% with an AUC=0.83; this analysis may be limited due to a small number of diabetes cases. Lastly, when we used hemoglobin A1c in place of fasting glucose for the diabetes definition, we obtained a sensitivity of 80 (91)%, specificity of 63 (47)%, PPV of 6 (5)%, and NPV of 99 (99)% with an AUC of 0.78. PPV is directly proportional to the prevalence of the disease/

condition (26, 32), explaining why our score, like previous methods, yielded lower PPV for these outcomes.

Finally, a sample questionnaire that can be used for community screening for undiagnosed diabetes or pre-diabetes is provided in Figure 2.

DISCUSSION

Clinical trials demonstrate that high risk individuals can reduce their risk of diabetes by more than half when they follow a well-structured, intensive, life style modification program(6, 8, 18). Therefore, early diagnosis could be crucial to reduce the global burden of diabetes. Widespread blood glucose testing may not be the best way to identify undiagnosed diabetes in large community or resource limited settings. Indeed, existing recommendations for diabetes screening that rely on blood testing are not widely followed, resulting in 30% of diabetics going undiagnosed(4).

Our goal was to develop a screening score that can be used in a wide variety of community settings and clinical encounters (including patient waiting rooms or internet) via a simple pencil-and-paper method. Our new diabetes score appeared to perform better than existing methods by quantitative criteria. We believe that it also has good feasibility characteristics – as a simple (with 6 easily answered health-related questions) and efficient (with minimal time needed for survey and no need for a calculator with the maximum score less than 10) screening score with which patients and health care providers can assess their or their patients' need for formal diabetes testing.

We found that the national guidelines for diabetes screening did not perform very well. The three diabetes risk assessment scores showed lower overall accuracy and tended to select larger proportions of people for diabetes screening compared to our new score. Low specificities of existing methods have been reported previously(33–35). The screening criteria recommended by different organizations were developed using different frameworks and purposes, e.g., to enhance efficiency of screening or to target those who could benefit most from screening. So although they differ in numerical performance characteristics (e.g., sensitivity and specificity) based on our analysis, they may be more appropriate for those purposes.

The primary endpoint in our study was undiagnosed diabetes rather than the composite outcome of undiagnosed diabetes and pre-diabetes, but the same questionnaire may well be justified for these closely related outcomes (a disease and its precursor) with different cutpoints (5 for diabetes and 4 for pre-diabetes) based on the evidence obtained from our ancillary analyses. In addition, our score is for prediction of currently undiagnosed diabetes and not for incident diabetes in the future. However, strong consistency in risk factors for the prediction of prevalent and incident events in diabetes and other chronic diseases has been reported(36–38), and we expect that the same set of risk factors in our model play important roles in the prediction of future diabetes or pre-diabetes. Nonetheless, other laboratory or behavioral/lifestyle variables could be useful in predicting future events rather than current events(18, 21, 31, 39–41).

A risk prediction approach that can capture a continuous risk spectrum is a popular tool that has been used to identify important risk factors and to estimate average risk; results can be used in decision making about public health and clinical care. Risk prediction has even been proposed as an alternative to diagnosis for some diseases(42). We believe that ideal risk assessment methods or prediction models should be derived from large representative samples of a target population and consist of fixed and modifiable risk factors together. Simplicity and user-friendliness (including optimal presentation), in addition to accuracy,

are keys for successful implementation and utilization, especially for lay persons(25, 39). To achieve these goals, we 1) adopted a statistical method that yields a systematic scoring system and accounts for design effects of the study appropriately (i.e., logistic regression model suited for complex survey data); 2) carefully selected a parsimonious set of predictors (guided not only by numerical and scientific evidence but also by feasibility perspectives); 3) chose categorized variables in intuitive or well-accepted ways (e.g., using deciles for age and obesity definition); and 4) emphasized an educational purpose of the screening score, highlighting the important risk factors to motivate high-risk people to be screened or to modify health behaviors (e.g., combining body mass index and waist circumference together, rather than using height, weight and waist separately). This combination of factors may explain the enhanced properties of our new score.

For this study, we tried to identify all existing screening guidelines or risk assessment scores for prevalent undiagnosed diabetes available for the U.S. population and one best-suited score for non-U.S. population for comparisons. We found that there are 3 national guidelines and 2 scores/questionnaires for diabetes screening in the U.S., whereas many prediction models exist for incident diabetes. Our search for the best suited non-U.S. model was guided by recent comparison studies(36, 43); we selected the Rotterdam model as it was developed for prevalent undiagnosed diabetes, has been externally validated in different samples, and only requires routinely available demographic or health information in its simple scoring system.

