American Journal of Epidemiology
Vol. 171, No. 9
DOI: 10.1093/aje/kwq030
Advance Access publication:
April 7, 2010

Original Contribution

Comparative Validity of 3 Diabetes Mellitus Risk Prediction Scoring Models in a Multiethnic US Cohort

The Multi-Ethnic Study of Atherosclerosis

Devin M. Mann*, Alain G. Bertoni, Daichi Shimbo, Mercedes R. Carnethon, Haiying Chen, Nancy Swords Jenny, and Paul Muntner

* Correspondence to Dr. Devin M. Mann, Department of Medicine, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, Box 1087, New York, NY 10029 (e-mail: devin.mann@mssm.edu).

Initially submitted September 8, 2009; accepted for publication January 25, 2010.

Several models for estimating risk of incident diabetes in US adults are available. The authors aimed to determine the discriminative ability and calibration of published diabetes risk prediction models in a contemporary multiethnic cohort. Participants in the Multi-Ethnic Study of Atherosclerosis without diabetes at baseline (2000–2002; $n = 5{,}329$) were followed for a median of 4.75 years. The predicted risk of diabetes was calculated using published models from the Framingham Offspring Study, the Atherosclerosis Risk in Communities (ARIC) Study, and the San Antonio Heart Study. The mean age of participants was 61.6 years (standard deviation, 10.2); 29.3% were obese, 53.1% had hypertension, 34.9% had a family history of diabetes, 27.5% had high triglyceride levels, 33.8% had low high density lipoprotein cholesterol levels, and 15.3% had impaired fasting glucose. There were 446 incident cases of diabetes (fasting glucose level $\geq$126 mg/dL or initiation of antidiabetes medication use) diagnosed during follow-up. *C* statistics were 0.78, 0.84, and 0.83 for the Framingham, ARIC, and San Antonio risk prediction models, respectively. There were significant differences between observed and predicted diabetes risks (Hosmer-Lemeshow goodness-of-fit chi-squared test for each model: $P < 0.001$). The recalibrated and best-fit models achieved sufficient goodness of fit (each $P > 0.10$). The Framingham, ARIC, and San Antonio models maintained high discriminative ability but required recalibration in a modern, multiethnic US cohort.

cohort studies; diabetes mellitus; models, statistical; risk; validation studies as topic

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CI, confidence interval; HDL, high density lipoprotein; MESA, Multi-Ethnic Study of Atherosclerosis.

The prevalence (1) and incidence (2) of diabetes mellitus are increasing at alarming rates in the United States and worldwide. The identification of adults at high risk for incident diabetes is important for implementing interventions in a cost-effective manner (3, 4). Clinicians use many risk factors, such as obesity, family history, hypertension, the metabolic syndrome, and impaired fasting glucose, to gauge a patient's risk for the development of diabetes.

Several equations for predicting incident diabetes among US and international populations have been developed (5–9). Using a variety of risk factors and weighting schemes, these equations have demonstrated strong test characteristics. However, these equations were developed in ethnically and geographically limited populations. For example, the Framingham Offspring Study included primarily whites, the Atherosclerosis Risk in Communities (ARIC) Study included whites and African Americans, and the San Antonio Heart Study was limited to whites and Mexican Americans. In addition, these cohort studies were initiated over 20 years ago, and their data may not reflect the latest trends in the diabetes epidemic. Furthermore, these equations have not been externally validated. The ability of a risk prediction model to identify persons at high risk for diabetes (i.e., discrimination) in a multiethnic

population is important, as there are substantial disparities in diabetes risk by ethnicity (10). Additionally, calibration provides important information beyond discrimination, since it assesses the accuracy of the predicted risks as compared with observed risks (11). As has been seen for the Framingham coronary heart disease risk scoring system, there can be significant over- or underestimation of events when the equations are applied to diverse populations without appropriate calibration (12, 13).

Our objective in this analysis was to determine the discriminative ability and calibration of 3 published diabetes risk prediction models derived from single (Framingham) or biethnic (ARIC, San Antonio) US populations in a contemporary, multiethnic population of US adults. To do so, we analyzed longitudinal data on the incidence of diabetes from the Multi-Ethnic Study of Atherosclerosis (MESA).

MATERIALS AND METHODS

Study population

Details regarding the design and objectives of MESA have been published previously (14). In brief, 6,814 white, African-American, Hispanic, and Chinese-American participants aged 45–84 years with no evidence of clinical cardiovascular disease were recruited from 6 geographically diverse US communities (Baltimore, Maryland; Chicago, Illinois; Forsyth County, North Carolina; Los Angeles, California; New York, New York; and St. Paul, Minnesota). Participants with atrial fibrillation, active cancer, cognitive impairment, or weight greater than 300 pounds (136 kg) and those who were pregnant were excluded. Among the 6,814 MESA participants enrolled, 859 had diabetes (fasting glucose level $\geq$126 mg/dL or use of hypoglycemic medication or insulin) at baseline and 626 were missing information needed to calculate their diabetes risk. After these exclusions, 5,329 participants were included in the current analyses.

