A New and Simple Questionnaire to Identify People at Increased Risk for Undiagnosed Diabetes

WILLIAM H. HERMAN, MD, MPH
PHILIP J. SMITH, PHD
THEODORE J. THOMPSON, MS
MICHAEL M. ENGELGAU, MD
RONALD E. AUBERT, PHD

OBJECTIVE — To develop a simple questionnaire to prospectively identify individuals at increased risk for undiagnosed diabetes.

RESEARCH DESIGN AND METHODS — People with newly diagnosed diabetes (n = 164) identified in the Second National Health and Nutrition Examination Survey and those with neither newly diagnosed diabetes nor a history of physician-diagnosed diabetes (n = 3,220) were studied. Major historical risk factors for undiagnosed non-insulin-dependent diabetes were defined, and classification trees were developed to identify people at higher risk for previously undiagnosed diabetes. The sensitivity, specificity, and predictive value of the classification trees were described and compared with those of an existing questionnaire.

RESULTS — The selected classification tree incorporated age, sex, history of delivery of a macrosomic infant, obesity, sedentary lifestyle, and family history of diabetes. In a representative sample of the U.S. population, the sensitivity of the tree was 79%, the specificity was 65%, and the predictive value positive was 10%.

CONCLUSIONS — This classification tree performed significantly better than an existing questionnaire and should serve as a simple, noninvasive, and potentially cost-effective tool for diagnosing diabetes in the U.S.

From the Epidemiology and Statistics Branch, Division of Diabetes Translation, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia.

Address correspondence and reprint requests to William H. Herman, MD, MPH, Epidemiology and Statistics Branch, Division of Diabetes Translation (K-10), National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, 4770 Buford Hwy., NE, Atlanta, GA 30341-3724.

Received for publication 18 July 1994 and accepted in revised form 17 November 1994.

ADA, American Diabetes Association; NHANES II, Second National Health and Nutrition Examination Survey; NIDDM, non-insulin-dependent diabetes mellitus; ROC, receiver-operator characteristic.

Worldwide, at least 20% of diabetes is undiagnosed, and in many communities >50% is undiagnosed (1). Although the proportion of undiagnosed diabetes has been reported to be low in several industrialized urban communities and in some traditional rural communities (1), the proportion of diabetes that is undiagnosed is high in many parts of the U.S. population, including non-Hispanic whites (50%), blacks (44%), and Hispanics (42%) (2). In the U.S., an estimated 4.5 million non-Hispanic whites, 1.0 million African-Americans, and 800,000 Hispanics have diabetes and do not know it (2).

The onset of non-insulin-dependent diabetes mellitus (NIDDM) may occur 9–12 years before its clinical diagnosis (3). As a result, undiagnosed diabetes is associated with substantial morbidity. Fifteen to 20% of Americans with undiagnosed diabetes have diabetic retinopathy (3), 8% have nephropathy (4), and 9% have clinical neuropathy (2). Cardiovascular risk factors are also prevalent. In the U.S., 61% of people 40–69 years of age who have undiagnosed diabetes have hypertension, 49% have total cholesterol levels >240 mg/dl, and 32% smoke cigarettes (2).

Early detection and treatment can reduce the burden of the complications of diabetes. The Diabetes Control and Complications Trial conclusively demonstrated that near-normal blood glucose control can prevent the development and slow the progression of the microvascular complications of insulin-dependent diabetes (5). Timely laser therapy can prevent visual loss from both macular edema (6) and proliferative retinopathy (7), angiotensin-converting enzyme inhibitors can significantly slow the progression of diabetic nephropathy (8), and foot care can prevent amputations in individuals with insensitive feet (9). Likewise, improved detection and treatment of cardiovascular risk factors can prevent cardiovascular disease (10).

Recognition of the public health

burden of undiagnosed diabetes and demonstration of the effectiveness of timely intervention have resulted in a change in attitude toward community screening for undiagnosed diabetes (11). In the late 1970s, a workshop on such screening was sponsored by the American Diabetes Association (ADA), the Centers for Disease Control, and the National Institutes of Health. The participants concluded that, except among pregnant women, community screening used to detect asymptomatic glucose intolerance could not be recommended because the evidence of benefit did not outweigh the deleterious effects and did not justify the expenditure of community resources (12). In 1989, the ADA changed its position on community screening to recommend that all people with one or more diabetes risk factors or having any of the symptoms of diabetes be identified and referred for medical evaluation (13). In 1993, the British Diabetic Association recommended screening for diabetes among individuals 40–75 years of age (14).