This study does have some limitations. First, some variables that are parts of existing methods (e.g., gestational diabetes) were not available in the databases we used. Therefore, some caution should be exercised in making comparisons between our and others' methods. Nonetheless, we believe that the vast majority of key information was available and utilized, minimizing the unfairness in the comparisons. Second, we could not incorporate oral glucose tolerance test results because these data were not collected in the newer NHANES (1999–2006) and in the baseline visits in ARIC and CHS. Thus, we defined the outcome based solely on the FPG. The FPG is a recommended screening test, however, and the lack of oral glucose tolerance test data has not been shown previously to affect the stability of diabetes risk assessment methods(4, 44). Our results seemed to be robust to different definitions of the endpoint, either based on FPG or Hemoglobin A1c (e.g., AUC=0.79 vs. 0.78).

Although the lay population is increasingly appreciating the danger of diabetes and its complications, more education is still needed in community and clinical settings. In that sense, although further validation of our screening score in other samples is important, this newly developed algorithm could still have immediate applications. In addition to its use in clinical encounters, targeted screenings, and health education programs, the screening score can be applied by health plans to existing databases for case-finding. The new algorithm can also potentially help identify optimal populations for enrollment in clinical trials that test new strategies to prevent or manage diabetes.

In conclusion, we envision our screening score to serve as a method for identifying individuals in need of formal diabetes screening and calling for more attention to pre-diabetes. A self-assessment method that helps people decide whether they should seek medical care for diabetes testing may serve as one way to address the lack of interaction with health care facilities/providers that may underlie the high percentage of the population with undiagnosed diabetes, particularly the underserved. Although a consensus on diabetes screening has not yet been reached(45, 46), we believe a priority in formal screening for undiagnosed diabetes should be given to those who are at high risk. This new diabetes

screening score could help identify these high risk individuals, while patients and caregivers alike await more definitive evidence-based recommendations(47, 48).

Acknowledgments

The authors thank the staff and participants of the NHANES, ARIC and CHS studies for their important contributions to research and data sharing.

Financial Support: Dr. Bang and Ms. Edwards were partially supported by the Clinical and Translational Science Center at Weill Cornell Medical College (UL1-RR024996).

REFERENCES

1. Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabetic Medicine*. 1997; 14(Suppl 5)(5):S1–S85. [PubMed: 9450510]
2. Cowie CC, Rust KF, Ford ES, et al. Full accounting of diabetes and pre-diabetes in the U. S population in 1988–1994 and 2005–2006. *Diabetes Care*. 2009; 32(2):287–294. [PubMed: 19017771]
3. Prevalence of diabetes and impaired fasting glucose in adults -- United States: 1999–2000. *MMWR Morb Mortal Wkly Rep*. 2003; 52(35):833–837. [PubMed: 12966357]
4. Heikes KE, Eddy DM, Arondekar B, Schlessinger L. Diabetes Risk Calculator: a simple tool for detecting undiagnosed diabetes and pre-diabetes. *Diabetes Care*. 2008; 31(5):1040–1045. [PubMed: 18070993]
5. Goyder EC, McNally PG, Drucquer M, Spiers N, Botha JL. Shifting of care for diabetes from secondary to primary care: 1990–5: review of general practices. *BMJ*. 1998; 316:1505–1506. [PubMed: 9582142]
6. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002; 346(6):393–403. [PubMed: 11832527]
7. Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. *J Am Coll Cardiol*. 2004; 44(3):720–732. [PubMed: 15358046]
8. Tuomilehto J, Lindström J, Eriksson JG, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*. 2001; 344(18):1343–1350. [PubMed: 11333990]
9. Chiasson J-L, Josse RG, Gomis R, Hanefeld M, Karasik A, Laakso M. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *The Lancet*. 2002; 359(9323):2072–2077.
10. Buchanan TA, Xiang AH, Peters RK, et al. Preservation of pancreatic beta-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk hispanic women. *Diabetes*. 2002; 51(9):2796–2803. [PubMed: 12196473]
11. The CDC Diabetes Cost-Effectiveness Study Group. The cost-effectiveness of screening for type 2 diabetes. *JAMA*. 1998; 280(20):1757–1763. [PubMed: 9842951]
12. American Diabetes Association. Standards of medical care in diabetes--2009. *Diabetes Care*. 2009; 32(Suppl 1)(1):S13–S61. [PubMed: 19118286]
13. US Preventive Services Task Force. Screening for type 2 diabetes mellitus in adults: U. S Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*. 2008; 148(11):846–854. [PubMed: 18519930]
14. Herman WH, Smith PJ, Thompson TJ, Engelgau MM, Aubert RE. A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes. *Diabetes Care*. 1995; 18(3):382–387. [PubMed: 7555482]
15. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *American Journal of Epidemiology*. 1989; 129:687–702. [PubMed: 2646917]
16. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Annals of Epidemiology*. 1991; 1(3):263–276. [PubMed: 1669507]

