Risk Assessment Tools for Identifying Individuals at Risk of Developing Type 2 Diabetes

Brian Buijsse, Rebecca K. Simmons, Simon J. Griffin, and Matthias B. Schulze*

* Correspondence to Professor Matthias B. Schulze, Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany (e-mail: mschulze@dife.de).

Accepted for publication November 12, 2010.

Trials have demonstrated the preventability of type 2 diabetes through lifestyle modifications or drugs in people with impaired glucose tolerance. However, alternative ways of identifying people at risk of developing diabetes are required. Multivariate risk scores have been developed for this purpose. This article examines the evidence for performance of diabetes risk scores in adults by 1) systematically reviewing the literature on available scores and 2) their validation in external populations; and 3) exploring methodological issues surrounding the development, validation, and comparison of risk scores. Risk scores show overall good discriminatory ability in populations for whom they were developed. However, discriminatory performance is more heterogeneous and generally weaker in external populations, which suggests that risk scores may need to be validated within the population in which they are intended to be used. Whether risk scores enable accurate estimation of absolute risk remains unknown; thus, care is needed when using scores to communicate absolute diabetes risk to individuals. Several risk scores predict diabetes risk based on routine noninvasive measures or on data from questionnaires. Biochemical measures, in particular fasting plasma glucose, can improve prediction of such models. On the other hand, usefulness of genetic profiling currently appears limited.

diabetes mellitus, type 2; predictive value of tests; risk assessment; ROC curve; sensitivity and specificity

INTRODUCTION

Type 2 diabetes is associated with increased risk of cardiovascular disease and premature mortality and is the leading cause of blindness, kidney failure, and nontraumatic amputations resulting from microvascular complications. The preventability or delay of onset of diabetes by lifestyle modifications that primarily promote weight loss or by pharmaceutical intervention has been demonstrated in randomized trials (1–5), prompting several countries to implement national diabetes programs (6) and to develop guidelines for diabetes prevention (7). However, to reduce costs, individual-level intervention programs are typically targeted at individuals at high risk of developing diabetes. To date, diabetes prevention trials included people with impaired glucose tolerance, who can be identified only by conducting an oral glucose tolerance test (8). Mass population screening by oral glucose tolerance test may be less feasible to identify people who might benefit from health promotion interven-

Abbreviations: ARIC, Atherosclerosis Risk in Communities; aROC, area under the receiver operating characteristic curve; EPIC, European Prospective Investigation into Cancer and Nutrition.
according to their absolute risk but would not necessarily 
need to provide accurate estimates of absolute risk. How-
ever, in many circumstances, risk scores will need to provide 
prognostic information and accurate estimation of the likely 
absolute benefit from an intervention for cost-benefit 
analyses. Here, a precise computation of absolute risk is 
important. Furthermore, the decision of an individual to 
participate in an intervention program may be influenced 
by providing information on the expected benefit of the in-
tervention program. Here again, accurate information on 
absolute risk is necessary but should primarily be based 
on modifiable risk factors.

In this review, we provide results from a systematic litera-
ture search on risk scores that have been developed or 
evaluated in general populations to predict future diabetes. 
Secondly, we assess whether risk scores developed and val-
ified in one cohort perform equally well in other cohorts. 
Finally, we explore methodological issues surrounding the 
development, validation, and comparison of diabetes risk 
scores.

METHODS

Search strategy

A comprehensive literature search for studies on diabetes 
risk prediction tools was performed using PubMed, Web of 
Science, and Cochrane Reviews from database inception 
until December 31, 2009. The search strategy focused 
on 4 key elements: type 2 diabetes, risk assessment/score/ 
prediction, specific names of known risk scores, and pros-
pective studies (refer to Web Table 1, the first of 5 supple-
mentary tables posted on the Epidemiologic Reviews Web 
site: http://epirev.oxfordjournals.org). We also screened 
the reference lists of papers identified from the initial 
search. No language restriction was applied.

Selection criteria

We included studies reporting diabetes risk assessment 
tools or scores that 1) were derived from or validated in 
prospective cohort studies, 2) were derived in the general 
adult population and were evaluated for individuals without 
diabetes at baseline, and 3) reported a measure of perfor-
ance of the risk score for predicting incident diabetes. We 
excluded studies that 1) derived or validated diabetes risk 
scores for the general adult population but did not evaluate 
them for individuals without diabetes; 2) derived or evalu-
ated risk prediction tools other than score-type tools, such as 
those using fasting plasma glucose or 2-hour glucose during 
oral glucose tolerance testing alone; and 3) evaluated fewer 
than 3 risk factors. If scores and their evaluation were re-
ported in multiple papers, we included the score only once 
by selecting the paper that reported the most information on 
predictive ability.

Data extraction

Two authors (B. B. and M. B. S.) independently reviewed 
the results from the primary search of titles, followed by the 
abstract and full paper searches (Figure 1). A form was used 
to extract data on the performance of the risk scores in 
a standardized manner for all articles. Included were the 
name of the risk score and study; country and setting; details 
on derivation and validation populations; follow-up for der-
ivation and validation cohorts; definition of diabetes; risk 
factors included in the scores; and measures of performance, 
including discrimination, calibration, sensitivity, specificity, 
and positive and negative predictive values. We also ex-
tracted data from original studies if no information on the 
development or validation of risk scores was available in the 
articles identified in the initial search. Information was gath-
ered from tables and figures as well as the text of manu-
scripts. When the reviewers disagreed with regard to the 
extracted models and details of performance, consensus 
was reached through discussion.

Means of model performance

MEASURES OF DISCRIMINATION. Receiver operating charac-
teristic (ROC) curves are frequently used to evaluate the 
 discriminatory accuracy of diagnostic or screening markers. 
This curve plots the sensitivity of a test against its false-
positive rate across all possible values. The area under the 
ROC curve (aROC or C statistic) is commonly reported as a 
summary measure. It gives the probability that the pre-
dicted risk for a participant with an event is higher than that 
for a participant without an event. An aROC of 0.5 reflects 
a random guess (null hypothesis), whereas an aROC of 1.0 
represents perfect discrimination. ROC curves do not
provide information about actual risks that the models predict or about the proportion of participants who have high-risk or low-risk values. Furthermore, for clinical or public health decision making, measuring classification accuracy (17) for a subset of meaningful thresholds for high risk might be more informative than the overall aROC.

