

Diabetes Risk Calculator

A simple tool for detecting undiagnosed diabetes and pre-diabetes

KENNETH E. HEIKES, PHD¹
DAVID M. EDDY, MD, PHD¹

BHAKTI ARONDEKAR, MBA, PHD²
LEONARD SCHLESSINGER, PHD¹

OBJECTIVE — The objective of this study was to develop a simple tool for the U.S. population to calculate the probability that an individual has either undiagnosed diabetes or pre-diabetes.

RESEARCH DESIGN AND METHODS — We used data from the Third National Health and Nutrition Examination Survey (NHANES) and two methods (logistic regression and classification tree analysis) to build two models. We selected the classification tree model on the basis of its equivalent accuracy but greater ease of use.

RESULTS — The resulting tool, called the Diabetes Risk Calculator, includes questions on age, waist circumference, gestational diabetes, height, race/ethnicity, hypertension, family history, and exercise. Each terminal node specifies an individual's probability of pre-diabetes or of undiagnosed diabetes. Terminal nodes can also be used categorically to designate an individual as having a high risk for 1) undiagnosed diabetes or pre-diabetes, 2) pre-diabetes, or 3) neither undiagnosed diabetes or pre-diabetes. With these classifications, the sensitivity, specificity, positive and negative predictive values, and receiver operating characteristic area for detecting undiagnosed diabetes are 88%, 75%, 14%, 99.3%, and 0.85, respectively. For pre-diabetes or undiagnosed diabetes, the results are 75%, 65%, 49%, 85%, and 0.75, respectively. We validated the tool using v-fold cross-validation and performed an independent validation against NHANES 1999–2004 data.

CONCLUSIONS — The Diabetes Risk Calculator is the only currently available noninvasive screening tool designed and validated to detect both pre-diabetes and undiagnosed diabetes in the U.S. population.

Diabetes Care 31:1040–1045, 2008

The objective of this study was to develop a simple, self-administered, paper-based screening tool that could be used by the public to determine their risk of having pre-diabetes or undiagnosed diabetes and to help people decide whether they should see a physician for further evaluation. To maximize its accessibility and ease of use, the tool should use only information that is commonly known to an average individ-

ual and preferably should not require any calculations.

The prevalence of diabetes is growing rapidly, with the total number of cases worldwide projected to increase from 171 million in 2000 to 366 million by 2030 (1). In the U.S. in 2002, the prevalence of diabetes was estimated to be 19.3 million, of which about 5.8 million cases were undiagnosed (2). An additional 41 million individuals are estimated to have pre-diabetes,

defined as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT). Pre-diabetes implies an increased risk of development of type 2 diabetes on the order of 30% over 4 years (3) and 70% over 30 years (4). Several studies have demonstrated that type 2 diabetes can be prevented or delayed with lifestyle modification or the use of pharmacotherapy in subjects with pre-diabetes (3,5). Studies have also indicated that preventing or delaying the onset of type 2 diabetes by lifestyle modification or the use of pharmacotherapy can be cost-effective (6) if costs of the interventions are controlled (4).

An important step in preventing or delaying type 2 diabetes and its complications is to identify people with pre-diabetes and undiagnosed diabetes so that they can be given appropriate care. The American Diabetes Association recommends screening for type 2 diabetes at 3-year intervals beginning at age 45, particularly in those with BMI ≥ 25 kg/m² (7). However these recommendations are not widely followed, as indicated by the fact that in 30% of people who have diabetes it is still undiagnosed. Major reasons for this problem are the cost and inconvenience of testing.

One way to address this problem is to develop a simple, inexpensive tool that can identify people who are at high risk of having pre-diabetes or undiagnosed diabetes and motivate them to be screened. Several investigators have developed diabetes risk assessment tools. However, most of those tools apply to non-U.S. populations and none were designed to detect pre-diabetes and undiagnosed diabetes. The objective of this study was to develop a simple tool for use in the U.S. to identify people who have a high probability of having pre-diabetes or undiagnosed diabetes, using only information that is commonly known to an average individual and preferably not requiring any calculations.