Although it is recognized that undiagnosed diabetes is prevalent and serious and that early treatment is sufficiently beneficial to make early detection appropriate, no screening test has been shown to be simple, inexpensive, and acceptable and to provide good sensitivity, specificity, and predictive value (14). Although the ADA has developed a risk-factor questionnaire to increase public awareness of diabetes and assess risk (15), the sensitivity, specificity, and predictive value of the questionnaire were not prospectively evaluated during its development (Richard Kahn, personal communication). Moreover, studies that have independently assessed using the existence of one or more risk factors (16) or the ADA questionnaire (17) as guides toward predicting risk of undiagnosed diabetes have not found these tools to be effective.

In this report, we have used information about diabetes risk factors (18) and knowledge of diabetes status from a representative sample of the U.S. population with no history of diabetes to develop and test a simple questionnaire to prospectively identify people at increased risk for undiagnosed diabetes.

RESEARCH DESIGN AND METHODS

— We analyzed data from the Second National Health and Nutrition Examination Survey (NHANES II). Through household interviews, this survey obtained demographic and medical information from a representative probability sample of the U.S. population (n = 15,357) (19). A subsample of respondents 20–74 years of age who reported having no history of physician-diagnosed diabetes were administered a 75-g oral glucose load on the morning after an overnight fast (n = 3,770). Subsequent diagnosis of diabetes was based on World Health Organization criteria (20).

The prevalence of undiagnosed diabetes and risk factors for undiagnosed diabetes in the U.S. population that is 20–74 years of age have been reported from NHANES II (2,21,22). The weighted prevalence of undiagnosed diabetes is 3.4% (21). The prevalence increases with age, tends to be higher among women than among men (3.8 vs. 2.9%), and is significantly higher among blacks than among whites (5.2 vs. 3.2%) (21). Obesity and parental history of diabetes are associated with a higher prevalence of undiagnosed diabetes (21).

We limited our study to the subsample of 164 people who tested positive and 3,220 people who tested negative for diabetes. Variables selected for study were defined in the NHANES II as follows: age (20–44, 45–64, and 65+ years); sex of participant and history of delivery of a macrosomic infant (male, female, female with history of delivering an infant >9 lbs); race (white, black, Hispanic, Native American); education (not a high school graduate, high school graduate or education after high school); obesity (self-reported weight [without clothes and shoes] for height ≥120% ideal body weight for medium frame) (23); sedentary lifestyle (little or no exercise during recreation and quite inactive during a usual day); general health assessment (excellent, good, fair, poor); history of physician-diagnosed hypertension; history of borderline or potential diabetes or prediabetes; and history of diabetes in either parents or siblings.

To identify groups at higher risk for undiagnosed diabetes, we used the tree function available in the S-Plus software package (Statistical Sciences, Seattle, WA) to construct classification trees. Classification trees are an exploratory method for identifying factors and interactions among factors that may explain variation in a binary outcome (24,25). Classification trees are constructed using recursive partitioning. At each node of the classification tree, the recursive partitioning algorithm identifies a predictor variable and a "split" by which cases may be subclassified. This predictor variable and split combination is chosen to have the greatest predictive power among all predictor-split combinations at the tree node. Once the cases at the node have been partitioned by the split, the algorithm is applied to both of the resulting subclassifications. This procedure is applied recursively until the tree has been grown to a specified number of terminal "leaves."

We chose tree sizes for their simplicity and agreement with current knowledge of diabetes risk factors (18) while ensuring that the number of cases in the terminal leaves was large enough to protect against being misled by vagaries of the data. Although the NHANES II data set included 3,384 cases with information on the variables of interest, there were only 164 people with undiagnosed diabetes to provide information about the discriminatory power of the predictors. One classification tree assessed basic demographic and historical variables (age, sex, history of delivery of macrosomic infant, race, obesity, sedentary lifestyle, and family history of diabetes). The other assessed these variables and history of hypertension and glucose intolerance.