Measures of calibration. Calibration measures the extent to which the model-predicted probability of an event for a person with a specified predictor value is the same as or very close to that for the proportion of all people in the population with those same predictor values who experience the event. For continuous predictors, people are commonly placed in categories of predicted risk, and the category values are compared with the observed event rates for participants in each category. More formally, the Hosmer-Lemeshow test compares observed event rates with average predicted risks, typically using deciles for categories of predicted risk, with statistically significant $P$ values indicating lack of calibration (18). Note that the $P$ value of the Hosmer-Lemeshow test is highly influenced by sample size and grouping (deciles vs. others).

Measures of overall model fit. Overall model fit can be assessed by using Nagelkerke $R^2$, which is analogous to the percentage of variation explained for linear models. Nagelkerke $R^2$ is the fraction of the log-likelihood explained by the predictors in the model, adjusted to a range of 0–1 (19). The Bayes Information Criterion is the value of the log-likelihood with an added penalty for the number of variables in the model; a lower number indicates a better fit (19).

Risk stratification and reclassification assessment. It has been suggested that it is necessary to evaluate performance of a prediction model in terms of its capacity to stratify the population into clinically relevant risk categories (17). The main assumption is that a better model would place more participants at the extremes of the risk distribution, with the upper category having clear implications for preventive interventions. It has further been suggested that the contribution of new markers to the performance of prediction models should also be evaluated based on risk stratification (17, 19, 20). ROC curves have been criticized in this context because they require a strong “independent” association of a new marker with the outcome to meaningfully increase aROCs compared with a model containing standard risk factors that already allow reasonably good discrimination (21).

The method of reclassification groups predicted risk estimates into clinically relevant categories and cross-classifies these categories for 2 different, but nested prediction models. In addition, event rates within categories of predicted risk before and after reclassification are frequently compared. The net reclassification improvement and the integrated discrimination improvement are statistical measures to quantify and test the statistical significance of the improvement in risk classification (21). Whether net reclassification improvement and integrated discrimination improvement are indeed more sensitive than the C statistic to detect small improvements in discrimination remains largely unknown thus far. We previously reported that improvement in discrimination by glycated hemoglobin (HbA1c) over the Framingham prediction model for coronary heart disease was significant comparing C statistics but not using net reclassification improvement (22) and that even small improvements in discrimination were reflected in C statistics, largely mirrored by the integrated discrimination improvement (23). Thus, despite recent statistical advances, there are still unanswered questions on how to best evaluate risk prediction models.

RESULTS

Our electronic search yielded 4,704 potentially relevant papers (Figure 1). After reviewing the titles and abstracts, 514 references remained; after further review of full texts, 40 articles from the literature search reporting the predictive performance of diabetes risk scores or models met the inclusion criteria. Reasons for exclusion of articles based on the review of full texts (24–45) are given in Web Table 2. The review of reference lists revealed 16 additional references; 3 of these studies derived prediction models cross-sectionally (46–48). However, because these risk scores have been evaluated in other prospective studies meeting inclusion criteria, we included the studies to describe the prediction scores. Thus, a total of 56 references were included in our review.

Development of risk scores

We identified 46 studies that derived risk prediction models for diabetes. Table 1 summarizes 10 studies (46–55) that developed risk prediction models and the performance of these models in external cohorts (47, 51, 53, 55–74). A more detailed description of study characteristics and model performance is given in Web Tables 3 (internal performance) and 4 (external performance). The other 36 studies reporting models not yet externally validated (23, 58, 59, 61–63, 65, 66, 68, 72–98) are described in Web Table 5. Of the total of 46 studies, the vast majority were carried out in either North American or European study populations. A few reports were based on Asian (48, 58, 61, 83, 98) populations, and only single reports were identified for study populations from Mauritius (74) and Australia (65). Cohort size ranged from 492 (88, 97) to 3,773,585 (62) and follow-up time from 3 years (58) to 28 years (89). Most studies included men and women, with the exception of 5 studies (49, 80, 93, 94, 98) that included men only. The majority of risk scores incorporated classic diabetes risk factors, such as age, sex, measures of obesity, family history of diabetes, and blood pressure status.

Prediction models including noninvasive measures only. Seventeen studies evaluated risk models involving noninvasively measured variables. The aROCs for these models generally ranged from 0.7 to 0.8 (52, 54, 55, 58, 59, 63, 68, 81, 84, 91, 94, 96). A few studies reported aROCs of <0.7 (48, 61, 91, 92), with risk models involving 3–4 variables. Only 2 studies reported aROCs of >0.8 in the derivation cohorts. The Finnish Diabetes Risk Score was based on the FINRISK studies and includes information on age; body mass index; waist circumference; history of
hypothesis of prevalent/latent diabetes; physical activity; and consumption of fruits, vegetables, and berries (aROC for integer point score: 0.85) (51). The study focused on drug-treated diabetes as the outcome; thus, cases who did not use medication were not excluded at baseline and were not identified as incident cases during follow-up. The German Diabetes Risk Score (aROC: 0.84) was derived from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study and includes information on age; waist circumference; height; history of hypertension; physical activity; and consumption of alcohol, coffee, whole grains, and red meat (53). This score was modified by categorizing variables to create an integer point score that had a slightly lower discriminatory ability (aROC: 0.83) (99).

Predicting models including biochemical measures. Models that include metabolic syndrome factors. Several prediction models have been proposed that include biochemical measures along with noninvasively measured variables. Studies have evaluated sensitivities, specificities, and predicted values for varying definitions of the metabolic syndrome (reviewed by Ford et al. (100)). ROC curves were reported in 7 studies, with the areas under the curve ranging from 0.68 to 0.85 (52, 74, 77, 78, 80, 97, 101). Some studies have evaluated models with the metabolic syndrome in addition to basic noninvasive parameters (66, 73, 78, 85–87). Although definitions of the metabolic syndrome vary, they generally include concentrations of blood lipids (high density lipoprotein cholesterol, triglycerides) and plasma glucose (either fasting or 2-hour) along with blood pressure and waist circumference. These biochemical parameters have also been evaluated in several other studies.

