DIABETES RISK CALCULATOR: A Simple Tool for Detecting Undiagnosed Diabetes and Prediabetes

Kenneth E. Heikes PhD, Archimedes, Inc.
David M. Eddy MD PhD, Archimedes, Inc.
Bhakti Arondekar MBA PhD, GlaxoSmithKline,
Leonard Schlessinger PhD, Archimedes, Inc.

Running title: Simple tool for pre- and undiagnosed diabetes

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

ABSTRACT

Objective: To develop a simple tool for the US population to calculate the probability that a person has either undiagnosed diabetes or prediabetes.

Research Design and Methods: We used data from NHANES III and two methods -- logistic regression, and classification tree analysis -- to build two models. We selected the classification tree model based on its equivalent accuracy but greater ease of use.

Results: The resulting tool, called DIABETES RISK CALCULATOR, includes questions on age, waist circumference, gestational diabetes, height, race/ethnicity, hypertension, family history, and exercise. Each terminal node specifies a person's probabilities of prediabetes or of undiagnosed diabetes. Terminal nodes can also be used categorically to designate a person as having high risk for (1) undiagnosed diabetes or prediabetes or (2) prediabetes or (3) neither undiagnosed diabetes or prediabetes. Using these classifications, the sensitivity, specificity, positive and negative predictive values, and ROC area for detecting undiagnosed diabetes are 88%, 75%, 14%, 99.3%, and 0.85, respectively. For prediabetes 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: DIABETES RISK CALCULATOR is the only currently available noninvasive screening tool designed and validated to detect both prediabetes and undiagnosed diabetes in the US population.

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 prediabetes or undiagnosed diabetes and to help people decide if 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 person 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 US in 2002 the prevalence of diabetes was estimated to be 19.3 million, of which about 5.8 million were undiagnosed (2). An additional 41 million are estimated to have prediabetes, defined as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT). Prediabetes implies an increased risk of developing type 2 diabetes (T2DM) on the order of 30% over 4 years (3) and 70% over 30 years (4). Several studies have demonstrated that T2DM can be prevented or delayed with the use of life-style modification or pharmacotherapy in subjects with prediabetes (3, 5). Studies have also indicated that preventing or delaying the onset of T2DM using life-style modification or pharmacotherapy can be cost-effective (6), if costs of the interventions are controlled (4).

An important step in preventing or delaying T2DM and its complications is to identify people with prediabetes and undiagnosed diabetes so that they can be given appropriate care. The American Diabetes Association (ADA) recommends screening for T2DM at 3-year intervals beginning at age 45, particularly in those with BMI $\geq$ 25 kg/m^2 (7). However these recommendations are not widely followed, as indicated by the fact that 30% of people who have diabetes are still undiagnosed. Major reasons for this 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 prediabetes 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-US populations and none are designed to detect prediabetes and undiagnosed diabetes. The objective of this study was to develop a simple tool for use in the US to identify people who have a high probability of having prediabetes or undiagnosed diabetes, using only information that is commonly known to an average person, and preferably not requiring any calculations.

RESEARCH DESIGN AND METHODS
Definitions. The definitions of prediabetes and diabetes are based on fasting plasma glucose (FPG) and glucose tolerance (GT), as measured by a 2-hour plasma oral glucose tolerance test (2hr-OGTT). Impaired fasting glucose (IFG) is defined as FPG 100-125 mg/dl. Impaired glucose tolerance (IGT) is defined as 2hr-OGTT 140-199 mg/dl. Diabetes is defined as FPG $\geq$ 126 mg/dl and/or 2hr-OGTT $\geq$ 200 mg/dl. Prediabetes is defined as IFG and/or IGT without diabetes. Undiagnosed diabetes is defined as the presence of actual diabetes based on FPG and/or 2hr-OGTT and the absence of a person having been told they have diabetes. We use the term "elevated plasma glucose" to define a

person who has either prediabetes or undiagnosed diabetes.

Data. We used data from NHANES III (1988-94) to build and internally validate the tool (8). This dataset was chosen for two main reasons. First it is a representative sample of the US population, which is the main population for which the tool is being designed. The survey methods over sampled certain subpopulations, such as race/ethnicity minorities, but provided weights to enable construction of a representative sample of the US. Second the NHANES III dataset is the most current NHANES survey that contains 2hr-OGTT results on a large sub-sample 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 7092 participants who were age ≥20 and had FPG results. Two hour-OGTT data were available for approximately half of those ages 40-75. For people for whom 2hr-OGTT results were missing, the diagnoses were based on FPG alone. An analysis of the group for whom both FPG and OGT data were available revealed that the lack of OGT 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 prediabetes and undiagnosed diabetes by approximately 2% and 1.5% respectively. Additional details about the data collection and analysis are described in a technical report (9).