RESEARCH DESIGN AND METHODS

Definitions

The definitions of pre-diabetes and diabetes are based on fasting plasma glucose

From ¹Archimedes, San Francisco, California; and ²GlaxoSmithKline, Philadelphia, Pennsylvania.

Corresponding author: David Eddy, MD, PhD, 201 Mission St., 29th Floor, San Francisco, CA 94105. E-mail: author@archimedesmodel.com.

Received for publication 18 June 2007 and accepted in revised form 5 December 2007.

Published ahead of print at <http://care.diabetesjournals.org> on 10 December 2007. DOI: 10.2337/dc07-1150.

Additional information for this article can be found in an online appendix at <http://dx.doi.org/10.2337/dc07-1150>.

K.E.H. has received consulting fees from GlaxoSmithKline, and D.M.E. and L.S. have received consulting fees from GlaxoSmithKline, Pfizer Inc., and Eli Lilly.

Abbreviations: CART, classification and regression tree; DRC, Diabetes Risk Calculator; FPG, fasting plasma glucose; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; NHANES, National Health and Nutrition Examination Survey; OGTT, oral glucose tolerance test; ROC, receiver operating characteristic.

© 2008 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

See accompanying editorial, p. 1084.

(FPG) and glucose tolerance, as measured by a 2-h plasma oral glucose tolerance test (OGTT). IFG is defined as FPG of 100–125 mg/dl. IGT is defined as 2-h OGTT result of 140–199 mg/dl. Diabetes is defined as FPG \geq 126 mg/dl and/or 2-h OGTT result \geq 200 mg/dl. Pre-diabetes is defined as IFG and/or IGT without diabetes. Undiagnosed diabetes is defined as the presence of actual diabetes based on FPG and/or a 2-h OGTT and the absence of an individual having been told that he or she has diabetes. We use the term “elevated plasma glucose” to define an individual who has either pre-diabetes or undiagnosed diabetes.

Data

We used data from the Third National Health and Nutrition Examination Survey (NHANES) (1988–1994) to build and internally validate the tool (8). This dataset was chosen for two main reasons. First, it is a representative sample of the U.S. population, which is the main population for which the tool is being designed. The survey methods oversampled certain subpopulations, such as race/ethnicity minorities but provided weights to enable construction of a representative sample of the U.S. Second, the NHANES III dataset is the most current NHANES survey that contains 2-h OGTT results for a large subsample of people, as well as FPG and the pertinent characteristics and risk factors needed to build a tool. Our analysis was based on the results of the 7,092 participants who were aged \geq 20 years and had FPG results. Two-hour OGTT data were available for approximately half of those aged 40–75 years. For people for whom 2-h OGTT results were missing, the diagnoses were based on FPG alone. An analysis of the group for whom both FPG and OGTT data were available revealed that the lack of OGTT data for some of the participants did not materially affect the stability of the results; the overall effect of the missing data was to underestimate the prevalences of pre-diabetes and undiagnosed diabetes by \sim 2 and 1.5%, respectively. Additional details about the data collection and analysis are described in a technical report available as an online appendix at <http://dx.doi.org/10.2337/dc07-1150>.

Explanatory variables

To build the tool, we examined 18 explanatory variables that would be known to the average individual and would not require laboratory results or special med-

ical definitions. These included BMI, height, weight, waist circumference, waist-to-hip ratio, age, sex, race/ethnicity, taking blood pressure medication, taking cholesterol medication, gestational diabetes, high blood pressure, high cholesterol, history of diabetes (any blood relative), history of diabetes (parent or sibling), history of diabetes (parent), history of diabetes (sibling), and exercise compared with peers. Not all variables were used in the final tool; their inclusion in the final models depended on their value as predictors of pre-diabetes and undiagnosed diabetes.

Strategy

To help ensure that we developed the best possible tool from the available data, we built two different tools using different methods, compared them, and selected the one that best served our objectives of simplicity and accuracy.

Analytical methods

Logistic regression. One tool was developed using logistic regression, in which the probability of pre-diabetes or undiagnosed diabetes was modeled by a logistic, or log-odds, transformation

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (1)$$

where the x_i 's are the continuous or dichotomous explanatory variables, the β_i 's are the regression coefficients estimated using maximum likelihood methods, and the response variable was assumed to have a binomial distribution. The probability can be rewritten as

$$p = (1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)})^{-1} \quad (2)$$

Logistic regressions were generated with SAS (version 9.1; SAS Institute, Cary, NC).