Sensitivity was defined as the

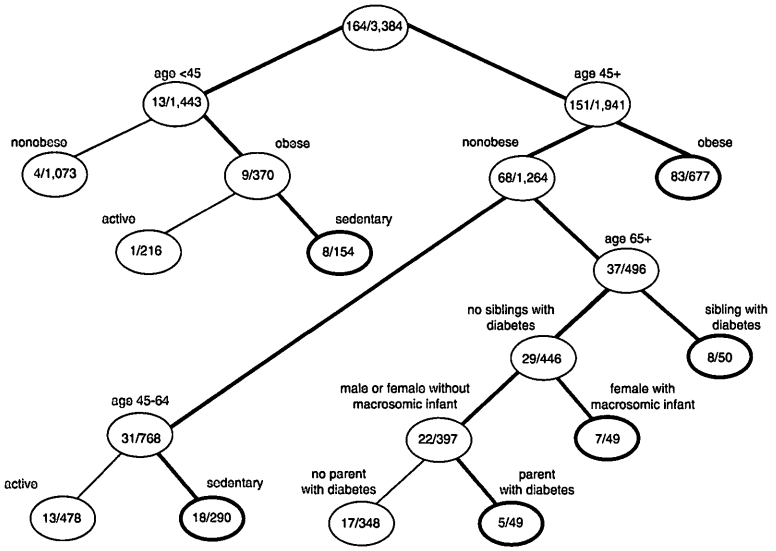

Figure 1—*Classification tree of risk factors for undiagnosed diabetes incorporating age, sex, history of macrosomic infant, race, obesity, sedentary lifestyle, and family history of diabetes among individuals not previously diagnosed with diabetes. For each risk factor profile, the numerator represents the number of survey participants with newly diagnosed diabetes and the denominator represents the total number with the risk factor profile. The dark ovals represent groups with a prevalence of undiagnosed diabetes >5%. For definitions of terms, see* METHODS.

probability of having a positive screening test given the presence of diabetes as defined by the oral glucose tolerance test. Specificity was defined as the probability of having a negative screening test given the absence of diabetes. The predictive value of a positive test was defined as the proportion of diabetic individuals among those who had positive screening tests.

To compare the performance of these two classification trees and the historical risk factors used in the ADA questionnaire (age, race, obesity, history of delivery of macrosomic infant(s), and family history of diabetes) we used the classification trees, the ADA questionnaire, and the data from the NHANES II, and we constructed receiver-operator characteristic (ROC) curves. These curves characterize the relationship between the true-positive ratio (sensitivity) and the false-positive ratio (1 − specificity). The points on the curves for the classification trees represent terminal nodes, and those on the curve for the ADA questionnaire represent the sum of the weighted scores assigned for each risk factor. Screening tests that discriminate well between diseased and living individuals aggregate toward the upper left corner of an ROC curve. The area under the curve quantifies how well the screening test correctly distinguishes a diseased from a living person; the greater the area under the curve, the better the performance of the screening test. Estimates of the percentage identified as being at higher risk were weighted to represent the U.S. population.

RESULTS — The classification tree incorporating demographic and selected historical variables (Fig. 1) shows that individuals at higher risk for undiagnosed diabetes are those ≥45 years of age who are obese (≥120% ideal body weight) (prevalence is 83 of 677, or 12%); people ≥65 years of age who are not obese but have siblings with diabetes (8 of 50; 16%), a history of delivering a macrosomic infant (7 of 49; 14%), or parents with diabetes (5 of 49; 10%); people 45–64 years of age who are not obese but are sedentary (18 of 290; 6%); and those who are younger than 45 and are obese and sedentary (8 of 154; 5%). This classification tree identifies a high-risk group of 1,269 people from a population of 3,384. Using weighted data, follow-up testing to establish a definitive diagnosis of diabetes would be required for only 31% of the total population. The high-risk population identified with this classification tree includes 129 of the 164 people with undiagnosed diabetes; the sensitivity of the classification tree is thus 79%. The specificity is 65% (2,080 of 3,220), and the predictive value positive is 10% (129 of 1,269).

The classification tree incorporating the same demographic and historical variables and also including history of hypertension and history of glucose intolerance (Fig. 2) shows that individuals at higher risk for undiagnosed diabetes are those ≥45 years of age who are obese (83 of 676; 12%); people ≥45 who are not obese but have a history of physician-diagnosed hypertension (40 of 407; 10%); people ≥45 who are not obese and do not have a history of physician-diagnosed hypertension but have a history of borderline or potential diabetes or prediabetes (5 of 40; 13%); and those who are younger than 45 and are obese and sedentary (8 of 154; 5%). This classification tree identifies a high-risk group of 1,277 individuals from a population of 3,380 (4 individuals did not respond to the NHANES II questions about hypertension or glucose intolerance). Using weighted data, follow-up testing to establish a definitive diagnosis of diabetes would thus be required for only 30% of the total population. The high-risk population identified includes 136 of the 164 people with undiagnosed diabetes; the sensitivity of this screening strategy is thus 83%. The specificity is 65% (2,075 of 3,216), and the predictive value positive is 11% (136 of 1,277).