Data collection

Data were collected during a baseline examination (2000–2002) and at 3 follow-up visits (MESA examinations 2, 3, and 4) occurring at 18-month intervals. During the baseline examination, standardized questionnaires were utilized to obtain data on demographic factors, tobacco use, medical conditions, and currently prescribed medications. Body weight, height, and waist circumference were measured by trained study staff. At baseline, height and weight were measured with participants wearing light clothing and no shoes. An Accu-Hite stadiometer (Seca GmbH & Company KG, Hamburg, Germany) was used to measure height, and a Detecto platform balance scale (Titus Home Health Care, Alhambra, California) was used to measure weight. Body mass index was calculated as weight (kg) divided by height (m) squared; overweight was defined as body mass index $\geq$25 and obesity as body mass index $\geq$30. Waist circumference was measured using a Gulick II anthropometric tape (Sammons Preston, Chicago, Illinois) applied horizontally at the level of the umbilicus and was rounded to the nearest centimeter. Family history of diabetes (parent or sibling) was not assessed at baseline but was assessed at examination 2, and this response was used for analysis.

Resting seated blood pressure was measured 3 times using an automated oscillometric sphygmomanometer (Dinamap PRO 100; Critikon, Tampa Bay, Florida); the last 2 measurements were averaged for analysis. Elevated blood pressure was defined as systolic blood pressure $\geq$130 mm Hg or diastolic blood pressure $\geq$85 mm Hg or use of antihypertensive medication. Participants were asked to fast overnight prior to their examination. Fasting blood glucose and lipid levels were analyzed at a central laboratory. Serum glucose level was measured using a Vitros analyzer (Johnson & Johnson Clinical Diagnostics, Rochester, New York). Among participants who were not using hypoglycemic drugs or insulin, impaired fasting glucose was defined as a glucose level of 100–125 mg/dL. Plasma lipids, including high density lipoprotein (HDL) cholesterol and triglycerides, were measured using a standardized kit (Roche Diagnostics Corporation, Indianapolis, Indiana). Low HDL cholesterol levels were defined as $<$40 mg/dL for men or $<$50 mg/dL for women, and high triglyceride levels were defined as $\geq$150 mg/dL. Metabolic syndrome was defined according to the revised National Cholesterol Education Program Adult Treatment Panel III criteria (15). The individual components of the metabolic syndrome include waist circumference $>$88 cm for women or $>$102 cm for men, serum glucose level $\geq$110 mg/dL, systolic/diastolic blood pressure $\geq$130/85 mm Hg or use of antihypertensive medication, HDL cholesterol level $<$40 mg/dL for men or $<$50 mg/dL for women, and triglyceride level $\geq$150 mg/dL. Participants with 3 or more of these components were categorized as having metabolic syndrome.

Risk prediction models

The primary exposures of interest included equations for predicting incident diabetes derived in the Framingham Offspring Study, the ARIC Study, and the San Antonio Heart Study (see Appendix). The baseline visits used for developing these models were conducted in 1971–1975 for the Framingham Offspring Study, 1987–1989 for the ARIC Study, and 1979–1988 for the San Antonio Heart Study, and diabetes incidence in these studies was determined over periods of 8, 9, and 7.5 years, respectively. Details on each of these cohort studies and the development of their diabetes risk prediction equations are available elsewhere (5–7).

In brief, the Framingham diabetes risk prediction model includes overweight and obesity, impaired fasting glucose, low HDL cholesterol, high triglycerides, elevated blood pressure, and parental history of diabetes. This model uses a point scoring algorithm with the risk of incident diabetes being correlated with a person's overall point score. The diabetes risk prediction model derived in the ARIC Study includes height, waist circumference, black race/ethnicity, systolic blood pressure, fasting glucose, HDL cholesterol, triglycerides, and a parental history of diabetes. The San Antonio diabetes risk prediction model includes age, sex, Mexican-American ethnicity, fasting glucose, systolic blood pressure, HDL cholesterol, body mass index, and family history of diabetes (parent or sibling). MESA included

Table 1. Mean Levels and Prevalences of Diabetes Risk Factors Among Participants Without Diabetes at Baseline (2000–2002), Overall and By Race/Ethnicity, Multi-Ethnic Study of Atherosclerosis

Characteristic	Total (n = 5,329)		Whites (n = 2,272)		African Americans (n = 1,341)		Hispanics (n = 1,101)		Chinese Americans (n = 615)	
	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%
Age, years	61.6 (10.2)		62.3 (10.2)		61.4 (10.2)		60.5 (10.3)		60.9 (10.2)	
Male sex		46.7		47.6		44.1		47.2		48.1
Body mass index[a]	28.0 (5.3)		27.5 (4.9)		29.8 (5.8)		29.0 (4.9)		23.9 (3.3)	
25–29		40.3		41.1		38.6		46.3		29.8
≥30		29.3		25.4		42.1		35.3		4.4
Waist circumference, cm		97.0		97.1		86.6		99.7		99.4
Abdominal obesity[b]		51.4		51.0		22.6		59.2		58.7
Systolic blood pressure, mm Hg	125.1 (20.9)		122.6 (20.0)		130.6 (21.2)		125.2 (21.2)		122.5 (21.0)	
Diastolic blood pressure, mm Hg	71.8 (10.2)		70.1 (9.9)		74.6 (10.0)		71.8 (10.1)		71.9 (10.3)	
Antihypertensive treatment		33.0		30.7		44.8		28.0		24.6
Elevated blood pressure[c]		53.1		49.9		66.0		48.5		45.4
Parental or sibling history of diabetes		34.9		29.4		42.7		41.5		26.2
Triglyceride level, mg/dL	127.0 (77.5)		129.1 (75.5)		99.1 (54.8)		149.9 (93.8)		138.7 (76.6)	
Triglyceride level ≥150 mg/dL		27.5		29.3		13.0		38.6		32.5
HDL cholesterol level, mg/dL	51.7 (14.9)		52.9 (15.7)		53.3 (15.3)		48.5 (13.4)		50.0 (12.8)	
Low HDL cholesterol level[d]		33.8		30.7		29.4		43.6		37.1
Fasting glucose level, mg/dL	89.4 (10.5)		87.7 (10.0)		90.1 (10.7)		91.0 (10.8)		91.4 (10.0)	
Fasting glucose level 100–125 mg/dL		15.3		11.3		17.4		18.8		19.3
Metabolic syndrome		25.5		25.9		22.4		32.6		18.4