Biochemical markers to improve model performance based on noninvasively measured risk factors could be particularly useful if diabetes risk screening involves a multistep procedure, with simple questionnaires or noninvasive information at the start and the more costly measurement of biochemical indicators in prescreened individuals during a second step. This process has rarely been assessed, however. In the Atherosclerosis Risk in Communities (ARIC) study, the aROC increased from 0.71 to 0.80 (P < 0.001) when fasting plasma glucose and lipids were added to noninvasively measured variables (52). Similarly, systolic blood pressure, fasting glucose, high density lipoprotein cholesterol, and triglycerides increased the aROC from 0.72 to 0.85 (P value not reported) after they were added to a model that included age, sex, family history, and body mass index in the Framingham Offspring Study (54). Improvements in discrimination were also observed in a Thai population (81). The German Diabetes Risk Score improved with inclusion of additional measurements of fasting glucose, glycated hemoglobin, triglycerides, high density lipoprotein cholesterol, and liver enzymes (aROC: 0.90 vs. 0.85, P < 0.001) (23).

Models containing measures of glucose and insulin control. Considerable attention has been paid to whether more sophisticated indexes of glucose and insulin control, for example, homeostasis model assessment and measures of insulin secretion and resistance from oral glucose tolerance tests, would improve prognostic ability. In the Framingham Offspring Study, the aROC did not improve over and above a model including noninvasively measured characteristics, fasting glucose, and lipids (54). Similarly, exchanging fasting glucose and lipids for measures of insulin secretion obtained from oral glucose tolerance tests yielded conflicting results in the Malmö Preventive Project and the Botnia Study (68). Fasting insulin did not appreciably increase the aROC in the ARIC study (52). However, adding 2-hour glucose (50) or 1-hour plasma glucose and insulin secretion/insulin resistance index based on the oral glucose tolerance test (82) to the San Antonio Heart Study model improved the aROC (0.86 vs. 0.84, P = 0.02 and 0.86–0.87 vs. 0.80, P < 0.001, respectively). Furthermore, adding impaired glucose tolerance to a noninvasive model yielded a slightly higher aROC (aROC: 0.78) compared with using impaired fasting glucose (aROC: 0.76) in a Thai population, although the statistical significance of this difference was not reported (81).

Models containing novel biomarkers. Other biochemical markers, although associated with diabetes risk, have rarely been investigated with regard to diabetes prediction. C-reactive protein did not improve discrimination beyond the metabolic syndrome in the Insulin Resistance Atherosclerosis study (78) or beyond the Framingham Offspring Study model (54). Similarly, in the EPIC-Potsdam Study, C-reactive protein did not add prognostic information beyond a more extended prediction model that includes the German Diabetes Risk Score, plasma glucose, glycated hemoglobin, triglycerides, high density lipoprotein cholesterol, and liver enzymes (23). Notably, liver enzymes—along with concentrations of blood lipids—significantly improved discrimination beyond the noninvasively measured variables and measures of glycemia in the EPIC-Potsdam Study (P = 0.002) (23). A risk score from Taiwan includes white blood cell count, although the overall discriminatory accuracy of the derived score was relatively low (61).

Plasma adiponectin concentrations, although strongly and consistently associated with a lower diabetes risk in prospective studies (102), only marginally improved discrimination beyond the German Diabetes Risk Score with standard biochemical variables in the EPIC-Potsdam Study (aROC: 0.902 vs. 0.900, P = 0.047) (23). Adiponectin was 1 of 6 biomarkers (besides C-reactive protein, ferritin, interleukin-2-receptor, fasting plasma glucose, insulin) selected for a biomarker risk score in the Inter99 cohort (96). The aROC was 0.78 and increased to 0.79 (P = 0.059) when family history, age, body mass index, and waist circumference were added.

Predicting models involving genetic information. Few prospective studies have investigated the value of multiple genetic variants in type 2 diabetes prediction (23, 55, 68, 79, 89, 92, 94). Only a small number of single nucleotide polymorphisms were tested in 2 of these studies, yielding no improvement in discrimination of type 2 diabetes beyond noninvasively measured characteristics (55, 79). Multiple single nucleotide polymorphisms only marginally improved discrimination beyond age, sex, and noninvasive characteristics in the Malmö Preventive Project and Botnia Study (68), the Framingham Offspring Study (89), the Rotterdam Study (92), the Health Professionals Follow-up Study (94), and the EPIC-Potsdam Study (23).
Table 1. Diabetes Risk Scores Developed in Populations Sampled Primarily From the General Population and Validated in External Populations

