Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women
The Reynolds Risk Score

Paul M Ridker, MD, MPH
Julie E. Buring, ScD
Nader Rifai, PhD
Nancy R. Cook, ScD

Context Despite improved understanding of atherothrombosis, cardiovascular prediction algorithms for women have largely relied on traditional risk factors.

Objective To develop and validate cardiovascular risk algorithms for women based on a large panel of traditional and novel risk factors.

Design, Setting, and Participants Thirty-five factors were assessed among 24,958 initially healthy US women 45 years or older who were followed up for a median of 10.2 years (through March 2004) for incident cardiovascular events (an adjudicated composite of myocardial infarction, ischemic stroke, coronary revascularization, and cardiovascular death). We used data among a random two thirds (derivation cohort, n = 16,400) to develop new risk algorithms that were then tested to compare observed and predicted outcomes in the remaining one third of women (validation cohort, n = 8,158).

Main Outcome Measure Minimization of the Bayes Information Criterion was used in the derivation cohort to develop the best-fitting parsimonious prediction models. In the validation cohort, we compared predicted vs actual 10-year cardiovascular event rates when the new algorithms were compared with models based on covariates included in the Adult Treatment Panel III risk score.

Results In the derivation cohort, a best-fitting model (model A) and a clinically simplified model (model B, the Reynolds Risk Score) had lower Bayes Information Criterion scores than models based on covariates used in Adult Treatment Panel III. In the validation cohort, all measures of fit, discrimination, and calibration were improved when either model A or B was used. For example, among participants without diabetes with estimated 10-year risks according to the Adult Treatment Panel III of 5% to less than 10% (n = 603) or 10% to less than 20% (n = 156), model A reclassified 379 (50%) into higher- or lower-risk categories that in each instance more accurately matched actual event rates. Similar effects were achieved for clinically simplified model B limited to age, systolic blood pressure, hemoglobin A1c, if diabetic, smoking, total and high-density lipoprotein cholesterol, high-sensitivity C-reactive protein, and parental history of myocardial infarction before age 60 years. Neither new algorithm provided substantive information about women at very low risk based on the published Adult Treatment Panel III score.

Conclusion We developed, validated, and demonstrated highly improved accuracy of 2 clinical algorithms for global cardiovascular risk prediction that reclassified 40% to 50% of women at intermediate risk into higher- or lower-risk categories.

JAMA. 2007;297:611-619

©2007 American Medical Association. All rights reserved.

Author Affiliations: Donald W. Reynolds Center for Cardiovascular Research and the Center for Cardiovascular Disease Prevention (Drs Ridker, Cook, and Buring), Division of Preventive Medicine (Drs Ridker, Buring, and Rifai), and the Division of Cardiovascular Diseases (Dr Ridker), Brigham and Women’s Hospital, Boston, Mass, and the Department of Laboratory Medicine, Children’s Hospital, Boston, Mass (Dr Rifai).

Corresponding Author: Paul M Ridker, MD, MPH, Center for Cardiovascular Disease Prevention, Brigham and Women’s Hospital, 900 Commonwealth Ave E, Boston, MA 02215 (pridker@partners.org).
tional risk determinants at baseline in a
cohort of 24,558 initially healthy US
women who were prospectively fol-
lowed up for a median 10.2 years for in-
cident myocardial infarction, stroke, coro-
nary revascularization, or cardio-
vacular death. In a random subset com-
prising two thirds of these women
(model derivation cohort, n=16,400), we
developed 2 novel algorithms for global
risk prediction. We then tested the ef-
fectiveness of these new prediction mod-
els in the remaining one third of the
women (test validation cohort, n=8158).

METHODS

Study Participants,
Laboratory Evaluation,
and End Point Ascertainment

Study participants were derived from the
Women’s Health Study (WHS), a nation-
wide cohort of US women 45 years and
older free of cardiovascular disease and
cancer at study entry initiated in Septem-
ber 1992.17 Women eligible for the cur-
cent analysis were those who provided
an adequate baseline plasma sample
(n=27,939) and had complete ascertain-
ment of all blood covariates of interest
(n=24,558). Exposure data were collected
for age, race/ethnicity, diabetes, blood
pressure, blood pressure treatment, smok-
ing status, cholesterol treatment, meno-
pausal status, postmenopausal hormone
therapy use, height, weight, alcohol use,
exercise frequency, parental history of
myocardial infarction before age 60 years,
and current multivitamin use. All partici-
ants self-reported race/ethnicity as white,
black, Hispanic American, Asian Ameri-
can, or other. All women were followed
up through March 2004 for a median
period of 10.2 years (interquartile range,
9.7-10.6 years) for incident myocardial
infarction, ischemic stroke, coronary re-
vascularization, and cardiovascular death;
these were adjudicated by an end-points
committee after medical record review.
All study participants provided written in-
fomed consent. The study protocol was
approved by the institutional review board
of Brigham and Women’s Hospital (Bos-
ton, Mass).