17. International Classification of Diseases, 9th Revision, Clinical Modification. 5th ed. Salt Lake City, Utah: Medicode; 1998.
18. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. *Diabetes Care*. 2005; 28(8):2013–2018. [PubMed: 16043747]
19. Nguyen NT, Varela E, Sabio A, Tran C-L, Stamos M, Wilson SE. Resolution of hyperlipidemia after laparoscopic Roux-en-Y gastric bypass. *Journal of the American College of Surgeons*. 2006; 203(1):24–29. [PubMed: 16798484]
20. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA*. 2001; 285(19):2486–2497. [PubMed: 11368702]
21. Duncan BB, Schmidt MI, Pankow JS, et al. Adiponectin and the development of type 2 diabetes: the atherosclerosis risk in communities study. *Diabetes*. 2004; 53(9):2473–2478. [PubMed: 15331562]
22. American Diabetes Association. Standards of medical care in diabetes--2008. *Diabetes Care*. 2008; 31(Suppl 1)(1):S12–S54. [PubMed: 18165335]
23. Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York: John Wiley & Sons, Inc.; 1987.
24. SAS Institute Inc. SAS/STAT User's Guide Version 8.2. Cary, NC: SAS Institute Inc.; 2000.
25. Grundy, SM.; D'Agostino, RB.; Mosca, L., et al. Summary of NHLBI workshop on cardiovascular risk assessment. 1999. Accessed at <http://www.nhlbi.nih.gov/resources/docs/cvrisk.pdf> on 31 August 2009.
26. Gonen, M. Analyzing Receiver Operating Characteristic Curves with SAS. SAS Publishing, Cary, NC: SAS Institute Inc.; 2007.
27. Sackett, D.; Straus, SE.; Richardson, WS.; Rosenberg, W.; Haynes, B. Evidence-Based Medicine: How to Practice and Teach EBM. 2nd ed. Philadelphia, PA: Churchill Livingstone; 2000.
28. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950; 3(1):32–35. [PubMed: 15405679]
29. Baan CA, Ruige JB, Stolk RP, et al. Performance of a predictive model to identify undiagnosed diabetes in a health care setting. *Diabetes Care*. 1999; 22(2):213–219. [PubMed: 10333936]
30. The International Expert Committee. International expert committee report on the role of the A1c assay in the diagnosis of diabetes. *Diabetes Care*. 2009; 32(7):1–8.
31. Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. *Diabetes Metab Res Rev*. 2000; 16(3):164–171. [PubMed: 10867715]
32. Nathan DM, Singer DE, Hurxthal K, Goodson JD. The clinical information value of the glycosylated hemoglobin assay. *N Engl J Med*. 1984; 310:341–346. [PubMed: 6690962]
33. Dunstan DW, Zimmet PZ, Welborn TA, et al. The rising prevalence of diabetes and impaired glucose tolerance: the Australian Diabetes, Obesity and Lifestyle Study. *Diabetes Care*. 2002; 25(5):829–834. [PubMed: 11978676]
34. Knudson PE, Turner KJ, Sedore A, Weinstock RS. Utility of the American Diabetes Association risk test in a community screening program. *Diabetes Care*. 1998; 21(6):1029–1031. [PubMed: 9614631]
35. Rolka DB, Narayan KM, Thompson TJ, et al. Performance of recommended screening tests for undiagnosed diabetes and dysglycemia. *Diabetes Care*. 2001; 24(11):1899–1903. [PubMed: 11679454]
36. Rathmann W, Martin S, Haastert B, et al. Performance of screening questionnaires and risk scores for undiagnosed diabetes: the KORA Survey 2000. *Archives of Internal Medicine*. 2005; 165(4): 436–441. [PubMed: 15738374]
37. Bang H, Vupputuri S, Shoham DA, et al. SCReening for Occult RENal Disease (SCORED): a simple prediction model for chronic kidney disease. *Archives of Internal Medicine*. 2007; 167(4): 374–381. [PubMed: 17325299]
38. Kshirsagar AV, Bang H, Bombardieri AS, et al. A simple algorithm to predict incident kidney disease. *Archives of Internal Medicine*. 2008; 168(22):2466–2473. [PubMed: 19064831]