Explanatory variables. To build the tool we examined 18 explanatory variables that would be known to the average person and would not require laboratory results or special medical definitions. These included body mass index, height, weight, waist circumference, waist-to-hip ratio, age, gender, race/ethnicity, taking blood pressure medication, taking cholesterol medication, had 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 to peers. Not all variables were used in the final tool; their inclusion in the final models depended on their value as predictors of prediabetes 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, where the probability *p* of prediabetes or undiagnosed diabetes was modeled by a logistic, or log-odds, transformation

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

where the x_i are the continuous or dichotomous explanatory variables, the β_i 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

$$p = \left(1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}\right)^{-1} \qquad (2)$$

Logistic regressions were generated with SAS 9.1 (10).

Analytical methods: Classification and Regression Tree (CART). We also built a model using a classification and regression tree method (CART) (11). 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 prediabetes) 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 based on 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 prediabetes 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 prediabetes. Since one of the goals was to create a simple model, body mass index (BMI) and waist-to-hip ratio (WHR) were dropped from the list of variables in favor of weight (WGT),

height (HGT), and waist circumference (WST), which required no calculation but still maintained the accuracy of the tool. We eliminated the cholesterol variables high cholesterol (HBC) and taking cholesterol medication (BCM) because of the large number of missing fields and low predictive value. We also eliminated history of diabetes in any blood relative (HST) in favor of the more specific diabetes history variables – history of diabetes in a parent or sibling (HDM), a parent only (HDP), or sibling only (HDS).

We used v-fold cross-validation to train and test the classification tree models, partitioning the data into equal-sized subsets (12). We then derived and tested the classification tree on all combinations of 9/10 training data and 1/10 test data.

RESULTS
Prevalence of prediabetes and diabetes. The prevalences of undiagnosed diabetes and prediabetes in the NHANES III dataset were 4.16% and 26.14%, respectively.

Logistic regression model. Complete results for the logistic regression approaches are in the Technical Report (9). The best solution specified the following coefficients for Equations (1) and (2): intercept, -21.6343; age at interview, 0.0402; gender, -0.5042; weight (kg), -0.029; Height (cm), 0.0730; waist/hip ratio, 5.3827; BMI, 0.2947; told has high BP, 0-.3449; 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 prediabetes 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 a person has elevated plasma glucose (either prediabetes or undiagnosed diabetes) and 0.453 as the threshold for saying a person has undiagnosed diabetes. Using these cut points, the sensitivity specificity and area under 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 conducted using split datasets, where 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.

Classification and regression tree (CART). Detailed results of the CART method can be found in the Technical Report (9). The particular result that best met our objectives is shown in Figure 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. A person moves through the tree in directions determined by answers to questions at each branch, until ending in a terminal node (oval). The probabilities a person at any terminal node has undiagnosed diabetes (DM) or prediabetes (PDM) are shown in the nodes. If thresholds are set for designating a person as having elevated plasma glucose or prediabetes, it is possible to calculate traditional measures

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

Using 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 prediabetes or undiagnosed diabetes is: sensitivity = 75%, specificity = 65%, positive predictive value = 49%, negative predictive value = 85%, and area under the ROC curve = 0.75. Based on these results, a positive result on the DRC for undiagnosed diabetes increases the odds a person 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 test's results. For increased plasma glucose, the difference in the odds of a positive versus negative result is a factor of about 6.

Validation of classification tree. The classification tree in Figure 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 decrease in sensitivity and specificity when applied to test data is

typical and the model appears to be robust. Positive predictive value (PPV) indicates the fraction of patients with a positive test having the condition. The PPV 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 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 to help identify people who are at increased risk for prediabetes or undiagnosed diabetes that uses only questions known to an average person and requires no calculations. DRC sorts people into fourteen different categories and reports for each category the probability that a person is at low-risk or high-risk for either undiagnosed diabetes or prediabetes. To develop the tool we applied two different methods: logistic regression and classification tree analysis (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 a person has prediabetes 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/no questions that any person would be able to answer. The final terminal node determines the person's risk of undiagnosed diabetes and/or prediabetes. The sensitivity of the DRC was 88% and 75%, and the specificity was 75% and 65% for individuals with "undiagnosed diabetes" and "prediabetes or undiagnosed diabetes", respectively.