All women had baseline plasma
samples, 76% of whom had fasting blood
samples. The plasma samples were mea-
sured in a core laboratory facility for total
cholesterol, HDL-C, low-density lipo-
protein cholesterol (LDL-C), lipoprotein
(a), apolipoproteins A-I and B-100,
hsCRP, sICAM-1, fibrinogen, creatine,
hemoglobin A1c, and plasma hom-
cysteine concentration. The core labor-
atory is certified by the National Heart,
Lung, and Blood Institute/Centers for
Disease Control and Prevention Lipid
Standardization Program. Assay charac-
teristics and coefficients of variation are
available upon request.

Derivation of Novel
Risk Prediction Algorithms

Two thirds of the study participants
(n=16,400) were randomly assigned to
a model derivation data set and one
third (n=8158) were reserved as an in-
dependent validation data set.

Among women allocated to the model
derivation set, the best overall predic-
tion algorithm (model A) was fit using
Cox proportional hazards models. All
available exposure variables and all blood
biomarkers were considered for this ini-
tial model, as were all potential trans-
formations and interactions between
them. Both stepwise selection proce-
dures and multiple additive regression
trees18 were used for variable selection,
assessment for interactions, and model
development. Partial dependence plots
were examined for evidence of interac-
tion, even in the absence of main ef-
tecs. These interaction terms were then
further tested in the Cox models.

The final criterion for inclusion in
model A was minimization of the Bayes
Information Criterion (BIC).19 The BIC
is a likelihood-based measure in which
lower values indicate better fit and in
which a penalty is paid for increasing the
number of variables. Thus, the vari-
ables selected for inclusion should pro-
vide not only the best fit but also a par-
simonious prediction model. The BIC is
not influenced by the number of covar-
iates, so models can be directly compared.

Once variables were selected for
model A, we created a second model
(model B) that was simplified for the
purpose of clinical application and ef-
ficiency. For example, in these data non-
HDL-C [total cholesterol −HDL-C] is
highly correlated with apolipoprotein
B-100 (r=0.87), and HDL-C is highly
correlated with apolipoprotein A-I
(r=0.80).20 Thus, model B substituted
total cholesterol and HDL-C. Simpli-
ified model B also eliminated lipopro-
tein(a) because prior work in this co-
hort has found the predictive utility of
lipoprotein(a) to be limited to those with
extremely high values (>90th percent-
tile) and concomitant hyperlipidemia.21

To allow for direct comparison, the
BIC was calculated using data from the
derivation cohort for models A and B,
as well as for models based exclu-
sively on covariates used in the cur-
rent Adult Treatment Panel III (ATP-
III) risk prediction algorithm7 or in the
Framingham Risk Score,9 but with co-
efficients reestimated in the WHS data.

Testing and Validation of Novel
Risk Prediction Algorithms

Once determined in final form, models
A and B were prospectively tested in the
validation data set of 8158 women. In this
validation stage, 3 global measures were
used to evaluate each prediction model:
Entropy (a likelihood-based function for
dichotomous outcomes for which smaller
values indicate better fit); the Yates slope
(the difference in predicted risk between
cases and noncases for which larger val-
es indicate better fit); and the Brier score
(which computes the sum of squared dif-
fences between the observed outcome
and fitted probabilities and for which
smaller values indicate better concordance
between predicted and observed out-
comes).22,23 Because all women were fol-
lowed up for at least 8 years, observed sta-
tus and predicted risk were evaluated and
compared as of 8 years of follow-up for
all measures.

In addition to these global mea-
sures, we assessed the predictive accu-
ricy of each derived model by looking at
2 components of accuracy: discrimi-
nation and calibration. Discrimina-
tion was evaluated using the C statis-
tic that represents the area under the
receiver operating characteristic curve
(for which larger values indicate bet-
ter discrimination). To assess model calibration (or how closely the predicted probabilities reflect actual risk), the Hosmer-Lemeshow calibration statistic comparing observed and predicted risk was computed based on categories defined by 2% increments in predicted risk.