Classification and regression tree. We also built a model using a classification and regression tree method (CART) (CART software version 5.0; Salford Systems, San Diego, CA). This technique separates data into mutually exclusive groups that concentrate a particular class of the target variable. In our analysis the target variable could take a value of 0 or 1 depending on whether undiagnosed diabetes (or pre-diabetes) is absent or present, respectively. The value of a target variable is referred to as its class. Starting at the tree root, the data are split into two

groups conditional on whether an explanatory variable or a linear combination of explanatory variables is greater than some value. The particular value of the explanatory variable selected for the split is the one that best separates the target classes, 0 or 1, at the root node into two child nodes. If the primary splitter variable field is missing, a surrogate splitter variable is used instead.

The process then repeats for each of the child nodes. Subsequent splits can involve another explanatory variable or a different value of a previously used variable. A node that is not further split is referred to as a terminal node. Each terminal node is assigned to a target class conditional on whether the prevalence of the target class exceeds a designated threshold. The same threshold is applied to all terminal nodes and determines overall sensitivity and specificity of the tree. The classification tree is grown to its maximum size and then pruned on the basis of a criterion that balances the number of terminal nodes (complexity) against the accuracy of the tree in classifying people, sometimes termed misclassification cost.

To develop a single tree that could be used to detect either pre-diabetes or undiagnosed diabetes, we used an approach analogous to that used for the regression model. Specifically, we first developed a tree to predict undiagnosed diabetes and then applied a different threshold to predict pre-diabetes. Because one of the goals was to create a simple model, BMI and waist-to-hip ratio were dropped from the list of variables in favor of weight, height, and waist circumference, which required no calculation but still maintained the accuracy of the tool. We eliminated the cholesterol variables high cholesterol and taking cholesterol medication because of the large number of missing fields and low predictive value. We also eliminated history of diabetes in any blood relative in favor of the more specific diabetes history variables (history of diabetes in a parent or sibling, in a parent only, or in a sibling only).

We used v-fold cross-validation to train and test the classification tree models, partitioning the data into equal-sized subsets (9). We then derived and tested the classification tree on all combinations of 9 of 10 training data and 1 of 10 test data.