<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Population, Country</th>
<th>Variables Included in the Risk Score</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Atherosclerosis Risk in Communities Study Diabetes Risk Score, United States</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schmidt, 2005 (52)</td>
<td>Atherosclerosis Risk in Communities study, United States</td>
<td>Clinical model: age, ethnicity, parental history, systolic BP, WC, height</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clinical model + fasting glucose</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clinical model + fasting glucose, triglycerides, HDL cholesterol</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metabolic syndrome National Cholesterol Education Program–Third Adult Treatment Panel definition (1 point for each high WC, high triglycerides, low HDL cholesterol, high BP/antihypertensive use, high fasting glucose)</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Augmented metabolic syndrome (1 point for each high WC, high triglycerides, low HDL cholesterol, high BP/antihypertensive use; 2 points for fasting glucose ( \geq 5.6 \text{ mmol/L} ) or 5 points for fasting glucose ( \geq 6.1 \text{ mmol/L} ); 1 point for BMI ( \geq 30 \text{ kg/m}^2 ))</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Validation in external populations:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainous, 2007 (56)</td>
<td>Coronary Artery Risk in Young Adults, United States</td>
<td>Augmented metabolic syndrome: WC, triglycerides, HDL cholesterol, hypertension, fasting glucose, BMI (6/6)</td>
<td>0.70</td>
</tr>
<tr>
<td>Stern, 2008 (57)</td>
<td>San Antonio Heart Study, United States</td>
<td>Not reported in detail</td>
<td>0.870</td>
</tr>
<tr>
<td>Sun, 2009 (58)</td>
<td>MJ Longitudinal health-check-up-based Population Database, Taiwan</td>
<td>Age, ethnicity, family history, fasting glucose, systolic BP, WC, height (7/7)</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age, ethnicity, family history, fasting glucose, systolic BP, WC, height, triglycerides, HDL cholesterol (9/9)</td>
<td>0.84</td>
</tr>
<tr>
<td>Sun, 2009 (58)</td>
<td>MJ Longitudinal health-check-up-based Population Database, Taiwan</td>
<td>Age, ethnicity, family history, fasting glucose, systolic BP, WC, height (7/7)</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age, ethnicity, family history, fasting glucose, systolic BP, WC, height, triglycerides, HDL cholesterol (9/9)</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Cambridge Diabetes Risk Score, United Kingdom</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Griffin, 2000 (46)</td>
<td>Population from general practices, United Kingdom</td>
<td>Model for predicting undiagnosed diabetes: age, sex, BMI, smoking status, corticosteroid use, antihypertensive use, family history</td>
<td>Independent sample: 0.80</td>
</tr>
<tr>
<td><strong>Validation in external populations:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simmons, 2007 (59)</td>
<td>EPIC-Norfolk, United Kingdom</td>
<td>Age, sex, prescribed antihypertensive medication, prescribed steroids, BMI, family history of diabetes, smoking (7/7)</td>
<td>0.76</td>
</tr>
<tr>
<td>Rahman, 2008 (60)</td>
<td>EPIC-Norfolk, United Kingdom</td>
<td>Age, sex, family history, smoking, prescribed antihypertensive medication, prescribed steroids, BMI (7/7)</td>
<td>0.745</td>
</tr>
<tr>
<td>Chien, 2009 (61)</td>
<td>Cohort, China</td>
<td>Not reported</td>
<td>0.581</td>
</tr>
<tr>
<td>Hippisley-Cox, 2009 (62)</td>
<td>Cohort from general practices, United Kingdom</td>
<td>Age, sex, BMI, smoking status, corticosteroid use, antihypertensive use, family history (7/7)</td>
<td>Men: 0.801; women: 0.813</td>
</tr>
</tbody>
</table>
| **Data From an Epidemiological Study on the Insulin Resistance Syndrome Diabetes Risk Score, France (55)**

Table continues
<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Population, Country</th>
<th>Variables Included in the Risk Score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Discrimination&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balkau, 2008 (55)</td>
<td>Data from an Epidemiological Study on the Insulin Resistance syndrome, France</td>
<td>Men: Clinical prediction model—current smoking, WC, hypertension</td>
<td>0.733</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Men: Clinical and biologic model—current smoking, WC, fasting glucose, fasting glucose squared, gamma-glutamyltransferase</td>
<td>0.850</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Men: above variables + risk alleles for transcription factor 7-like 2 and interleukin 6</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Men: Integer clinical risk score of WC, current smoking, hypertension</td>
<td>0.713</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women: Clinical prediction model—family history, WC, hypertension</td>
<td>0.839</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women: Clinical and biologic model—family history, BMI, fasting glucose, fasting glucose squared, triglycerides</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women: above variables + risk alleles for transcription factor 7-like 2 and interleukin 6</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women: integer clinical risk score of WC, family history, hypertension</td>
<td>0.827</td>
</tr>
<tr>
<td>Validation in external populations:</td>
<td>Atherosclerosis Risk in Communities study, United States</td>
<td>WC, hypertension, current smoker (men), family history (women) (3/3)</td>
<td>0.66</td>
</tr>
<tr>
<td>Kahn, 2009 (63)</td>
<td></td>
<td>Finnish Diabetes Risk Score, Finland (51)</td>
<td></td>
</tr>
<tr>
<td>Lindström, 2003 (51)</td>
<td>FINRISK, Finland</td>
<td>Concise model: age, BMI, WC, history of antihypertensive use, previous diabetes</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Full model: concise model + physical inactivity, fruit and vegetable intake</td>
<td>0.860</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Score model: age, BMI, WC, antihypertensive use, previous diabetes, physical activity, fruit and vegetables intake</td>
<td>0.852</td>
</tr>
<tr>
<td>Validation in external populations:</td>
<td>FINRISK, Finland</td>
<td>Full model: age, BMI, WC, antihypertensive use, previous diabetes, physical activity, fruit and vegetables intake (7/7)</td>
<td>0.87</td>
</tr>
<tr>
<td>Lindström, 2003 (51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alsema, 2008 (64)</td>
<td>Hoom Study, the Netherlands</td>
<td>Concise model: age, BMI, WC, antihypertensive medication, previous diabetes, family history (6/5): an extra age category of ≥65 years created and includes family history</td>
<td>0.71</td>
</tr>
<tr>
<td>Alsema, 2008 (64)</td>
<td>Prevention of renal and vascular end-stage disease study, the Netherlands</td>
<td>Concise model: age, BMI, WC, antihypertensive medication, previous diabetes, family history (6/5): an extra age category of ≥65 years created and includes family history</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 1. Continued

<sup>a</sup>Discrimination is calculated as c-index,

<sup>b</sup>Variables are listed in order of entry in the risk score.

<sup>c</sup>Discrimination is calculated as c-index.