To our knowledge there are no other tools designed to find people likely to have prediabetes as defined by "IFG or IGT". Other tools have been developed for detecting people with undiagnosed diabetes (13, 14, 15, 16, 17, 18, 19, 20). The sensitivities for undiagnosed diabetes ranged from 72%-86%, with the highest sensitivity observed in individuals who had one or more cardiovascular risk factors (14). The specificities for the same tools ranged from 41%-77%. Other tools have been built to calculate the risk of future development of diabetes (13, 21, 22). One of the tools designed to predicting future drug-treated diabetes (13) has been used in people with one or more risk factor for cardiovascular disease to try to identify those who have "either undiagnosed diabetes or IGT" (14). These tools and applications were all designed for different purposes than 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 DRC were 14% and 49% for diabetes and elevated plasma glucose respectively, and the negative predictive values were 99.3% and 85%. The positive predictive

value for the other screening tools for undiagnosed diabetes ranged from 8%-13% for non-high risk populations, and 23% for individuals who have one or more cardiovascular risk factors (14). Thus in terms of overall performance, the DRC appears to compare favorably to other available tools for detecting people with undiagnosed diabetes, in addition to its ability to detect people at high risk of prediabetes.

Another important distinction of DRC is that it has been constructed for and tested in a US population. The NHANES III dataset is a weighted survey and includes individuals from different ethnicities as represented in the U.S. population. An analysis found 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 (15).

To our knowledge the only other tool for undiagnosed diabetes developed using NHANES data was based on an older version of the NHANES (NHANES II) (17). 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% (9).

Models generally perform best on the data on which they are developed, and they perform better on training data than test data. That the NHANES II model does not perform as well on NHANES III data as 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) applying 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 prediabetes 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 US. Finally, development of a patient-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 a person would choose to undergo more definitive screening with appropriate lab tests. In the case of prediabetes 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 ADA 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 prediabetes or diabetes to be sufficiently high to justify the inconvenience and cost. The DRC we

describe in this paper is intended to give them a simple method for determining if 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 conclusion, we have described a simple, validated, paper-based screening tool that can calculate the probability that an individual has either undiagnosed diabetes or prediabetes using information known to an average person, 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 people to assess their own risks. Use of this tool enables the identification of people who might benefit from confirmatory tests and treatment to delay or prevent the onset of T2DM 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. population: 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-Kiukaanniemi 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.   Screening for type 2 diabetes. *Diabetes Care* 26 Suppl 1:S21-S24, 2003

8.   Introduction   to   NHANES. http://www.cdc.gov/nchs/about/major/nhanes/intro_mec.htm

9.   Heikes, K and Eddy DM. A Simple tool for detecting unrecognized prediabetes and diabetes. Technical Report. Archimedes Inc. October 30, 2006. Available as an online appendix .

10. SAS Institute Inc., Cary, NC, USA

11. All classification trees were created using CART software v5.0, Salford Systems, San Diego, CA 92123

12. Breiman L, Friedman J, Olshen RA, Stone CJ. *Classification And Regression Trees.* 1998, Chapman & Hall / CRC, Boca Raton, FL 33431

13. Lindström J, Tuomilehto J: The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 26:725-731, 2003

14. 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

15. 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

16. 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

17. 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

18. 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

19. Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of Type 2 diabetes in general practice. *Diabetes/Metabolism Res Rev* 16:164-171, 2000

20. 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

21. 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? *Annals of Internal Medicine* 136:575-581, 2002

22. Simmons RK, Harding A-H, WarehamNJ, Griffin SJ on behalf of the EPIC-Norfolk project team. Do simple questions about diet and physical activity help to identify those a risk of Type 2 diabetes? *Diabetic Medicine* 24: 830-835, 2007

TABLE 1. Specificity, sensitivity, positive and negative predictive values (PPV, NPV) and ROC for classification tree in 2 applied to elevated plasma glucose and undiagnosed diabetes.

Variable	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

FIGURE 1. Classification tree for detecting prediabetes or undiagnosed diabetes.