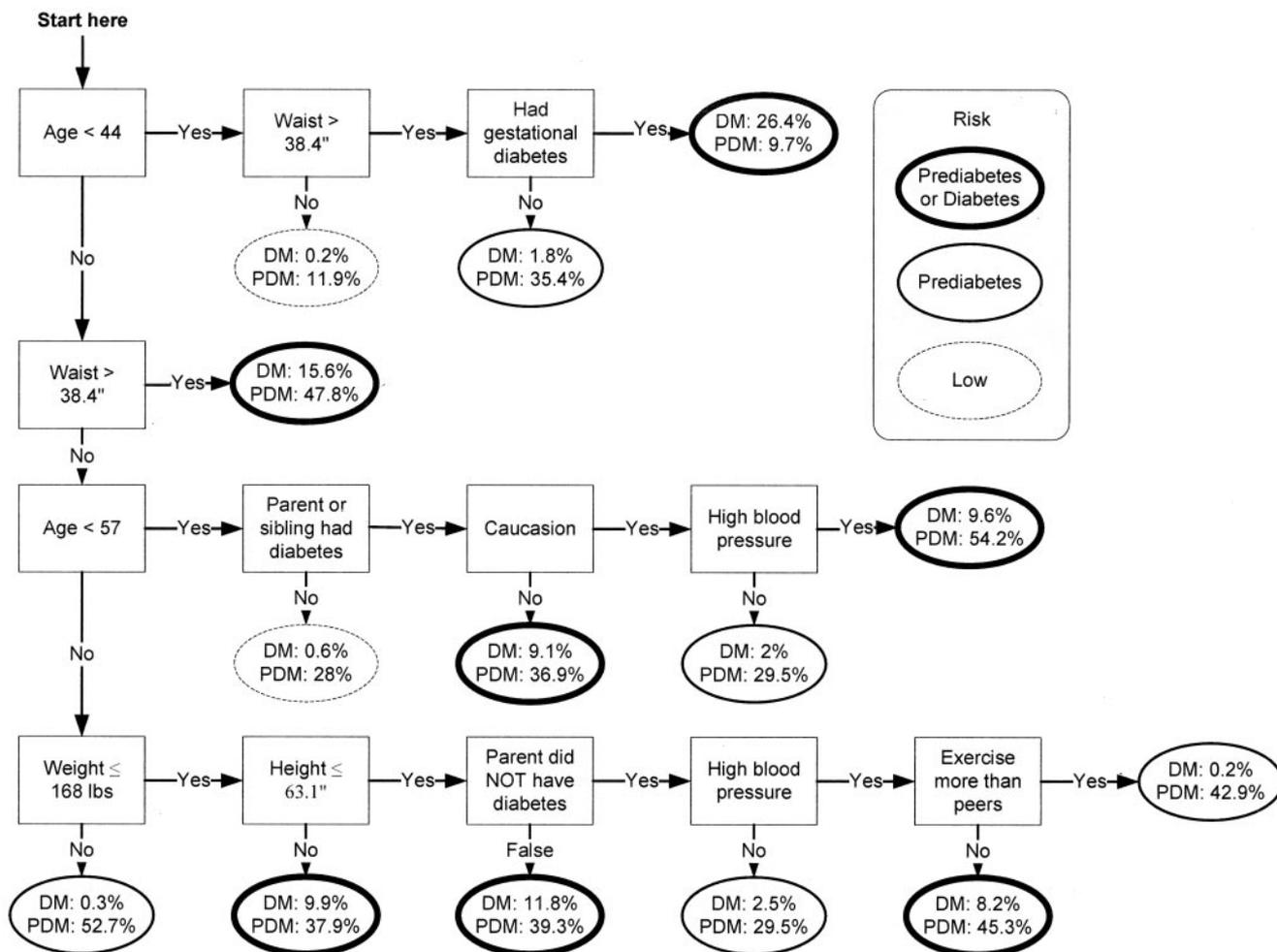


Figure 1—Classification tree for detecting pre-diabetes (PDM) or undiagnosed diabetes (DM).

RESULTS

Prevalence of pre-diabetes and diabetes

The prevalences of undiagnosed diabetes and pre-diabetes in the NHANES III dataset were 4.16 and 26.14%, respectively.

Logistic regression model

Complete results for the logistic regression approaches are in found in the online appendix. The best solution specified the following coefficients for eqs. 1 and 2: intercept, -21.6343; age at interview (years), 0.0402; sex, -0.5042; weight (kilograms), -0.029; height (centimeters), 0.0730; waist-to-hip ratio, 5.3827; BMI (weight in kilograms divided by the square of height in meters), 0.2947; told has high blood pressure, 0-0.3449; and parent has diabetes, 0.3981. Our strategy for including two target conditions in a single tool was to first choose the model and threshold for detecting elevated plasma glucose (either pre-diabetes or

diabetes) that had 80% sensitivity and maximum specificity and then determine a second threshold for the same model that achieved 80% sensitivity for detecting undiagnosed diabetes. This led to setting 0.254 as the threshold probability for saying an individual has elevated plasma glucose (either pre-diabetes or undiagnosed diabetes) and 0.453 as the threshold for saying an individual has undiagnosed diabetes. With use of these cut points, the sensitivity, specificity, and area under receiver operating characteristic (ROC) for detecting elevated plasma glucose were 80%, 64%, and 0.793, respectively. For detecting undiagnosed diabetes the values were 80%, 76.4%, and 0.86, respectively.

Validations were performed using split datasets, in which the model was "trained" on a randomly selected subset of the data and tested on the remaining data. Validation tests were repeated for different selections of training and test data. These tests all produced models that were

very similar to the original and performed nearly as well on test data as on training data.