39. Herman WH. Predicting risk for diabetes: Choosing (or building) the right model. *Annals of Internal Medicine*. 2009; 150:812–814. [PubMed: 19487716]
40. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years. *Annals of Internal Medicine*. 2009; 150:741–751. [PubMed: 19487709]
41. Schulze MB, Hoffmann K, Boeing H, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care*. 2007; 30(3):510–515. [PubMed: 17327313]
42. Vickers AJ, Basch E, Kattan MW. Against diagnosis. *Annals of Internal Medicine*. 2008; 149(3): 200–203. [PubMed: 18678847]
43. Glümer C, Vistisen D, Borch-Johnsen K, Colagiuri S. and on behalf of the DETECT-2 Collaboration. Risk scores for type 2 diabetes can be applied in some populations but not all. *Diabetes Care*. 2006; 29:410–414. [PubMed: 16443896]
44. Norris SL, Kansagara D, Bougatsos C, Fu R. Screening adults for type 2 diabetes: a review of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*. 2008; 148(11): 855–868. [PubMed: 18519931]
45. Sheehy AM, Coursin DB, Gabbay RA. Back to Wilson and Jungner: 10 good reasons to screen for type 2 diabetes mellitus. *Mayo Clinic Proceedings*. 2009; 84(1):38–42. [PubMed: 19121253]
46. Lawler FH. Reasons to exercise caution when considering a screening program for type 2 diabetes mellitus. *Mayo Clinic Proceedings*. 2009; 84(1):34–36. [PubMed: 19121252]
47. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. *Archives of Internal Medicine*. 2004; 164(20):2223–2228. [PubMed: 15534158]
48. Kaplah, JP.; Liverman, CT.; Kraak, VI., editors. *Preventing Childhood Obesity: Health in the Balance*. Washington, DC: Institute of Medicine; 2004.

Appendix Table 1

Final regression model fitted to external validation datasets, ARIC/CHS (N=19,728, AUC=0.74)

Variable	Odds Ratio (95% CI)	P-value	Log (Odds Ratio)
Age in years *			
<50	reference	---	---
50–59	1.4 (1.1–1.8)	0.0014	0.36
60	2.5 (2.0–3.1)	<0.0001	0.91
Sex			
Female	reference	---	---
Male	1.8 (1.6–2.0)	<0.0001	0.57
Family history of diabetes			
No	reference	---	---
Yes	1.9 (1.7–2.2)	< 0.0001	0.66
History of hypertension			
No	reference	---	---
Yes	2.3 (2.0–2.6)	< 0.0001	0.83
Obesity **			
Not overweight or obese	reference	---	---
Overweight	1.8(1.4–2.3)	< 0.0001	0.57
Obese	3.6 (2.9–4.5)	< 0.0001	1.28

Variable	Odds Ratio (95% CI)	P-value	Log (Odds Ratio)
Extremely obese	8.8 (6.2–12.4)	<0.0001	2.18
Physically active?			
No	reference	---	---
Yes	0.8 (0.7–0.9)	0.0004	–0.23

* Minimum age is 45 in ARIC/CHS.

** if (BMI ≥ 40 kg/m²) or (waist ≥ 50 inches for male) or (waist ≥ 49 inches for female) then extremely obese;
 else if (30 \leq BMI < 40) or (40 \leq waist < 50 for male) or (35 \leq waist < 49 for female) then obese;
 else if (25 \leq BMI < 30) or (37 \leq waist < 40 for male) or (31.5 \leq waist < 35 for female) then overweight;
 else not overweight or obese.

ARIC = Atherosclerosis Risk in Communities; AUC = area under the receiver-operating-characteristic curve; BMI = body mass index; CHS = Cardiovascular Health Study.