The second decision tree performed the same as the first as judged by areas under the ROC curves (Fig. 3) (area under both curves = 0.78), and both performed significantly better ($P < 0.001$)

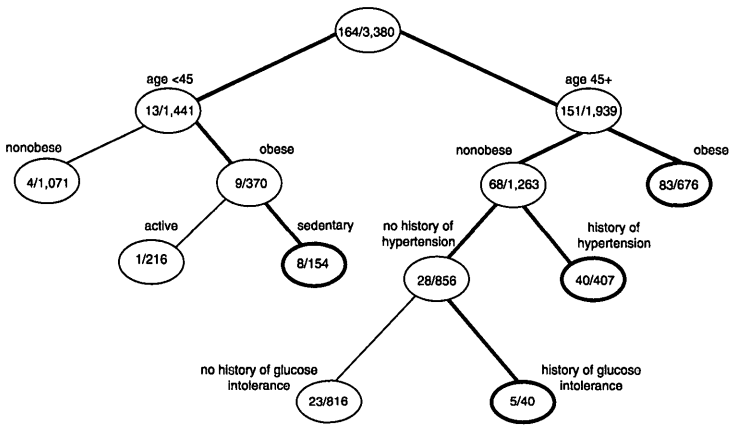

Figure 2—*Classification tree of risk factors for undiagnosed diabetes incorporating age, sex, history of macrosomic infant, race, obesity, sedentary lifestyle, family history of diabetes, history of hypertension, and history of glucose intolerance among individuals not previously diagnosed with diabetes. For each risk factor, the numerator represents the number of survey participants with newly diagnosed diabetes and the denominator represents the total number with the risk factor profile. The dark ovals represent groups with a prevalence of undiagnosed diabetes >5%. For definitions of terms, see* METHODS.

than the ADA questionnaire's historical risk factors (area under the curve = 0.71).

CONCLUSIONS— Screening is indicated for the detection of disease in asymptomatic, apparently healthy individuals if the burden of suffering caused by the disease is large, the natural history of the disease process is understood, effective treatment exists, early treatment is more effective than later treatment, and the screening procedure is good. With the development of a good screening procedure, these criteria would be met for diabetes.

Our results show that information about risk factors can identify groups at higher risk for undiagnosed diabetes and that such an approach has good sensitivity and specificity and reduces the number of subjects who require definitive testing. Because the first classification tree (Fig. 1) performed as well as the second (Fig. 2) and did not depend on respondents' having had prior medical evaluation, we recommend using the first classification tree.

Race and a number of other important risk factors for undiagnosed diabetes did not enter into the classification trees. This does not mean that race or other risk factors have no role in predicting undiagnosed diabetes. Rather, it means that after adjusting for the risk factors in the classification trees, other risk factors have diminished predictive power. When applied to the racial- and ethnic-minority subpopulations included in NHANES II, the classification tree in Fig. 1 performed as well as it did among whites. Among blacks, Hispanics, and Native Americans, sensitivity was 80% and specificity was 61%. Among whites, sensitivity was 78% and specificity 65%.

The tree in Fig. 1 can be formatted as a simple questionnaire for community screening (Table 1). Like the ADA questionnaire, it can be adapted to educate the public and increase awareness of the risks of diabetes (15).