CART

Detailed results of the CART method can be found in the online appendix. The particular result that best met our objectives is shown in Fig. 1, which is formatted as a screening tool and called the Diabetes Risk Calculator (DRC). It includes questions on age, waist circumference, gestational diabetes, height, race/ethnicity, hypertension, family history, and exercise. The tree begins in upper left corner. An individual moves through the tree in directions determined by answers to questions at each branch, until ending in a terminal node (oval). The probabilities an individual at any terminal node has undiagnosed diabetes or pre-diabetes are shown in the nodes. If thresholds are set for designating an individual as having elevated plasma glucose or pre-diabetes, it is possible to calculate traditional mea-

Table 1—Specificity, sensitivity, positive and negative predictive values, and ROC for the classification tree applied to elevated plasma glucose and undiagnosed diabetes

	Sensitivity (%)	Specificity (%)	PPV	NPV	ROC
Undiagnosed diabetes					
Training	88.16	74.92	0.1369	0.9929	0.8508
v-fold cross-validation	78.22	74.13			0.8219
NHANES 1999–2004	81.02	66.81	0.0627	0.9923	0.7685
Prediabetes or undiagnosed diabetes					
Training	75.36	64.59	0.4940	0.8511	0.7503
NHANES 1999–2004	77.65	51.36	0.4053	0.8433	0.6991

NPV, negative predictive value; PPV, positive predictive value.

asures of accuracy and predictive value. Thus, each terminal node can designate an individual to be at high risk of either 1) diabetes or pre-diabetes (if the probability of undiagnosed diabetes is >8%) or 2) pre-diabetes (if the risk of pre-diabetes is >29% and the risk of undiagnosed diabetes is \leq 2.5%), or 3) neither diabetes or pre-diabetes (if the risk of pre-diabetes is \leq 29% and the risk of undiagnosed diabetes is <1%). These three designations are represented by heavy, medium, and dashed borders around the terminal nodes.

With use of these classifications, the accuracy of the classification tree for undiagnosed diabetes is sensitivity 88%, specificity 75%, positive predictive value 14%, negative predictive value 99.3%, and area under the ROC curve 0.85. The accuracy for pre-diabetes or undiagnosed diabetes is sensitivity 75%, specificity 65%, positive predictive value 49%, negative predictive value 85%, and area under the ROC curve 0.75. On the basis of these results, a positive result on the DRC for undiagnosed diabetes increases the odds that an individual has undiagnosed diabetes by a factor of 3.5, whereas a negative result decreases the odds by a factor of 6, for an 18-fold difference in the odds depending on the results of the test. For increased plasma glucose, the difference in the odds of a positive versus a negative result is a factor of \sim 6.

Validation of classification tree

The classification tree in Fig. 1 was validated in two ways. One was a split-sample validation using data in NHANES III and v-fold cross-validation methods. The other was an external validation using independent data from NHANES 1999–2004. The results are shown in Table 1. The decreases in sensitivity and specificity when applied to test data are typical, and the model appears to be robust. Positive predictive value indicates the fraction of

patients with a positive test having the condition. The positive predictive value for undiagnosed diabetes is much lower for NHANES 1999–2004 than for NHANES III data because the prevalence of undiagnosed diabetes is lower in the NHANES 1999–2004 dataset.

Comparison of logistic regression model and classification tree

The classification tree performed slightly better than the logistic regression model for undiagnosed diabetes in the range of greatest interest and was almost as accurate for detecting elevated plasma glucose. Because it is considerably simpler to apply, requiring no calculations at all, we selected it as the preferred tool for our objectives.

CONCLUSIONS— We have developed a simple tool that uses only questions known to an average individual and requires no calculations to help identify people who are at increased risk for pre-diabetes or undiagnosed diabetes. The DRC sorts people into 14 different categories and reports for each category the probability that an individual is at low risk or high risk for either undiagnosed diabetes or pre-diabetes. To develop the tool we applied two different methods: logistic regression and CART. The versions produced by the two methods had similar accuracies, predictive values, and areas under ROC curves. We selected the tool developed by the CART method because it could be translated into a simpler tool, and it provided information about the actual probabilities that an individual has pre-diabetes or undiagnosed diabetes. The tool developed by the CART method is in the form of a tree that can be easily navigated from the root to terminal nodes through a series of branches, where the path followed depends on the answers to simple yes or no questions that any indi-

vidual would be able to answer. The final terminal node determines the individual's risk of undiagnosed diabetes and/or pre-diabetes. The sensitivity of the DRC was 88 and 75% and the specificity was 75 and 65% for individuals with undiagnosed diabetes and pre-diabetes or undiagnosed diabetes, respectively.