Abbreviations: HDL, high density lipoprotein; SD, standard deviation.
[a] Weight (kg)/height (m)^2.
[b] Defined as >88 cm for women and >102 cm for men.
[c] Defined as blood pressure ≥130/85 mm Hg or antihypertensive treatment.
[d] Defined as <40 mg/dL for men and <50 mg/dL for women.

a broader population with respect to age and race/ethnicity. If a participant's race/ethnicity was not included in the equation being evaluated, it was coded as the referent value. For example, a Chinese-American participant in MESA was coded as not Hispanic in the San Antonio model and not African-American in the ARIC model. Similarly, the ages of MESA participants were used in the San Antonio and ARIC models even if their ages were outside the range of participants included in those studies (age is not a component of the Framingham model). Since MESA did not assess Latino subgroups, we coded all Latinos from MESA as Mexican-American when applying the San Antonio model. For the ARIC and San Antonio diabetes risk prediction models, these variables are incorporated into a logistic regression model used to calculate the probability of developing diabetes.

Study outcome

The outcome for the current study was the incidence of diabetes. Incident diabetes was determined using data collected during 3 in-person MESA follow-up examinations. The incidence of diabetes was defined as the first visit at which a MESA participant self-reported use of oral hypoglycemic drugs or insulin or had a fasting serum glucose level greater than or equal to 126 mg/dL.

Statistical analysis

The mean level and prevalence of each component of each diabetes risk model was calculated overall and by ethnicity. The probability of developing diabetes over a period of 4.75 years (the median duration of follow-up of participants in this analysis) was determined for the overall study population and for participants with and without individual components of each risk prediction model, using interval-censored regression models (16). The distributions of the predicted probabilities of diabetes for the Framingham Offspring Study, ARIC Study, and San Antonio Heart Study risk equations were calculated. Since MESA provided a median of 4.75 years of follow-up, the predicted probability for each diabetes risk prediction equation was estimated for this

time period. We achieved this by dividing each MESA participant's predicted probability of developing diabetes by the number of years of follow-up used in each of the prior studies (e.g., 9 years in the ARIC Study) and multiplying this number by 4.75 years. This approach assumes that the risk of diabetes is constant over time. Results were markedly similar when nonconstant rates of diabetes incidence (i.e., logarithmic and exponential rates) were modeled.

The predicted probability of diabetes was divided into quintiles, and probabilities of and hazard ratios for incident diabetes were calculated by quintile using interval-censored regression models. Next, using logistic regression models with incident diabetes as the outcome, we calculated the *c* statistic for each of the diabetes risk prediction models. *C* statistics were calculated for the overall MESA population and for each racial/ethnic group separately. We compared *c* statistics for the diabetes risk equations, pairwise, using the method of DeLong et al. (17) for correlated data. We utilized *t* tests, using Taylor series approximation to calculate standard errors, to determine the statistical significance of differences in *c* statistics across ethnic groups for each prediction equation. We also calculated *c* statistics for individual diabetes risk factors.

To assess calibration of the risk prediction equations, we calculated the observed and predicted incidence rates of diabetes by quintile of predicted risk for each equation and compared them using the Hosmer-Lemeshow goodness-of-fit chi-squared test (18). Next, we calculated the predicted risk of diabetes after recalibrating each equation using the method described by D'Agostino et al. (13). This procedure involves replacing the average diabetes incidence rate from each cohort (i.e., the intercept) with the average diabetes incidence rate from MESA and replacing the mean values of risk equation components from each of the prior studies with the mean values from MESA.

Finally, using the variables in each prediction equation, we generated best-fit models. The best-fit process produces predicted values for the incidence of diabetes as close to the diabetes rates (observed values) as possible. The predicted risks of diabetes after recalibration and using the best-fit model, separately, were compared with the observed incidence of diabetes via the Hosmer-Lemeshow goodness-of-fit chi-squared test. Sensitivity analyses were conducted using a Framingham model which applies an exponential function incorporating a linear combination of each of the variables to estimate the probability of developing diabetes (see Appendix Table 1). Analyses were conducted using Stata, version 9 (Stata Corporation, College Station, Texas) and SAS, version 9.2 (SAS Institute Inc., Cary, North Carolina).