Studies that have independently assessed using the existence of one or more diabetes risk factors or the complete ADA questionnaire to predict the risk of undiagnosed diabetes have reported sensitivity to be 59–69% and specificity to be 34–46% (16,17). In our analysis, the classification tree in Fig. 1 performed signifi-

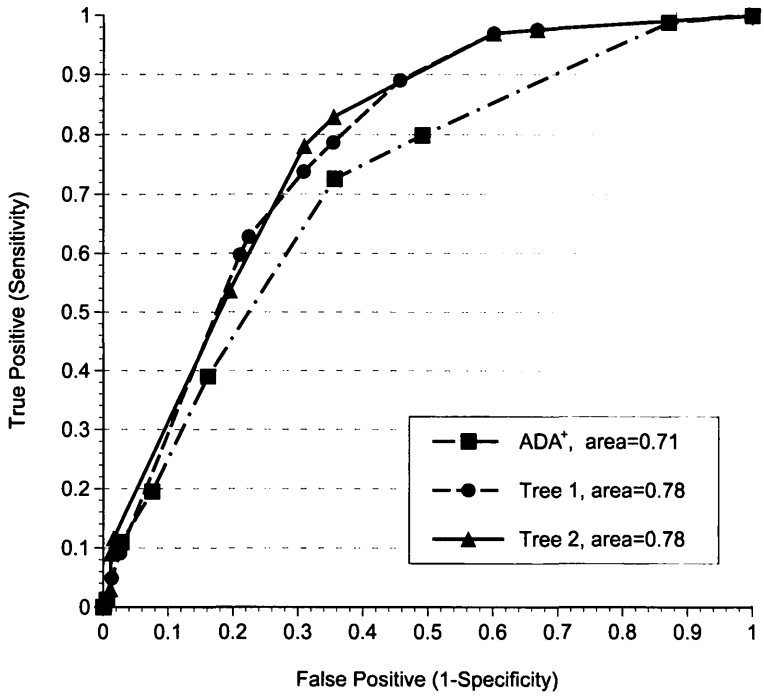

Figure 3—*ROC curves showing performance of three screening questionnaires. ADA+, the ADA questionnaire (15) using historical risk factors but not symptoms (see* METHODS.

Table 1—*A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes*

Are you at risk for diabetes?

Select the column that corresponds to your current age and answer the questions.

I am 20–44 years of age:	I am 45–64 years of age:	I am 65+ years of age:
my weight is equal to or above that listed in the chart at the right*	my weight is equal to or above that listed in the chart at the right*	my weight is equal to or above that listed in the chart at the right*
____ yes	____ yes	____ yes
I get little or no exercise during a usual day	I get little or no exercise during a usual day	my mother, father, sister, or brother had diabetes
____ yes	____ yes	____ yes
		I delivered a baby > 9 lbs
		____ yes
if you answered yes to <u>both</u> questions, you are at risk!	if you answered yes to <u>either</u> question, you are at risk!	if you answered yes to <u>any</u> question, you are at risk!

Women			Men		
Height		Weight	Height		Weight
Feet	Inches	Pounds	Feet	Inches	Pounds
4	9	134	5	1	157
4	10	137	5	2	160
4	11	140	5	3	162
5	0	143	5	4	165
5	1	146	5	5	168
5	2	150	5	6	172
5	3	154	5	7	175
5	4	157	5	8	179
5	5	161	5	9	182
5	6	164	5	10	186
5	7	168	5	11	190
5	8	172	6	0	194
5	9	175	6	1	199
5	10	179	6	2	203
5	11	182	6	3	209

Height is without shoes and weight is without clothes.
*These charts show weights that are 20% heavier than what is recommended for women and men with a medium frame. If your weight is above the amount shown for your height, you may be at risk for developing diabetes.

cantly better than the ADA questionnaire when clinical symptoms were not included. We could not use NHANES II to rigorously evaluate the performance of the ADA questionnaire, however, because NHANES II did not ask about clinical symptoms of undiagnosed diabetes. On the other hand, the symptoms of diabetes—such as fatigue, weight loss, thirst, frequent urination, and blurry vision—are difficult to define, quantify, and assess in self-administered questionnaires and add little to a questionnaire's ability to identify those at higher risk for undiagnosed diabetes (17).

Our proposed questionnaire may prove useful as part of a sequential screening strategy that also uses biochemical testing to determine diabetes risk. Random capillary blood glucose testing has been shown to be 73% sensitive and 95% specific in detecting early NIDDM (26). Our suggested questionnaire will render a general population to a smaller group that will have a higher prevalence of diabetes. Subsequent application of biochemical screening tests will be more efficient in this population and will reduce the need for definitive diagnostic testing. The trade-off is that ~20% of individuals with undiagnosed diabetes will be missed by the initial screening questionnaire. Further studies combining this questionnaire with random capillary blood glucose testing are needed.