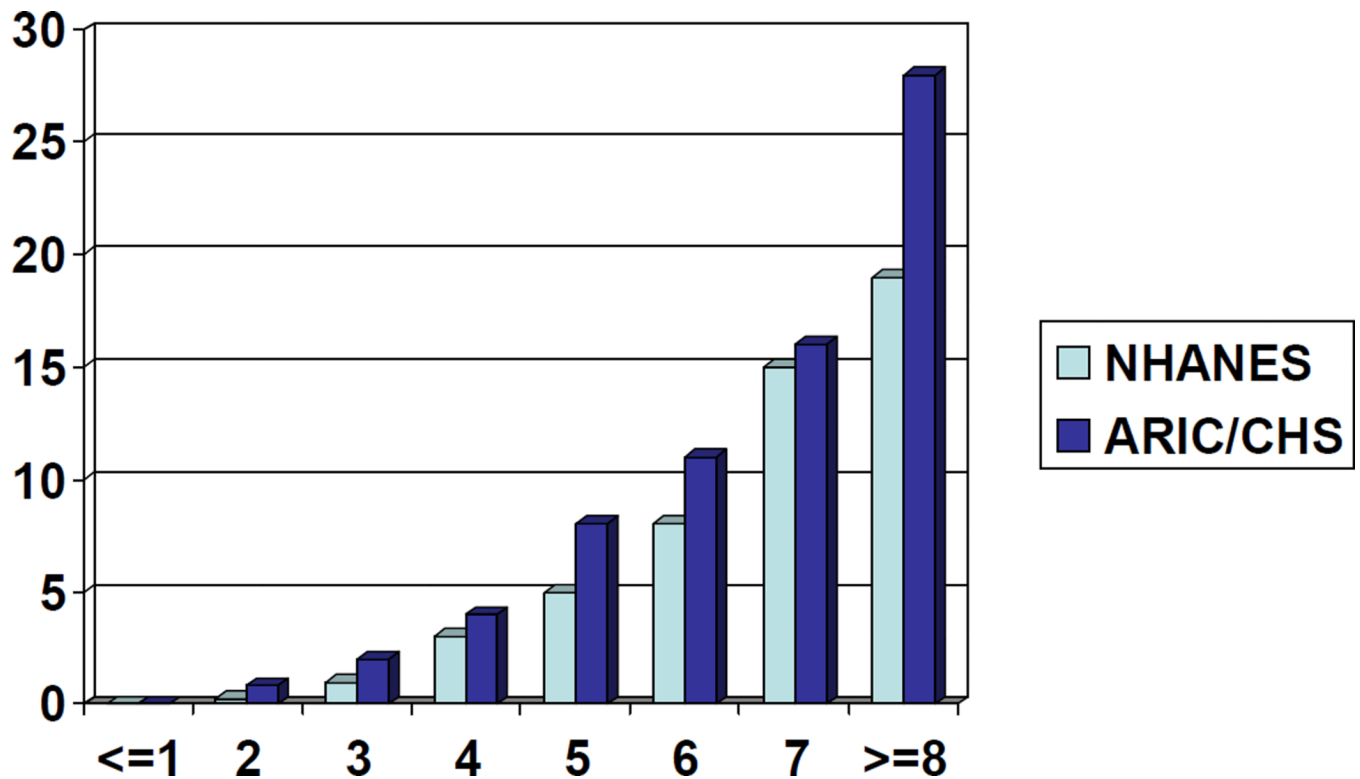


Figure 1. The estimated prevalence of undiagnosed diabetes by screening score
 NHANES 1999–2006 has N=6,898 and ARIC/CHS has N=19,728.

Proportions of individuals for scores of (–1,0,1,2,3,4,5,6,7,8,9) correspond to (1%,4%,10%,14%,16%,17%,16%,13%,7%,2%,0.1%) in NHANES and (2%,5%,11%,19%,24%,22%,13%,4%,0.4%,0.01%) in ARIC/CHS.

ARIC = Atherosclerosis Risk in Communities; CHS = Cardiovascular Health Study;
 NHANES = National Health and Nutrition Examination Survey.