RESULTS

Characteristics of the MESA participants included in the current analysis are displayed overall and by race/ethnicity in Table 1. Overall, the mean age was 61.6 years (standard deviation, 10.2); 46.7% of the participants were male, and the racial/ethnic breakdown was 42.6% white, 25.2% African-American, 11.5% Chinese-American, and 20.7%

Table 2. Probability of Developing Diabetes Over a 4.75-Year Period[a] According to Demographic Factors and Diabetes Risk Factors, Multi-Ethnic Study of Atherosclerosis

	No. of Incident Diabetes Cases	No. of Participants at Risk	Probability[b] of Diabetes, %
Total	446	5,329	8.7
Age, years			
<65	262	3,127	8.4
≥65	184	2,202	9.0
Sex			
Male	207	2,489	8.7
Female	239	2,840	8.6
Race/ethnicity			
White	137	2,272	6.2
African-American	140	1,341	11.0
Hispanic	124	1,101	11.7
Chinese-American	45	615	7.5
Body mass index[c]			
<25	58	1,625	3.7
25–29	159	2,145	7.7
≥30	229	1,559	15.1
Blood pressure, mm Hg			
<130/85 and not receiving treatment	142	2,498	5.8
≥130/85 or receiving treatment	304	2,831	11.3
Parental or sibling history of diabetes			
No	215	3,470	6.5
Yes	231	1,859	12.8
Smoking status			
Never smoker	216	2,692	8.3
Former smoker	179	1,966	9.5
Current smoker	51	671	7.9
High density lipoprotein cholesterol level, mg/dL			
≥40 (men) or ≥50 (women)	224	3,529	6.6
<40 (men) or <50 (women)	222	1,800	12.8
Triglyceride level, mg/dL			
<150	268	3,865	7.2
≥150	178	1,464	12.6
Fasting glucose level, mg/dL			
<100	178	4,512	4.1
100–125	268	817	34.7
Abdominal obesity[d]			
No	125	2,592	5.0
Yes	321	2,737	12.1
Metabolic syndrome			
No	213	3,969	4.5
Yes	233	1,360	19.1

[a] Data were collected during the baseline examination (2000–2002) and at 3 follow-up visits occurring at 18-month intervals.
[b] Probability was calculated for 4.75 years of follow-up using interval-censored methods.
[c] Weight (kg)/height (m)^2.
[d] Defined as >88 cm for women and >102 cm for men.

Table 3. Probabilities and Hazard Ratios for Incident Diabetes in the Multi-Ethnic Study of Atherosclerosis According to Quintile of Predicted Diabetes Risk, Determined Using Risk Prediction Equations From the Framingham Offspring Study, the Atherosclerosis Risk in Communities Study, and the San Antonio Heart Study

Risk Prediction Model	Quintile of Predicted Risk					*P*-Trend
	1	2	3	4	5	
Framingham Offspring Study[a]						
Range of scores	0–2	3–5	6–10	11–13	14–28	
Range of predicted risk[b]	≤4	≤4	≤4	5–6	7–44	
Probability[c], %	3.3	3.3	3.3	6.7	28.9	<0.001
HR (95% CI)	1.0 (referent)	1.0 (referent)	1.0 (referent)	2.32 (1.50, 3.58)	14.3 (9.72, 21.0)	<0.001
Atherosclerosis Risk in Communities Study						
Range of predicted risk[a,b]	0.02–2.9	3.0–5.9	6.0–11.1	11.2–23.4	23.5–102.0	
Probability[c], %	0.8	2.1	3.6	8.3	29.1	<0.001
HR (95% CI)	1.0 (referent)	2.75 (1.14, 6.61)	5.18 (2.24, 12.0)	13.5 (5.91, 30.7)	64.5 (26.7, 156)	<0.001
San Antonio Heart Study						
Range of predicted risk[a,b]	0.02–5.9	6.0–12.2	12.3–22.6	22.7–44.3	44.4–128.1	
Probability[c], %	1.0	1.9	3.8	8.1	29.1	<0.001
HR (95% CI)	1.0 (referent)	1.98 (0.86, 4.46)	4.35 (2.01, 9.43)	10.5 (4.91, 22.4)	51.4 (22.8, 116)	<0.001

Abbreviations: CI, confidence interval; HR, hazard ratio.

[a] The bottom 3 quintiles of risk in the Framingham Offspring Study are equivalent because of the constraints of the point scoring system, in which all persons with a point score of 10 or less are assigned a ≤3% risk over a period of 8 years (which is converted to ≤4 cases per 1,000 person-years). The top quintile is also constrained by a maximum possible risk of 35% over 8 years for any score of 25 or more.

[b] Per 1,000 person-years.

[c] Probability of developing diabetes over a period of 4.75 years (the median length of follow-up available for analysis).

Hispanic. Over the course of 23,995 person-years of follow-up (median, 4.75 years; interquartile range: 4.5–5.0 years), 446 new cases of diabetes were diagnosed. The probability of developing diabetes over 4.75 years of follow-up was 8.7% (Table 2). The probability of developing diabetes was higher among African Americans and Hispanics (vs. whites and Chinese Americans) and among persons with a higher body mass index, elevated blood pressure, a family history of diabetes (parent or sibling), low HDL cholesterol levels, high triglyceride levels, impaired fasting glucose, and the metabolic syndrome. The probabilities of and hazard ratios for diabetes incidence increased by quintile of predicted risk for each equation (Table 3).