To our knowledge, there are no other tools designed to find people likely to have pre-diabetes as defined by IFG or IGT. Other tools have been developed for detecting people with undiagnosed diabetes (10–17). The sensitivities for undiagnosed diabetes ranged from 72 to 86%, with the highest sensitivity observed in individuals who had one or more cardiovascular risk factors (18). The specificities for the same tools ranged from 41 to 77%. Other tools have been built to calculate the risk of future development of diabetes (10,18,19). One of the tools designed to predict future drug-treated diabetes (10) has been used in people with one or more risk factors for cardiovascular disease to try to identify those who have either undiagnosed diabetes or IGT (11). These tools and applications were all designed for different purposes than was the DRC.

Because one of the objectives of a good screening tool is to minimize the need for unnecessary testing and therefore reduce the economic impact of testing, the predictive value is important to consider for performance of the tool. The positive predictive values of the DRC were 14 and 49% for diabetes and elevated plasma glucose, respectively, and the negative predictive values were 99.3 and 85%, respectively. The positive predictive value for the other screening tools for undiagnosed diabetes ranged from 8 to 13% for non-high-risk populations and was 23% for individuals who have one or more cardiovascular risk factors (11). Thus, in terms of overall performance, the DRC appears to compare favorably with

other available tools for detecting people with undiagnosed diabetes, in addition to its ability to detect people at high risk for pre-diabetes.

Another important distinction of the DRC is that it has been constructed for and tested in a U.S. population. The NHANES III dataset is a weighted survey and includes individuals from different ethnicities as represented in the U.S. population. An analysis showed that a risk score for undiagnosed diabetes developed originally in a strictly Caucasian population could not be applied reliably to other populations with diverse ethnic origins (12).

To our knowledge, the only other tool for undiagnosed diabetes developed with NHANES data was based on an older version of the NHANES (NHANES II) (14). For convenience we will call this the "NHANES II model." When this model was applied to the NHANES II data on which it was developed, its reported sensitivity of 79% and specificity of 65% were lower than the sensitivity and specificity calculated for the DRC applied to NHANES III data on which that model was developed (88 and 75%, respectively). Furthermore, when the NHANES II model is applied to NHANES III data, its sensitivity drops to 71.7% and its specificity decreases to 54.1% (see online appendix).

Models generally perform best on the data on which they were developed, and they perform better on training data than on test data. That the NHANES II model does not perform as well on NHANES III data as the DRC is not unexpected. However, the fact that the DRC performs better than the NHANES II model, when each is tested against the data used to develop it, indicates a significant improvement in predictability for the DRC.

Finally, the DRC has been validated using two methods, 1) split sample cross-validation methods applied to NHANES III data and 2) application of the classification tree to the NHANES 1999–2004 dataset. As expected, the sensitivity, specificity, and predictive values were somewhat lower in the validations than for the training datasets. Nonetheless, the DRC still appears to compare favorably to other tools for detecting pre-diabetes and undiagnosed diabetes. Future research will include validating the tool using an independent dataset from a diabetes prevention clinical trial, as well as determining its applicability to populations outside the U.S. Finally, development of a pa-

tient-friendly, electronic version is underway for broader use in clinical practice.

It is not possible to determine precisely the clinical value of any risk-calculating tool or any diagnostic test for that matter. Their purpose is to provide information that would tip the balance that an individual would choose to undergo more definitive screening with appropriate laboratory tests. For pre-diabetes, it is well documented that treatment can postpone and in some cases prevent the onset of diabetes. It is also well established that treatment of diabetes helps prevent complications. For these reasons, several organizations, such as the American Diabetes Association, recommend screening. Yet a high proportion of people do not receive the recommended screening tests. It is reasonable to assume that some people do not perceive their risk of pre-diabetes or diabetes to be sufficiently high to justify the inconvenience and cost. The DRC we describe in this article is intended to give them a simple method for determining whether they might have a higher risk than they perceive. For undiagnosed diabetes, a positive versus a negative result spreads the odds of having that condition by a factor of 18. For increased plasma glucose, the spread in odds of that condition is by a factor of 6. It seems reasonable to believe that for many people this information may aid in their decision to seek care.