Label of Y-axis: Prevalence, %

Label of X-axis: Total Screening Score

Question	Score	Enter your score (enter 0, if you don't know)
1. How old are you?	<40 years (0 point) 40-49 years (1 point) 50-59 years (2 points) 60 years or older (3 points)	
2. Are you a woman or man?	Woman (0 point) Man (1 point)	
3. Do your family members (parent or sibling) have diabetes?	No (0 point) Yes (1 point)	
4. Do you have high blood pressure or are you on medication for high blood pressure?	No (0 point) Yes (1 point)	
5. Are you overweight or obese? (see chart below to answer this question more accurately)	Not overweight or obese (0 point) Overweight (1 point) Obese (2 points) Extremely obese (3 points)	
6. Are you physically active?	No (0 point) Yes (-1 point)	
TOTAL SCORE (add points from questions 1-6)		
If your TOTAL SCORE is ≥ 4 , you are at high risk of having undiagnosed diabetes or pre-diabetes. If your TOTAL SCORE is ≥ 5 , you are at high risk of having undiagnosed diabetes. See your doctor for a blood test to look for diabetes if your score is high.		

Obesity definition:
if (BMI ≥ 40 kg/m²) or (waist ≥ 50 inches for male) or (waist ≥ 49 inches for female) then extremely obese;
else if (30 \leq BMI < 40) or (40 \leq waist < 50 for male) or (35 \leq waist < 49 for female) then obese;
else if (25 \leq BMI < 30) or (37 \leq waist < 40 for male) or (31.5 \leq waist < 35 for female) then overweight;
else not overweight or obese.

<Body Mass Index (BMI) Chart>

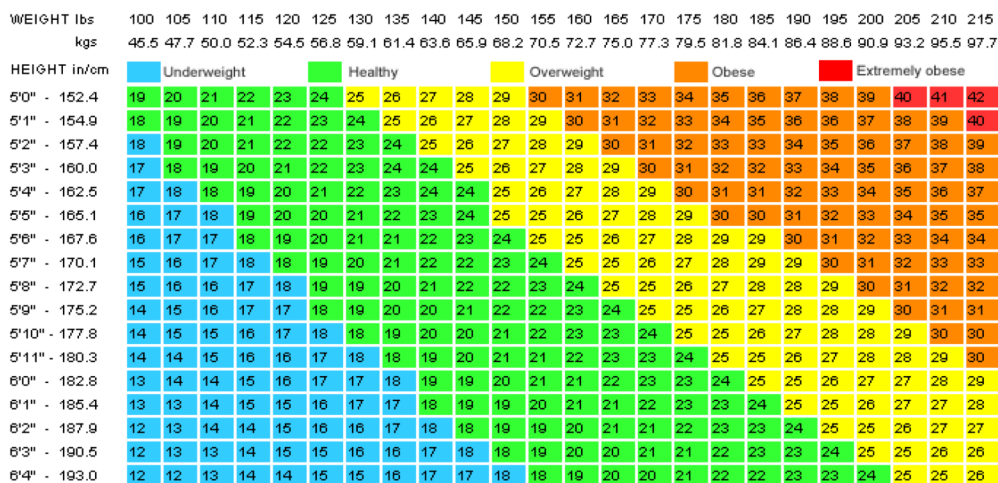


Figure 2. Self-assessment screening score for undiagnosed diabetes or pre-diabetes

Table 1

Characteristics of participants according to diabetes status in NHANES 1999–2004: Mean (standard error) for continuous variables and percentage (%) for categorical variables

Characteristic	Normal glucose (N=3,396)	Impaired glucose (N=1,652)	Undiagnosed diabetes (N=210)	Known diabetes (N=482)
Age, years	42.4 (0.49)	51.8 (0.54)	58.3 (1.65)	57.6 (0.88)
Male	44	60	62	51
White race	72	75	74	61
Education, high school	83	76	67	70
Annual household income, \$25K	65	63	47	49
Married	58	64	57	62
Having health insurance	81	83	83	92
Number of health care visits in last year	1.9 (0.03)	2.0 (0.04)	2.1 (0.13)	3.0(0.06)
History of hypertension	28	44	67	69
History of cardiovascular disease	3	7	12	14
Hyperlipidemia	52	67	70	64
High cholesterol	34	50	57	78
Family history of diabetes	46	48	61	75
Mother or father	20	27	35	47
Brother or sister	7	11	18	31
Family history of cardiovascular disease	40	35	33	49
Smoking status				
Current smokers	26	21	20	23
Former smokers	22	32	43	32
Being physically active	36	36	30	29
Systolic blood pressure, mmHg	120 (0.44)	126.7 (0.55)	135.1 (2.23)	129.6 (1.22)
Diastolic blood pressure, mmHg	72.2 (0.26)	73.5 (0.43)	72.9 (1.37)	70.9 (0.86)
Body mass index, kg/m ²	26.9 (0.11)	29.5 (0.20)	32.3 (0.83)	31.6 (0.57)
Waist circumference, inches	36.4 (0.12)	39.9 (0.20)	43.1 (0.65)	42.3 (0.57)
Total cholesterol, mmol/L	5.14 (0.03)	5.36 (0.03)	5.41 (0.11)	5.16 (0.07)
mg/dL	198.6 (1.10)	207.1 (1.22)	209.2 (4.32)	199.6 (2.58)
Triglycerides, mmol/L	1.45 (0.03)	1.91 (0.05)	2.82 (0.37)	2.22 (0.09)
mg/dL	128.9 (2.90)	169.0 (4.57)	249.4 (33.0)	196.6 (8.38)
High density lipoprotein, mmol/L	1.39 (0.01)	1.27 (0.01)	1.18 (0.03)	1.22 (0.03)
mg/dL	53.7 (0.38)	49.1 (0.56)	45.6 (1.15)	47.0 (0.99)
Low density lipoprotein, mmol/L	3.09 (0.02)	3.25 (0.03)	3.10 (0.08)	2.98 (0.05)
mg/dL	119.5 (0.85)	125.7 (1.17)	119.8 (3.07)	115.1 (1.90)
Hemoglobin A1c, %	5.20 (0.01)	5.27 (0.01)	7.06 (0.20)	7.16 (0.07)