Discriminative validity

C statistics for the incidence of diabetes were 0.78 (95% confidence interval (CI): 0.74, 0.82), 0.84 (95% CI: 0.82, 0.86), and 0.83 (95% CI: 0.81, 0.85) for the Framingham, ARIC, and San Antonio diabetes risk models, respectively (Table 4). Discrimination was significantly better for the ARIC and San Antonio models than for the Framingham model ($P < 0.01$ for each pairwise comparison). In sensitivity analyses using a logistic regression equation for the Framingham model rather than the point scoring system, the *c* statistic was 0.81 (95% CI: 0.79, 0.84), which was not significantly different from the *c* statistics derived from the ARIC and San Antonio models ($P > 0.05$). For the Framingham and ARIC models, *c* statistics were similar for all racial/ethnic groups. However, using the San Antonio model, the *c* statistic was significantly lower among African-American participants than among white participants.

Discriminatory value of individual risk factors

The risk prediction components, modeled individually, demonstrated significantly inferior discriminative ability when compared with the full risk prediction models (each $P < 0.01$). For example, the *c* statistic was 0.74 (95% CI: 0.73, 0.76) for impaired fasting glucose, 0.73 (95% CI: 0.71, 0.75) for metabolic syndrome, 0.69 (95% CI: 0.66, 0.71) for waist circumference, 0.65 (95% CI: 0.63, 0.67) for body mass index, and 0.59 (95% CI: 0.56, 0.62) for family history of diabetes.

Model calibration and best-fit models

The Framingham diabetes risk model underestimated the risk of diabetes in all quintiles of predicted diabetes risk (Figure 1 and Appendix Table 2). The ARIC prediction model underestimated risk of diabetes in the highest quintile but was accurate in all other quintiles, while the San Antonio model overestimated diabetes risk in all quintiles. The Hosmer-Lemeshow goodness-of-fit test result was significant for each model ($P < 0.001$). After recalibration, all of the

Table 4. *C* Statistics for Prediction of Incident Diabetes in the Multi-Ethnic Study of Atherosclerosis Using Risk Prediction Equations From the Framingham Offspring Study, the Atherosclerosis Risk in Communities Study, and the San Antonio Heart Study, Overall and By Race/Ethnicity

Risk Prediction Model	Total		Whites		African Americans		Hispanics		Chinese Americans	
	C	95% CI	*C*	95% CI	*C*	95% CI	*C*	95% CI	*C*	95% CI
Framingham Offspring Study	0.78	0.74, 0.82	0.80	0.72, 0.87	0.78	0.70, 0.85	0.75	0.72, 0.78	0.78	0.65, 0.90
Atherosclerosis Risk in Communities Study	0.84	0.82, 0.86	0.86	0.82, 0.90	0.81	0.77, 0.85	0.82	0.78, 0.86	0.83	0.77, 0.89
San Antonio Heart Study	0.83	0.81, 0.85	0.85	0.81, 0.89	0.80	0.76, 0.84*	0.81	0.77, 0.85	0.85	0.79, 0.90

Abbreviation: CI, confidence interval.
* $P < 0.05$ (African Americans vs. whites).

prediction models demonstrated good estimation of diabetes risk (Hosmer-Lemeshow goodness-of-fit test: $P > 0.10$). In addition, using the best-fit model, the predicted incidence rates of diabetes were not statistically significantly different from the observed rates of diabetes, indicating a good model fit for each prediction model (Hosmer-Lemeshow goodness-of-fit test: $P > 0.10$).

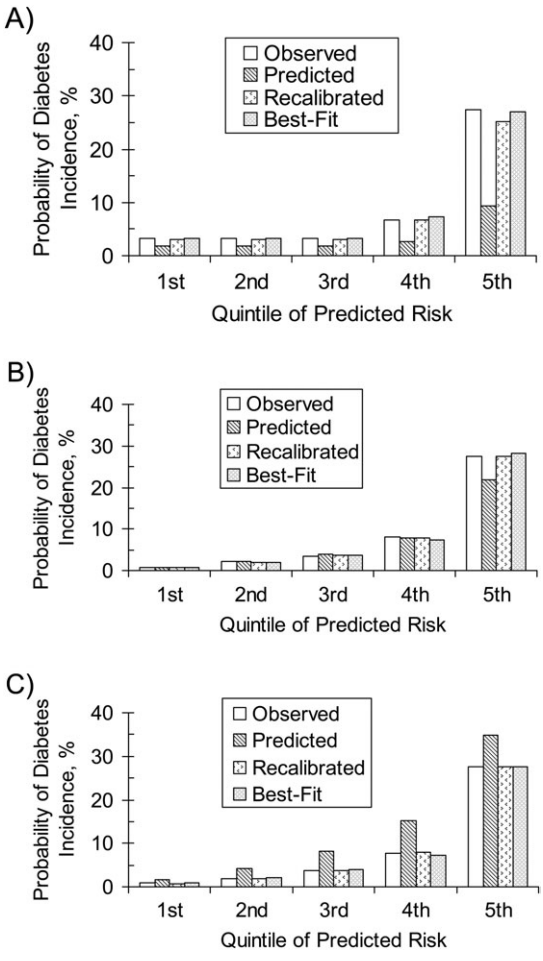

Figure 1. Observed, predicted, recalibrated, and best-fit probabilities of incident diabetes (%) over a 4.75-year period by quintile of predicted risk, calculated using risk prediction equations from A) the Framingham Offspring Study, B) the Atherosclerosis Risk in Communities Study, and C) the San Antonio Heart Study.