Am J Epidemiol 2011;33:46–62
<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Population, Country</th>
<th>Variables Included in the Risk Score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Discrimination&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alssema, 2008 (64) Monitoring Project on Chronic Disease Risk Factors Study, the Netherlands</td>
<td>Concise model: age, BMI, WC, antihypertensive medication, previous diabetes, family history (6/5); an extra age category of &gt;65 years created and includes family history</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>Balkau, 2008 (55) Data from an Epidemiological Study on the Insulin Resistance syndrome, France</td>
<td>Full model: age, BMI, WC, antihypertensive medication, physical activity (5/7); excludes previous diabetes and diet</td>
<td>Men: 0.678; women: 0.809</td>
<td></td>
</tr>
<tr>
<td>Cameron, 2008 (65) Australian Diabetes, Obesity and Lifestyle Study, Australia</td>
<td>Deviations from the full original score: includes parental history, activity excludes occupational activity</td>
<td>0.727</td>
<td></td>
</tr>
<tr>
<td>Abdul-Ghani, 2009 (66) Botnia Study, Finland</td>
<td>Concise model: age, BMI, WC, use of antihypertensive medications, family history (5/5); excludes prevalent diabetes, includes family history</td>
<td>0.646</td>
<td></td>
</tr>
<tr>
<td>Wilson, 2007 (54) Framingham Offspring Study, United States</td>
<td>Personal model: age, sex, parental history, BMI</td>
<td>0.724</td>
<td></td>
</tr>
<tr>
<td>Framingham Offspring Diabetes Risk Score, United States</td>
<td>Simple clinical model with categorical variables: age, sex, parental history, BMI, WC, fasting glucose, HDL cholesterol, triglycerides, hypertension</td>
<td>0.852 (repeated random samples: 0.73–0.91)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple point score system: parental history, BMI, fasting glucose, HDL cholesterol, triglycerides, hypertension</td>
<td>0.850</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple clinical model with continuous variables: age, sex, parental history, BMI, systolic BP, WC, fasting glucose, HDL cholesterol, triglycerides</td>
<td>0.881</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex clinical model: age, sex, parental history, BMI, WC, fasting glucose, HDL cholesterol, triglycerides, 2-hour glucose, fasting insulin, C-reactive protein</td>
<td>0.854</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Best biologic model: complex clinical model + hormone therapy, current smoking, alcohol intake, aspirin or nonsteroidal antiinflammatory drug use, glycated hemoglobin, homeostatic model assessment of insulin resistance, Gutt insulin sensitivity index, homeostatic model assessment beta-cell index</td>
<td>0.869</td>
<td></td>
</tr>
<tr>
<td>Validation in external populations:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li, 2007 (67) Cohort, Germany</td>
<td>Reestimated simple clinical model: age, sex, family history, BMI, hypertension, HDL cholesterol, triglycerides, fasting glucose (8/9); excludes WC</td>
<td>0.86 (validated: 0.828)</td>
<td></td>
</tr>
<tr>
<td>Lyssenko, 2008 (68) Malmö Preventive Project, Sweden</td>
<td>Personal model: age, sex, family history, BMI (4/4)</td>
<td>Categorical: 0.69; continuous 0.707</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple clinical model with categorical variables: age, sex, family history, BMI, BP, triglycerides, fasting glucose (7/9); excludes WC, HDL cholesterol</td>
<td>Categorical: 0.729; continuous: 0.743</td>
<td></td>
</tr>
<tr>
<td>Lyssenko, 2008 (68) Botnia Study, Finland</td>
<td>Personal model: age, sex, family history, BMI (4/4)</td>
<td>Categorical: 0.736; continuous: 0.769</td>
<td></td>
</tr>
</tbody>
</table>

Table continues
<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Population, Country</th>
<th>Variables Included in the Risk Score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Discrimination&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nichols, 2008 (69)</td>
<td>Kaiser Permanente Northwest, United States</td>
<td>Simple clinical model with categorical variables: age, sex, family history, BMI, BP, triglycerides, fasting glucose (7/9); excludes WC, HDL cholesterol</td>
<td>Categorical: 0.755; continuous: 0.786</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Family history used as proxy for parental history</td>
<td></td>
</tr>
<tr>
<td>Personal model: age, sex, parental history, BMI (4/4); reestimated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple clinical model with categorical variables: age, sex, parental history, BMI, fasting glucose, HDL cholesterol, triglycerides, hypertension (8/9); reestimated model excludes WC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple clinical model with continuous variables: age, sex, parental history, BMI, fasting glucose, HDL cholesterol, triglycerides, hypertension (8/9); reestimated model excludes WC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple point score system: parental history, BMI, fasting glucose, HDL cholesterol, triglycerides, hypertension (6/6)</td>
<td></td>
<td></td>
<td>Not reported</td>
</tr>
<tr>
<td>Chien, 2009 (61)</td>
<td>Cohort, China</td>
<td>Not reported</td>
<td>0.662</td>
</tr>
<tr>
<td>Kahn, 2009 (63)</td>
<td>Atherosclerosis Risk in Communities study, United States</td>
<td>Simple point score: fasting glucose, BMI, HDL cholesterol, parental diabetes, triglycerides, hypertension (6/6)</td>
<td>0.76</td>
</tr>
<tr>
<td>Schulze, 2007 (53, 99)</td>
<td>EPIC-Potsdam, Germany</td>
<td>Full model: age, WC, height, hypertension, physical activity, smoking, and consumption of whole-grain bread, red meat, coffee, moderate alcohol</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simplified model with categorical variables: age, WC, height, hypertension, physical activity, smoking, and consumption of whole-grain bread, red meat, coffee, moderate alcohol</td>
<td>0.83</td>
</tr>
<tr>
<td>Validation in external populations:</td>
<td></td>
<td>Full model: age, WC, height, hypertension, physical activity, smoking, and consumption of whole-grain bread, red meat, coffee, moderate alcohol (10/10)</td>
<td>0.82</td>
</tr>
<tr>
<td>Schulze, 2007 (53)</td>
<td>EPIC-Heidelberg, Germany</td>
<td>Model for predicting undiagnosed diabetes: age, WC, family history, physical activity</td>
<td>0.698</td>
</tr>
<tr>
<td>Mohan, 2005 (48)</td>
<td>Chennai Urban Rural Epidemiology Study, India</td>
<td>Age, WC, family history, physical activity (4/4)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Mohan, 2008 (70)</td>
<td>Chennai Urban Population Study, India</td>
<td>Age, BMI, fasting glucose, HDL cholesterol, family history, hypertension</td>
<td>0.793</td>
</tr>
<tr>
<td>von Eckardstein, 2000 (49)</td>
<td>Prospective Cardiovascular Münster Diabetes Risk Score, Germany</td>
<td>Model for predicting persons with 2-hour glucose ≥140 mg/dL: sex, age ≥70 years, triglycerides ≥150 mg/dL, fasting glucose</td>
<td>Continuous: 0.73; categorical: 0.71; score points: 0.70</td>
</tr>
</tbody>
</table>