Normal glucose = fasting plasma glucose (FPG)<5.5 mmol/L (100 mg/dL); impaired glucose=FPG 5.6–6.9 mmol/L (100–125 mg/dL); and undiagnosed diabetes = FPG ≥7.0 mmol/L (126 mg/dL).

Sample sizes were unweighted, while summary statistics were weighted in fasting subsamples.

Actual sample sizes are reduced for some variables. For example, income data are missing for 10% participants and other variables had 5% missing data.

Table 2

Risk factors of undiagnosed diabetes: Final regression model in the development dataset, NHANES 1999–2004 (N = 5,258, AUC=0.79)

Risk factor	Odds Ratio (95% CI)	P-value	Log (Odds Ratio)	Score assigned
Age in years				
<40	reference	---	---	0
40–49	2.6 (1.3–5.0)	0.004	0.95	1
50–59	4.8 (2.2–10.6)	0.0001	1.57	2
60	8.1 (3.9–16.9)	<0.0001	2.09	3
Sex				
Female	reference	---	---	0
Male	2.6 (1.8–3.7)	<0.0001	0.96	1
Family history of diabetes				
No	reference	---	---	0
Yes	2.0 (1.5–2.6)	< 0.0001	0.67	1
History of hypertension				
No	reference	---	---	0
Yes	1.9 (1.2–2.9)	0.004	0.64	1
Obesity*				
Not overweight or obese	reference	---	---	0
Overweight	1.3 (0.6–2.8)	0.47	0.27	1
Obese	3.1 (1.6–5.8)	0.0006	1.12	2
Extremely obese	7.3 (4.0–13.4)	<0.0001	1.99	3
Physically active?				
No	reference	---	---	0
Yes	0.7 (0.5–1.0)	0.06	–0.34	–1

* if (BMI ≥ 40 kg/m²) or (waist ≥ 50 inches for male) or (waist ≥ 49 inches for female) then extremely obese; else if (30 \leq BMI < 40) or (40 \leq waist < 50 for male) or (35 \leq waist < 49 for female) then obese; else if (25 \leq BMI < 30) or (37 \leq waist < 40 for male) or (31.5 \leq waist < 35 for female) then overweight; else not overweight or obese.

See Figure 2 for user-friendly questionnaire and BMI chart.

AUC = area under the receiver-operating-characteristic curve; BMI = body mass index; NHANES = National Health and Nutrition Examination Survey.

Table 3

Performance of different cutpoints for detecting undiagnosed diabetes in development and validation datasets in NHANES

Total score	% high risk	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Development data (NHANES 1999–2004, N=5,258, AUC=0.79)							
6	23	61	79	11	98	2.90	0.49
5	39	82	63	8	99	2.22	0.29
4	55	93	46	7	99	1.72	0.15
3	71	98	30	5	100	1.40	0.07
Validation data (NHANES 2005–2006, N=1,640, AUC=0.83)							
6	20	62	82	13	98	3.44	0.46
5	35	79	67	10	99	2.39	0.31
4	51	97	51	8	100	1.98	0.06
3	68	99	33	6	100	1.48	0.03

AUC = area under the receiver-operating-characteristic curve; LR = likelihood ratio; NPV = negative predictive value; PPV = positive predictive value; NHANES = National Health and Nutrition Examination Survey.