DISCUSSION

In the current study, each of 3 previously published diabetes risk prediction models maintained high levels of discriminative validity when applied in a contemporary, multiethnic cohort of US adults. The growing epidemic of diabetes makes this ability to identify persons at high risk critically important. Over the past 2 decades, the prevalence of diabetes risk factors such as obesity has increased substantially (19). Obesity among US adults increased from 13% to 32% between the 1960s and 2004, with 41% predicted to be obese by 2015 (20). It is reassuring that the ability to differentiate relative levels of diabetes risk using these prediction models has not been mitigated by the increasing prevalence of risk factors. The high levels of discriminative ability also support the use of these models among multiethnic populations for identifying subgroups at high risk of diabetes.

While the discriminative ability of impaired fasting glucose was higher than that of all of the other individual risk factors, it did not predict diabetes as well as the risk prediction models. In fact, no single risk factor was as powerful as the full prediction models for discriminating the risk of incident diabetes. This highlights the utility of risk models above estimation based on any single risk factor alone. It also illustrates that readily available data on clinical risk factors can be combined to target high-risk adults for diabetes prevention interventions.

While the Framingham model using a point scoring system performed well, its predictive ability, as judged by the *c* statistic, was significantly inferior to that of the ARIC and San Antonio risk prediction models. However, we remedied this difference by using a logistic equation model rather than the point scoring system. This indicates that the grouping of low- and high-risk participants into single risk estimation categories ($\leq$3% and $>$35%, respectively) and the use of dichotomous predictors led to a small decrement in discriminative ability. This is balanced by the enhanced potential of the point scoring system in current clinical practice.

Although a core set of variables, including fasting glucose, HDL cholesterol, blood pressure, family history, and a measure of adiposity (either body mass index or waist circumference), are common to all 3 equations evaluated in the current study, each prediction model has variations in its components. For example, the ARIC equation includes black race as a variable, while the San Antonio equation includes Mexican ethnicity. The San Antonio equation is

also unique in its inclusion of age and sex, while the ARIC equation is the only one to include waist circumference rather than body mass index. Despite variations in the inclusion of variables, each of the risk models provided similar *c* statistics. This suggests that the core set of risk prediction variables is stable in terms of the variables' discriminative ability, regardless of small alterations in the relative importance of each individual predictor between risk equations.

When individual ethnic groups in MESA were analyzed separately, there were only small differences in the discriminative abilities of the risk prediction models. The continued ability of prediction models derived from mono- and biethnic cohorts to discriminate levels of risk in a multiethnic population that included Chinese Americans as well as additional Hispanic groups beyond Mexicans provides evidence for their applicability in the increasingly diverse US population.

The ARIC prediction model maintained good calibration in the bottom 4 quintiles but underestimated diabetes risk in the highest quintile, while the Framingham and San Antonio models systematically under- or overestimated the observed rates of incident diabetes in MESA. Recalibration remedied this issue for all 3 models, with a more dramatic effect in the highest quintile, as this group had the largest disparity between observed and predicted rates in all 3 models. The need for recalibration is not surprising given the differences in the sample characteristics (time, race/ethnicity, diabetes incidence, etc.) between MESA and the prior cohorts. Using the best-fit models, the predicted rates of diabetes incidence were markedly similar to the observed rates in MESA for each prediction model. The accuracy of the recalibrated and best-fit models suggests that the participant characteristics included in the diabetes risk prediction models have retained their importance in a contemporary, multiethnic cohort, but the underlying incidence of diabetes and the relative importance of each risk factor may have changed over time from those used in the derivation cohorts. Therefore, without recalibration, the ability of these equations to generate an accurate point estimate of an individual's diabetes risk may be inadequate.

The current study had both strengths and limitations. Its strengths included the large, ethnically diverse population, the detailed clinical and metabolic characterization of the cohort, and active follow-up for incident diabetes. MESA is remarkable for its ability to provide in-depth data on sociodemographic, physiologic, and medical characteristics that reflect an ethnically diverse population. Limitations included the inability to isolate persons of Mexican ethnicity in MESA in order to match the San Antonio equation; the shorter follow-up time in MESA as compared with the studies in which the diabetes risk prediction models were developed; and the lack of oral glucose tolerance testing results in MESA, which were used as part of the definition of diabetes in the other cohort studies. In addition, analyses comparing the observed incidence of diabetes with the predicted incidence of diabetes used quintiles rather than deciles, since there were only 446 cases of incident diabetes.

The maximum age at the baseline MESA examination was higher (84 years) than that of the populations included when the risk models were developed (e.g., the maximum age in ARIC was 64 years), and MESA, by design, included a more racially/ethnically diverse population. While this was a major strength of MESA that allowed us to test the performance characteristics of the previous models in a contemporary cohort, it was also a limitation, since these same characteristics (e.g., Chinese ethnicity) were often not accounted for in the original models and their impact on validity could not be fully explored. The mean follow-up in MESA was 4.5 years as compared with 7.5–9 years in the previous models. As Hippisley-Cox and Coupland noted previously (21), the importance of each risk factor may vary over time, and the differential follow-up may explain some of the calibration issues between the previous models and MESA.