Table continues
Validation of risk scores in independent cohorts

Ten risk scores were evaluated in different validation cohorts (Table 1, Web Table 3). The majority of validation cohorts consisted of European populations, and sample size varied from 100 (57) to 1,232,832 (62) individuals. The number of incident diabetes cases varied considerably, from 37 in a German cohort (67) to 37,535 in a British cohort.
Most studies identified diabetes cases by using fasting blood glucose measurements and—less frequently—2-hour glucose values during an oral glucose tolerance test. Some studies used alternative strategies to identify cases, for example, registries of medication use, clinical registers, electronic health records, or verified self-reports (14, 51, 59, 60, 62).

Only a few studies reported complete measures of predictive performance, including discrimination, calibration, and sensitivity, specificity, and positive predicted value or negative predicted value for potential cutoffs (14, 58, 61, 69). The majority of studies reported a measure of discrimination (aROC) but lacked information on calibration. Risk scores showed variable discriminatory power in validation cohorts (aROC range: 0.58 (61) to 0.87 (51, 57)).

Several risk scores based solely on noninvasive measurements have been validated in independent populations. The most frequently validated score is the Finnish Diabetes Risk Score, validated in 8 independent cohorts (51, 55, 64–66). The discrimination was very good (aROC: 0.87) in another Finnish study involving similar methodology compared with the cohort study from which the score was derived (51), but it was lower in other populations (aROC range: 0.65–0.81) (55, 64–66). These later studies included some modifications of the risk score, in particular the addition of family history and the omission of diet and activity as predictors, and they involved different endpoint definitions. Calibration measures were not reported.

The Cambridge Diabetes Risk Score was initially developed to identify individuals with undiagnosed diabetes based on information on age, sex, antihypertensive medication use, steroid use, body mass index, family history of diabetes, and smoking status (46). It has been validated in 2 United Kingdom studies: the prospective EPIC-Norfolk Study yielding an aROC of 0.75 (60) and in a large sample of people recruited from general practices (aROC: 0.80 among men and aROC: 0.81 among women) (62), although discrimination was lower in a cohort of Chinese from Taiwan (aROC: 0.58) (61). The Framingham personal model yielded an aROC of 0.68 in a US cohort in which coefficients have been reestimated (69). In the Malmo Preventive Project and the Botnia Study, the aROCs were 0.69 and 0.74, respectively (68). The German Diabetes Risk Score was validated in another German cohort—EPIC-Heidelberg (aROC: 0.82) (53). Calibration analysis suggested accurate estimation of absolute risk in this external cohort.

One model with biochemical measures that has been frequently validated in independent populations is the San Antonio Heart Study model (50). It includes information on age, gender, ethnicity, body mass index, family history of diabetes, systolic blood pressure, fasting glucose, and high density lipoprotein cholesterol. The aROCs were 0.76–0.79 for Japanese Americans (71), 0.785 in the Insulin Resistance Atherosclerosis study (72), 0.765 in the Mexico City Diabetes Study (73), and 0.743 in the Botnia study (66), and graphic display of the ROC curve suggests good discrimination in the Mauritius study (74). However, discrimination was considerably lower among Chinese in Taiwan (aROC: 0.675) (61). The Framingham Offspring Study clinical model (54) includes age, sex, parental history, body mass index, waist circumference, fasting glucose, high density lipoprotein cholesterol, triglycerides, and hypertension. It has been validated in several studies with differing levels of discrimination; aROCs were 0.86 in a German population (67); 0.73 in the Malmo Preventive Project and 0.76 in the Botnia Study (68); 0.84 in Kaiser Permanente Northwest (69); 0.66 in a Chinese population (61); and 0.76 in the ARIC study (63)).

A number of prediction models with relatively similar components have been validated in other cohorts, for example, the PROCAM score (61), the ARIC clinical model plus glucose (52, 57, 58), and the Rancho Bernardo model (47, 66). Although aROCs (mostly in the range of 0.7–0.8) suggest overall acceptable to good discrimination by most of these latter scores, the vast majority of studies did not report measures of calibration.

**DISCUSSION**

This systematic review shows that the predictive ability of diabetes risk scores, which have been developed in populations of varying ethnic backgrounds, differs considerably between populations. Several risk scores exist that enable prediction of type 2 diabetes based on information readily available in routine clinical practice or that can be gathered by questionnaires.

Although collecting data from a questionnaire is likely less costly and more acceptable than methods of screening involving biochemical measures such as blood glucose, difficulties in distributing questionnaires, the time required to complete them, the complexity of computing the results, issues related to misreporting (reporting bias), and unavailability of some required information may hamper their population-wide application. Questionnaires may also create anxiety or false reassurance.

Risk scores based entirely on routine health service data have the advantage that all necessary information has already been collected, but this approach may also create false reassurance or anxiety if test results are communicated to patients. Furthermore, these risk scores focus mainly on nonmodifiable risk factors such as age and family history or on the consequences of adverse health behaviors such as high body mass index and waist circumferences, high blood pressure, and medication use. In addition, available risk factor information might differ between health services.

The feasibility of implementing any screening model will depend on the availability and completeness of the required risk factor data (103). Furthermore, the context in which prediction models are used may largely determine the degree of complexity of their calculation. Some models involve categorization of noninvasively measured variables and do not require a calculator (51, 99) and are thus applicable as paper questionnaires; other prediction models involve considerable computational effort. Thus, performance of alternative models needs to be weighed against the feasibility of their application. However, current technology
can be used to calculate more complicated risk scores. Thus, increasing accessibility of computerized calculators (e.g., software applications, Web tools) may allow future development of risk prediction tools with more emphasis on accuracy than on simplicity of calculation.