Table 4

Performance of diabetes screening methods in validation datasets

Method	% high risk	Sensitivity	Specificity	PPV	NPV	LR+	LR-	Youden index	Number of variables*	Requires clinician's input?
<i>NHANES 2005–2006 (N=1,640)</i>										
CDC	90	100	10	5	100	1.11	0	10	1	No
ADA	82	97	19	5	99	1.20	0.16	16	14 or 15	Yes
USPSTF	28	44	73	7	97	1.63	0.77	17	2	Yes
ADA questionnaire I	57	83	44	6	98	1.48	0.39	27	6 or 7	No
ADA questionnaire II	42	76	59	8	98	1.85	0.41	35	10	No
Rotterdam model	47	89	55	8	99	1.98	0.20	44	4 or 5	No
New screening score	5	79	67	10	99	2.39	0.31	46	6 or 8	No
New screening score	4	51	51	8	100	1.98	0.06	48		
<i>ARIC/CHS (N=19,728)</i>										
CDC	100	100	NA	5	NA	NA	NA	NA	NA	
ADA	100	100	NA	5	NA	NA	NA	NA	NA	
USPSTF	35	53	66	8	96	1.56	0.71	19		Same as above
ADA questionnaire I	60	80	41	7	97	1.36	0.49	21		
ADA questionnaire II	64	86	38	7	98	1.39	0.37	24		
Rotterdam model	55	81	46	8	98	1.50	0.41	27		
New screening score	5	40	62	10	98	1.89	0.45	34		
New screening score	4	64	38	8	99	1.47	0.24	29		

Definitions of screening methods are as follows:

CDC: all adults at 25 years of age or older.

ADA: all adults 45 years of age and older and those younger adults who are overweight or obese [BMI ≥25] and have at least one other risk factor, including physical inactivity, family history of diabetes, minority ethnicity, history of gestational diabetes or delivery of a baby weighing >9 lbs, hypertension, HDL <35 mg/dl and/or triglyceride >250 mg/dl, women with polycystic ovarian syndrome, history of impaired fasting glucose or impaired glucose tolerance, other clinical conditions associated with insulin resistance (e.g., severe obesity), or history of cardiovascular disease (ADA 2009).

USPSTF: all adults with sustained blood pressure (either treated or untreated) greater than 135/80 mm Hg (USPSTF 2008).

ADA diabetes questionnaire I: score_Herman 10, where score_Herman=(woman who delivered a baby weighing 9 lbs)*1+(parental history of diabetes)*1+(sibling's history of diabetes)*1+(BMI 27)*5+(age<65 and not physically active)*5+(45 age<65)*5+(age 65)*9 (Herman et al. 1995).

ADA diabetes questionnaire II: It is a function of age, waist, weight, height, gestational diabetes, parental and sibling's diabetes history, race, high blood pressure, and exercise. See Heikes et al. (2008) for graphical presentation of this algorithm.

Rotterdam model: score_Rotterdam>6, where score_Rotterdam=2 per 5-year increment from 55 years+(male)*5+(use of antihypertensive medications)*4+(BMI 30)*5 (Baan et al. 1999).

New screening score: $\text{score_new} = (40 \text{ age} < 50) * 1 + (50 \text{ age} < 60) * 2 + (\text{age} \geq 60) * 3 + (\text{male}) * 1 + (\text{family history of diabetes}) * 1 + (\text{history of hypertension}) * 1 + (\text{obese}) * 2 + (\text{extreme obese}) * 3 - \text{exercise} * 1$.

Note: Data on history of polycystic ovarian syndrome, history of impaired fasting glucose or impaired glucose tolerance and pregnancy were not collected so that these conditions were omitted for the ADA guideline. ARIC and NHANES 2005–6 collected the data on family history of diabetes but not separately for parents and sibling. So we replaced parental history by family history for the ADA questionnaires.

* A smaller number indicates when a person knows his/her obesity status or BMI value.

ADA = American Diabetes Association; ARIC = Atherosclerosis Risk in Communities; CDC = Centers for Disease Control and Prevention; CHS = Cardiovascular Health Study; LR = likelihood ratio; NHANES = National Health and Nutrition Examination Survey; NPV = negative predictive value; PPV = positive predictive value; USPSTF = U.S. Preventive Services Task Force.