Diabetes risk prediction models derived from single or biethnic cohorts, including those from the Framingham Offspring Study, the ARIC Study, and the San Antonio Heart Study, maintained high levels of discriminative ability when applied to a modern, multiethnic cohort. However, the calibration of these models was not adequate, requiring recalibration to avoid significant over- or underestimation of the observed diabetes risks. Clinicians should be encouraged to use these diabetes risk prediction models to stratify their patient populations. However, recalibration is needed before these equations can be used to estimate the risk of diabetes for individual patients.

ACKNOWLEDGMENTS

Author affiliations: Department of Medicine, Division of General Internal Medicine, Mount Sinai School of Medicine, New York, New York (Devin M. Mann); Department of Epidemiology and Prevention, Division of Public Health Sciences, and Department of Biostatistical Sciences, School of Medicine, Wake Forest University, Winston-Salem, North Carolina (Alain G. Bertoni, Haiying Chen); Department of Medicine, Division of General Internal Medicine, Columbia University Medical Center, New York, New York (Daichi Shimbo); Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Mercedes R. Carnethon); Department of Pathology, College of Medicine, University of Vermont, Burlington, Vermont (Nancy Swords Jenny); and Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama (Paul Muntner)

The Multi-Ethnic Study of Atherosclerosis was supported by contracts NO1-HC-95159 through NO1-HC-95165 and contract NO1-HC-95169 from the National Heart, Lung, and Blood Institute. This research was also supported by grant 1K23DK081665, a Patient-Oriented Mentored Scientist Award, from the National Institute of Diabetes and Digestive and Kidney Diseases (Devin M. Mann).

Conflict of interest: none declared.

REFERENCES

1. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care*. 1998;21(9):1414–1431.
2. Fox CS, Pencina MJ, Meigs JB, et al. Trends in the incidence of type 2 diabetes mellitus from the 1970s to the 1990s: the Framingham Heart Study. *Circulation*. 2006;113(25):2914–2918.
3. American Diabetes Association. Executive summary: standards of medical care in diabetes 2008. *Diabetes Care*. 2008; 31(suppl 1):S5–S11.
4. American Diabetes Association and National Institute of Diabetes and Digestive and Kidney Diseases. The prevention or delay of type 2 diabetes. *Diabetes Care*. 2002;25(4): 742–749.
5. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: the Atherosclerosis Risk in Communities Study. *Diabetes Care*. 2005;28(8):2013–2018.
6. Stern MP, Williams K, Haffner SM. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann Intern Med*. 2002;136(8): 575–581.
7. Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med*. 2007;167(10): 1068–1074.
8. Aekplakorn W, Bunnag P, Woodward M, et al. A risk score for predicting incident diabetes in the Thai population. *Diabetes Care*. 2006;29(8):1872–1877.
9. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care*. 2003;26(3): 725–731.
10. Harris MI, Flegal KM, Cowie CC, et al. Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care*. 1998;21(4): 518–524.
11. McGeechan K, Macaskill P, Irwig L, et al. Assessing new biomarkers and predictive models for use in clinical practice: a clinician's guide. *Arch Intern Med*. 2008;168(21):2304–2310.
12. Weiner DE, Tighiouart H, Griffith JL, et al. Kidney disease, Framingham risk scores, and cardiac and mortality outcomes. *Am J Med*. 2007;120(6):552.e1–552.e8.
13. D'Agostino RB Sr, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA*. 2001; 286(2):180–187.
14. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*. 2002; 156(9):871–881.
15. National Cholesterol Education Program (NCEP) Expert Panel on Detection. Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*. 2002;106(25):3143–3421.
16. Allison P. *Survival Analysis Using SAS: A Practical Guide*. Cary, NC: SAS Institute Inc; 1995.
17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845.
18. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982;115(1):92–106.
19. Cowie CC, Rust KF, Ford ES, et al. Full accounting of diabetes and pre-diabetes in the U.S. population in 1988–1994 and 2005–2006. *Diabetes Care*. 2009;32(2): 287–294.
20. Wang Y, Beydoun MA. The obesity epidemic in the United States—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiol Rev.* 2007;29:6–28.
21. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores [electronic article]. *BMJ*. 2009;339:b4229.

APPENDIX

Risk Prediction Models

Framingham Offspring Study point scoring model

As per the original model development paper (7), the Framingham point scoring system was recreated using the following dichotomous risk factors: 10 points for a fasting glucose level of 100–125 mg/dL, 2 points for a body mass index (weight (kg)/height $(m)^2$) of 25–29.9, 5 points for a body mass index of $\geq$30, 5 points for a high density lipoprotein (HDL) cholesterol level of $<$40 mg/dL in men or $<$50 mg/dL in women, 3 points for a parental history of diabetes, 3 points for a triglyceride level of $\geq$150 mg/dL, and 1 point for elevated blood pressure (systolic blood pressure $\geq$130 mm Hg or diastolic blood pressure $\geq$85 mm Hg or use of antihypertensive medication). Points were summed for each participant, and the risk of diabetes over 8 years was calculated as shown in Appendix Table 1.