References

1. King H, Rewers M, WHO Ad Hoc Diabetes Reporting Group: Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. *Diabetes Care* 16:157–177, 1993
2. Harris MI: Undiagnosed NIDDM: clinical and public health issues. *Diabetes Care* 16:642–652, 1993
3. Harris MI, Klein R, Welborn TA, Knuiman MW: Onset of NIDDM occurs at least 4–7 yr before clinical diagnosis. *Diabetes Care* 15:815–819, 1992
4. Ballard DJ, Humphrey LL, Melton LJ, Frohnert PP, Chu PC, O'Fallon WM, Palumbo PJ: Epidemiology of persistent proteinuria in type II diabetes mellitus: population-based study in Rochester, Minnesota. *Diabetes* 37:405–412, 1988
5. The Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993
6. Early Treatment of Diabetic Retinopathy Study Research Group: Early photocoagulation for diabetic retinopathy (ETDRS report number 9). *Ophthalmology* 98: 766–785, 1991
7. Diabetic Retinopathy Study Research Group: Photocoagulation treatment of proliferative diabetic retinopathy: clinical

application of Diabetic Retinopathy Study (DRS) findings (DRS report number 8). *Ophthalmology* 88:583–590, 1981
8. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD, for the Collaborative Study Group: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. *N Engl J Med* 329:1456–1462, 1993
9. Rith-Najarian SJ, Stolusky T, Gohdes DM: Identifying diabetic patients at high risk for lower-extremity amputation in a primary health care setting: a prospective evaluation of simple screening criteria. *Diabetes Care* 15:1386–1389, 1992
10. Division of Diabetes Translation: *The Prevention and Treatment of Complications of Diabetes Mellitus: A Guide for Primary Care Practitioners*. Atlanta, GA, Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, 1991
11. Davidson JK: Screening for diabetes mellitus. In *Clinical Diabetes Mellitus: A Problem-Oriented Approach*. Davidson JK, Ed. New York, Thieme, 1986, p. 107–113
12. Herron CA: Screening in diabetes mellitus: report of the Atlanta workshop. *Diabetes Care* 2:357–362, 1979
13. American Diabetes Association: Screening for diabetes (Position Statement). *Diabetes Care* 12:588–590, 1989
14. Paterson KR, for Professional Advisory Committee of the British Diabetic Association: Population screening for diabetes mellitus. *Diabetic Med* 10:777–781, 1993
15. American Diabetes Association: American diabetes alert. *Diabetes Forecast* 46 (3): 54–55, 1993
16. Duncan WE, Linville N, Clement S: Assessing risk factors when screening for diabetes mellitus. *Diabetes Care* 16:1403–1404, 1993
17. Burden ML, Burden AC: The American Diabetes Association screening questionnaire for diabetes: is it worthwhile in the U.K.? (Letter) *Diabetes Care* 17:97, 1994
18. Herman WH, Geiss LS: Epidemiology of non-insulin-dependent diabetes mellitus. In *Diabetes: Clinical Science in Practice*. Leslie RDG, Robbins DC, eds. Cambridge, U.K., Cambridge Univ. Press, in press
19. National Center for Health Statistics: *Plan and Operation of the Second National Health and Nutrition Examination Survey*. Vital and Health Statistics series 1, No. 15. Washington, DC, U.S. Govt. Printing Office, 1981 (DHHS publ. no. 81–1317)
20. World Health Organization: *WHO Expert Committee on Diabetes Mellitus. Second Report*. Geneva, World Health Org., 1980 (Tech. Rep. Ser., no. 646) p. 9–14
21. Harris MI, Hadden WC, Knowler WC, Bennett PH: Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in US population aged 20–74 yr. *Diabetes* 36:523–534, 1987
22. Harris MI: Impaired glucose tolerance in the US population. *Diabetes Care* 12:464–474, 1989
23. Society of Actuaries and Association of Life Insurance Medical Directors of America: *Build study, 1979*. Chicago, Society of Actuaries and Association of Life Insurance Medical Directors of America, 1980
24. Breiman L, Friedman JH, Olshen R, Stone CJ: *Classification and Regression Trees*. Belmont, CA, Wadsworth International Group, 1984
25. Cook EF, Goldman L: Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. *J Chron Dis* 37: 721–731, 1984
26. Andersson DKG, Lundblad E, Svardsudd K: A model for early diagnosis of type II diabetes mellitus in primary health care. *Diabetic Med* 10:167–173, 1993