Biochemical measures, in particular fasting plasma glucose, can strongly improve the performance of models based on noninvasive measures. Although other markers that are relatively easily obtained in clinical practice—such as high density lipoprotein cholesterol, triglyceride, and liver enzymes—add a small increase in predictive value, there is little evidence for less commonly measured parameters, such as C-reactive protein or adiponectin. The overall sensitivity and specificity of a simple prediction model using routine data might exceed that of one involving a blood test if the response rate for attendance at a blood test is low and the routine data are available for the majority of the population. Indeed, risk factor questionnaires (51) and risk scores generated from data routinely available in general practice (46) are increasingly being used to stratify populations before inviting those at high risk to undergo blood glucose testing. Recent data from the United Kingdom suggest that an approach of population stratification prior to inviting people to be screened for cardiovascular disease risk factors is likely to be more efficient than inviting all adults (104). In the Diabetes Prevention Program, older age and higher body mass index increased the yield of screening (105).

The usefulness of genetic profiling currently appears limited. Because the discriminative accuracy of genetic profiling depends on the number of genes involved, the frequency of the risk alleles, and the risks associated with the genotypes (106, 107), a large number of additional common variants with small effect sizes or rare variants with stronger effect sizes need to be identified. Novel diabetes genes identified by genome-wide association studies, requiring tens of thousands of cases for sufficient statistical power, confer a very modest increase in risk of each risk allele (odds ratios: 1.1–1.2) (108). Even if attempts to identify enough genetic variants were made, it remains unclear how such information can be communicated and whether it will motivate people to adopt healthy lifestyles and to seek medical interventions (109).

Diabetes risk scores demonstrated good discrimination in the study populations in which they were derived. However, their predictive value was usually reduced in external populations. Studies that derive risk scores in one-half of the cohort and validate them in the other half, or validate risk scores in cohorts with very similar methodology (e.g., endpoint definition, exposure information collection) or source populations, are likely to report better predictive abilities. This might, for example, be true for scores developed and validated in the FINRISK studies (Finnish Diabetes Risk Score (51)) and the EPIC-Germany studies (German Diabetes Risk Score (53)). Conversely, validating risk scores in different populations and ethnic groups is likely to result in relatively poorer performance, as has been observed for the Finnish Diabetes Risk Score (55, 64–66).

Thus, risk prediction models should not be assumed to perform comparably well but may rather need to be validated within the population in which they are intended to be used, particularly if ethnicities and countries differ from the derivation cohorts. Furthermore, reestimation of regression coefficients for existing models may result in better performance when models are evaluated in external populations (71). It may also be more useful to develop population-specific risk prediction tools (103) rather than try to find a universal risk score that will work in all populations. Although validation studies have been undertaken in the United States, Australia, several European countries, India, and China, such data are largely lacking from African, South-American, southern and eastern European, and most Asian countries.

Information on sensitivities, specificities, and predicted values is essential for deciding appropriate cutoffs based on cost-benefit considerations. Such data were unavailable for several prediction models identified in this review. Furthermore, most evaluation studies did not assess model calibration. Thus, whether absolute risk is estimated accurately remains unclear for most existing diabetes risk scores, which has implications for the applicability of scores in the context of prevention programs focusing on motivation of individuals to change their behavior, where accurate estimation of absolute risk is necessary. Although modifiable risk might be more informative than absolute risk in this context, most evaluated risk scores are dominated by non-modifiable factors such as age, sex, ethnicity, and family history. Modifiable risk factors usually include measures of obesity (body mass index, waist circumference) but, less frequently, smoking and, rarely, others such as diet and physical activity (51, 53, 58).

To our knowledge, this systematic review is the first to assess the ability of risk scores to estimate risk of incident type 2 diabetes in healthy individuals from general populations. Different definitions of the diabetes endpoint as well as differences in follow-up time, source population, and methods of collection and modeling of risk factors make it difficult to compare the performance of risk scores. Furthermore, the majority of published diabetes prediction models were not validated in independent studies, and, if a prediction model was validated, the original risk model was frequently modified. Although a variety of statistical approaches were used to describe the performance of risk models, they were mostly limited to a global measure of discrimination (aROC). Identification of different prediction models and extraction of model information was based on tables and figures as well as on text in the results section of papers. Although data were extracted independently by 2 reviewers and disagreement required consensus between them, we cannot rule out the possibility that information was falsely extracted or missed.

Methodological issues

Study design and population. Prediction models for incident diabetes should be prospectively derived and validated in initially disease-free populations in observational studies. Epidemiologists have generally used large-scale cohort studies for this purpose. However, some investigators have used different approaches with weaker designs, for example, without excluding prevalent cases at baseline (35, 110). Evaluation of patients undergoing intervention...
(41, 111) frequently involves prescreening, which hampers extrapolation to general populations. Furthermore, linking the baseline risk factor profile to incidence is distorted by the intervention. In addition, case-control designs have been used to evaluate genetic markers as predictors of diabetes (112, 113). This design might be appropriate to evaluate genetic risk alone if controls and cases are population based. However, case-control studies are hampered by several sources of bias involved in analysis of lifestyle risk factors, including differential reporting based on disease status (recall bias) and reverse causation, making it problematic to evaluate genetic markers beyond lifestyle or metabolic risk factors. Some investigators did not evaluate the performance of risk prediction models in general population samples but rather among individuals after an initial prescreening, for example, individuals with a positive family history of diabetes (24) or prevalent impaired glucose tolerance (28). Such studies did not meet our predefined inclusion criteria and were thus excluded from our review.

**Case definition.** Several studies relied on self-reported diabetes. The validity of self-reported data may distort relative risk estimates and corresponding prediction models, particularly in the presence of false-positive self-reports. This misclassification can be reduced if studies apply thorough validation procedures. Although there might still be misclassification present because of undiagnosed diabetes, assuming this misclassification is not dependent on risk factor status, this does not bias estimates of relative risk (114). Still, false-negative self-reports may distort estimates of discrimination and calibration.