Appendix Table 1. Point scoring system used for predicting diabetes risk in the Framingham Offspring Study

Point Score	8-Year Risk of Diabetes, %
$\leq$10	$\leq$3
11 or 12	4
13	5
14	6
15	7
16	9
17	11
18	13
19	15
20	18
21	21
22	25
23	29
24	33
$\geq$25	$\geq$35

Framingham Offspring Study logistic regression model

$X = -5.517 + 1.98 \times$ (1 if impaired fasting glucose, else 0) $+ 0.30 \times$ (1 if overweight, else 0) $+ 0.92 \times$ (1 if obese, else 0) $+ 0.94 \times$ (1 if low HDL cholesterol, else 0) $+ 0.58 \times$ (1 if high triglycerides, else 0) $+ 0.50 \times$ (1 if elevated blood pressure, else 0) $+ 0.57 \times$ (1 if family history of diabetes, else 0).

Best-fit Multi-Ethnic Study of Atherosclerosis (MESA)/ Framingham logistic regression model

$X = -4.281 + 2.26 \times$ (1 if impaired fasting glucose, else 0) $+ 0.157 \times$ (1 if overweight, else 0) $+ 0.189 \times$ (1 if obese, else 0) $+ 0.063 \times$ (1 if low HDL cholesterol, else 0) $+ 0.082 \times$ (1 if high triglycerides, else 0) $+ 0.157 \times$ (1 if elevated blood pressure, else 0) $+ 0.211 \times$ (1 if family history of diabetes, else 0).

Atherosclerosis Risk in Communities (ARIC) Study logistic regression model

$X = -9.9808 + 0.0173 \times$ age in years $+ 0.4433 \times$ if black $+ 0.4981 \times 1$ if family history of diabetes is present $+ 0.0880 \times$ fasting glucose in mg/dL $+ 0.0111 \times$ systolic blood pressure in mm Hg $+ 0.0273 \times$ waist circumference in cm $- 0.0326 \times$ height in cm $- 0.0122 \times$ HDL cholesterol in mg/dL $+ 0.00271 \times$ triglycerides in mg/dL.

Best-fit MESA/ARIC logistic regression model

$X = -12.911 - 0.305 \times$ age in years $+ 0.181 \times$ if black $+ 0.578 \times 1$ if family history of diabetes is present $+ 0.119 \times$ fasting glucose in mg/dL $+ 0.006 \times$ systolic blood pressure in mm Hg $+ 0.028 \times$ waist circumference in cm $- 0.015 \times$ height in cm $- 0.009 \times$ HDL cholesterol in mg/dL $+ 0.001 \times$ triglycerides in mg/dL.

San Antonio Heart Study regression model

$X = -13.415 + 0.028 \times$ age in years $+ 0.661 \times$ sex (1 if female, else 0) $+ 0.412 \times 1$ if Mexican $+ 0.079 \times$ fasting glucose in mg/dL $+ 0.018 \times$ systolic blood pressure in mm Hg $- 0.039 \times$ HDL cholesterol in mg/dL $+ 0.070 \times$ body mass index $+ 0.481 \times 1$ if family history of diabetes, else 0.

Best-fit MESA/San Antonio Heart Study regression model

$X = -14.836 - 0.239 \times$ age in years $+ 0.367 \times$ sex (1 if female, else 0) $- 0.129 \times 1$ if Mexican $+ 0.122 \times$ fasting glucose in mg/dL $+ 0.006 \times$ systolic blood pressure in mm Hg $- 0.016 \times$ HDL cholesterol in mg/dL $+ 0.034 \times$ body mass index $+ 0.567 \times 1$ if family history of diabetes, else 0.

The probability of developing diabetes was calculated as $\exp(X)/(1 + \exp(X))$.

Once the probability of diabetes was calculated, it was adjusted to represent 4.75 years of follow-up by dividing the probability by duration of follow-up in each of the prior cohorts (8, 9, and 7.5 years for Framingham, ARIC, and San Antonio, respectively) and multiplying this fraction by 4.75.

Appendix Table 2. Probability[a] (%) of Incident Diabetes in the Multi-Ethnic Study of Atherosclerosis According to Quintile of Predicted Risk, Determined Using Risk Prediction Equations From the Framingham Offspring Study, the Atherosclerosis Risk in Communities Study, and the San Antonio Heart Study

Risk Prediction Model	Quintile of Predicted Risk				
	1	2	3	4	5
Framingham Offspring Study					
Observed	3.2	3.2	3.2	6.6	27.5
Predicted	1.8	1.8	1.8	2.6	9.3
Recalibrated	2.8	2.8	2.8	9.0	27.7
Best-fit	3.2	3.2	3.2	7.4	27.1
Atherosclerosis Risk in Communities Study					
Observed	0.8	2.1	3.5	8.0	27.5
Predicted	0.8	2.1	3.9	7.8	21.8
Recalibrated	0.8	2.0	3.8	7.8	27.6
Best-fit	0.8	2.0	3.7	7.3	28.1
San Antonio Heart Study					
Observed	1.0	1.9	3.7	7.7	27.6
Predicted	1.6	4.2	8.1	15.3	34.8
Recalibrated	0.7	1.8	3.8	7.9	27.6
Best-fit	1.0	2.1	4.0	7.3	27.5

[a] Probability of developing diabetes over a period of 4.75 years.