In summary, we have described a simple, validated, paper-based screening tool that can be used to calculate the probability that an individual has either undiagnosed diabetes or pre-diabetes using information known to an average individual, without requiring any calculations. The screening tool can be used by physicians to assess the risks of their patients or can be self-administered by individuals to assess their own risks. Use of this tool enables the identification of individuals who might benefit from confirmatory tests and treatment to delay or prevent the onset of type 2 diabetes and its complications.

References

1. Wild S, Roglic G, Green A, Sicree R, King H: Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 27:1047–1053, 2004
2. Cowie CC, Rust KF, Byrd-Holt DD, Eberhardt MS, Flegal KM, Engelgau MM, Saydah SH, Williams DE, Geiss LS, Gregg EW: Prevalence of diabetes and impaired fasting glucose in adults in the U.S. pop-

- ulation: National Health and Nutrition Examination Survey 1999–2002. *Diabetes Care* 29:1263–1268, 2006
3. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346:393–403, 2002
4. Eddy DM, Schlessinger L, Kahn R: Clinical outcomes and cost-effectiveness of strategies for managing people at high risk for diabetes. *Ann Intern Med* 143:251–264, 2005
5. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, Keinanen-Kiukkaanniemi S, Laakso M, Louheranta A, Rastas M, Salminen V, Uusitupa M: Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 344:1343–1350, 2001
6. Herman WH, Hoerger TJ, Brandle M, Hicks K, Sorensen S, Zhang P, Hamman RF, Ackermann RT, Engelgau MM, Ratner RE: The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. *Ann Intern Med* 142:323–332, 2005
7. American Diabetes Association: Screening for type 2 diabetes (Position Statement). *Diabetes Care* 26 (Suppl. 1):S21–S24, 2003
8. Introduction to NHANES. Available from http://www.cdc.gov/nchs/about/major/nhanes/intro_mec.htm. Accessed 15 January 2007
9. Breiman L, Friedman J, Olshen RA, Stone CJ: *Classification and Regression Trees*. Boca Raton, FL, Chapman & Hall/CRC, 1998
10. Lindström J, Tuomilehto J: The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 26:725–731, 2003
11. Franciosi M, De Berardis G, Rossi MC, Sacco M, Belfiglio M, Pellegrini F, Tognoni G, Valentini M, Nicolucci A: Use of the diabetes risk score for opportunistic screening of undiagnosed diabetes and impaired glucose tolerance: the IGLOO (Impaired Glucose Tolerance and Long-Term Outcomes Observational) study. *Diabetes Care* 28:1187–1194, 2005
12. Glumer C, Carstensen B, Sandbaek A, Lauritzen T, Jorgensen T, Borch-Johnsen K: A Danish diabetes risk score for targeted screening: the Inter99 study. *Diabetes Care* 27:727–733, 2004
13. Baan CA, Ruige JB, Stolk RP, Witteman JC, Dekker JM, Heine RJ, Feskens EJ: Performance of a predictive model to identify undiagnosed diabetes in a health care setting. *Diabetes Care* 22:213–219, 1999
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* 18:382–387, 1995
15. Ruige JB, Neeling JN, Kostense PJ, Bouter LM, Heine RJ: Performance of an NIDDM screening questionnaire based on symptoms and risk factors. *Diabetes Care* 20: 491–496, 1997
 16. 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* 16:164–171, 2000
 17. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M, Pfeiffer AFH, Spranger J, Thamer C, Häring H-U, Fritsche A, Joost HG: An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 30:510–515, 2007
 18. Stern MP, Williams K, Haffner SM: Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann Intern Med* 136: 575–581, 2002
 19. Simmons RK, Harding A-H, Wareham NJ, Griffin SJ, EPIC-Norfolk Project Team: Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes? *Diabet Med* 24: 830–835, 2007