Most studies used glucose screening to detect prevalent cases at baseline and incident cases during follow-up. Although undiagnosed diabetes might not be an issue in such studies, the results of prediction models would apply to similarly screened populations. Universal glucose screening, either fasting or by oral glucose tolerance test, is, however, not presently carried out, so studies based on self-reports only might more accurately reflect “real-world” conditions of diabetes diagnostics in general populations. In addition, studies involving glucose measurements usually base identification of cases on a single measurement, resulting in false-positive screens (115, 116). Little is known about whether the performance of risk scores depends on the method of case identification. The Cambridge Risk Score (46) was more strongly related to diabetes risk in the EPIC-Norfolk study when prevalent and incident cases were identified based on self-reports, clinical registers, and death certificates compared with also using glycated hemoglobin measurements (60). Perhaps even more important than choosing either self-report only or additional glucose screening is that studies use similar definitions of case status at baseline and at follow-up.

**Model derivation.** Modeling risk factors to derive prediction models in cohort studies most frequently involved logistic regression, although some studies used Cox regression models, which might better reflect the prospective nature of these studies. Variables were usually retained in a prediction model if they were significantly associated with diabetes risk, a process highly dependent on statistical power. Some investigators also considered variables that were not significant predictors (51).

Calculation of a graded risk score is usually based on the set of chosen variables and corresponding beta-coefficients from regression models. For example, beta-coefficients from logistic or Cox regression models were used directly or were transformed to assign points in the San Antonio diabetes model (50), ARIC models (52), Framingham Offspring model (54), EPIC-Norfolk risk score (59), Cambridge Score (46), and German Diabetes Risk Score (53). However, other investigators translated observed beta-coefficients into relatively crude score points, not matching observed weights from regression (51).

Choosing appropriate cutoffs to determine “high risk.” The use of risk classification and reclassification is based on the assumption that individuals should be stratified into clinically relevant risk categories. This assumption seems logical because screening for subpopulations is a prerequisite for the high-risk approach of prevention or for selection of persons to include in clinical trials. One approach for selecting cutoffs is to base decisions on existing thresholds above which risk increases sharply with increasing risk factor profiles. Unfortunately, diabetes risk factors generally do not provide evidence for such thresholds. For example, although clinical categories for waist circumference are in use, diabetes risk appears to increase with each centimeter of waist circumference, even within the range of values considered normal (45). The same applies to predicted risk estimates from more complex prediction models such as diabetes risk scores. Thus, justification of cutoffs based on observed risk associations is challenging.

Another approach for defining risk categories is based on ROC curves: the pair of sensitivity and false-positive rates closest to the upper left corner is considered optimal here because the slope of the curve indicates that any cutoff yielding higher sensitivity (benefit) would result in disproportionally higher costs in terms of a false-positive rate, and vice versa. This approach has been, in part, the rationale for lowering the cutoff for impaired fasting glucose from 110 mg/dl to 100 mg/dl, for example (117).

National Cholesterol Education Program–Adult Treatment Panel III guidelines consider different therapeutic approaches based on cost-effectiveness analyses for different categories of absolute cardiovascular disease risk based on the Framingham algorithm (118). These risk categories have been the basis for evaluating reclassification after including novel cardiovascular disease biomarkers (119, 120). However, it is clear that the cost-effectiveness of cholesterol-lowering therapy increases with increasing baseline risk (121) and may change depending on changes in drug costs, efficacy of interventions, costs of treating new cases and sequelae, or compliance characteristics of the population. Thus, risk categories may satisfy clinicians’ requests for thresholds to trigger certain interventions, but they are largely arbitrary (122).

Furthermore, population-based screening for high-risk individuals might assign lower relative costs to false-positive screens compared with clinical intervention studies, where the primary goal might be to select individuals with a high risk of developing diabetes within a relatively short time period. For example, in the Diabetes Prevention Program, only about 5% of those initially contacted were eligible for
the intervention study after several steps of screening (105). If population-based screening either is based on a simple paper questionnaire only or also involves subsequent biomarker evaluation, such as fasting blood glucose, cutoffs would need to be defined quite differently to yield similar overall sensitivities.

These examples highlight the point that cutoffs for a diabetes risk score may vary greatly depending on the specific objectives for using it and the related costs and benefit. However, all these approaches require that sensitivities, specificities, and predicted values for different potential cutoffs for prediction models be known. The varying sensitivities and specificities associated with similar cutoffs across different populations observed suggest that cost-benefit analyses are uncertain unless the prediction model is validated within the specific population in which it is intended to be used. Furthermore, regardless of screening and prevention strategies for high-risk individuals, population-based approaches targeting modifiable diabetes risk factors such as physical activity, diet, obesity, and smoking should be supported (123).

Conclusions

Computation of diabetes risk based on multivariate risk models is useful in the context of targeting prevention interventions to high-risk groups. Several risk scores have been validated in independent populations, frequently showing good discriminatory ability. However, discrimination is generally lower than in the populations in which the scores were developed, and the validation results are more heterogeneous. This finding suggests that risk scores should not simply be expected to perform comparably well but rather may need to be validated within the population in which they are intended to be used. Data on whether risk scores enable accurate estimation of absolute risk are largely lacking from validation studies, which currently limits the use of diabetes risk scores in the context of providing prognostic information to individuals.

Risk scores based on noninvasive measurements can be improved by adding commonly measured biochemical markers, in particular, measures of glycemia. Thus, scores based on noninvasive information—which might be available from routine clinical data or collected by questionnaires—should increasingly be used to identify individuals or population subgroups that might benefit from more comprehensive risk assessment, for example, additional determination of blood glucose levels, or to even start directly with preventive action. A stepwise stratification approach would reduce the number of individuals requiring blood sampling. However, the degree to which existing risk scores can be improved by using novel biochemical markers or genetic information is questionable.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany (Brian Buijsse); MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom (Rebecca K. Simmons, Simon J. Griffin); and Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany (Matthias B. Schulze).

This study was partly funded by the European Union (LSHM-CT-2006-037197) and the NIHR Programme (RP-PG-0606-1259).

Conflict of interest: none declared.

REFERENCES


49. von Eckardstein A, Schulle H, Assmann G. Risk for diabetes mellitus in middle-aged Caucasian male participants of the