To compare the performance of models A and B to current risk prediction algorithms, we also computed each of these summary statistics in the test cohort using models limited exclusively to covariates defined in the current ATP-III or Framingham Risk Scores, but with coefficients reestimated in the WHS cohort. We additionally computed each of these summary statistics for predicted outcomes based on formal application of the published ATP-III and Framingham Risk scoring systems as estimated from Framingham data.6,7

### Risk Stratification, Reclassification, and Clinical Application

For ease of interpretation and to address the clinical critical issues of reclassification and risk stratification, we divided all participants in the test cohort into the 10-year risk groups of less than 5%, 5% to less than 10%, 10% to less than 20%, and 20% or higher using covariates currently included in the ATP-III risk prediction model. We then calculated the proportion of participants in either higher- or lower-risk categories using models A or B rather than the covariates in the ATP-III model and then compared observed to predicted events during the follow-up period.

Finally, to mimic clinical practice, we repeated these latter analyses using the published ATP-III risk prediction score to determine 10-year risk groups rather than the refitted model using the ATP-III covariates, because diabetes is considered a coronary risk equivalent in current ATP-III guidelines, this final analysis was restricted to nondiabetic study participants.

Analyses were conducted using SAS version 9.1 (SAS Institute Inc, Cary, NC), SPlus version 7.0 (Insightful Corp, Seattle, Wash), and Treenet version 2.0 (Salford Systems, San Diego, Calif).

### RESULTS

Table 1 shows baseline characteristics and biomarker levels for women in the derivation and validation cohorts.

During follow-up, 504 cardiovascular events occurred in the derivation cohort and 262 in the validation cohort.

#### Table 1. Baseline Clinical Characteristics and Plasma Biomarker Levels for Women Initially Free of Cardiovascular Disease and Cancer in the Model Derivation Cohort and the Model Testing and Validation Cohort

<table>
<thead>
<tr>
<th></th>
<th>Derivation Cohort (n = 16,400)</th>
<th>Validation Cohort (n = 8,158)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR), y</td>
<td>52 (48-58)</td>
<td>52 (49-59)</td>
</tr>
<tr>
<td>Race, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>15,500 (95.2)</td>
<td>7,710 (95.3)</td>
</tr>
<tr>
<td>Black</td>
<td>310 (1.9)</td>
<td>151 (1.9)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>169 (1.0)</td>
<td>82 (1.0)</td>
</tr>
<tr>
<td>Asian</td>
<td>220 (1.4)</td>
<td>123 (1.5)</td>
</tr>
<tr>
<td>Other</td>
<td>77 (0.5)</td>
<td>23 (0.3)</td>
</tr>
<tr>
<td>Smoking status, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>1,865 (11.6)</td>
<td>927 (11.4)</td>
</tr>
<tr>
<td>Past</td>
<td>5,961 (36.4)</td>
<td>3,007 (36.9)</td>
</tr>
<tr>
<td>Never</td>
<td>8,544 (52.1)</td>
<td>4,424 (56.8)</td>
</tr>
<tr>
<td>Height, median (IQR), in</td>
<td>65 (63-66)</td>
<td>65 (63-66)</td>
</tr>
<tr>
<td>Weight, median (IQR), lb</td>
<td>148 (132-170)</td>
<td>148 (132-170)</td>
</tr>
<tr>
<td>Body mass index, median (IQR)</td>
<td>24.9 (22.5-28.3)</td>
<td>24.9 (22.5-28.3)</td>
</tr>
<tr>
<td>Alcohol use, &gt;once/wk, No. (%)</td>
<td>6890 (42.0)</td>
<td>3571 (43.8)</td>
</tr>
<tr>
<td>Exercise, &gt;once/wk, No. (%)</td>
<td>7110 (43.4)</td>
<td>3492 (42.8)</td>
</tr>
<tr>
<td>Blood pressure, median (IQR), mm Hg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>125 (115-135)</td>
<td>125 (115-135)</td>
</tr>
<tr>
<td>Diastolic</td>
<td>80 (70-80)</td>
<td>80 (70-80)</td>
</tr>
<tr>
<td>Risk factors, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>442 (2.7)</td>
<td>238 (2.9)</td>
</tr>
<tr>
<td>History of hypertension</td>
<td>4,061 (24.8)</td>
<td>2,061 (25.3)</td>
</tr>
<tr>
<td>Parental history of MI</td>
<td>2112 (12.9)</td>
<td>1,039 (12.7)</td>
</tr>
<tr>
<td>Menopausal</td>
<td>8,911 (54.4)</td>
<td>4,423 (54.3)</td>
</tr>
<tr>
<td>Medication use, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>7,233 (44.2)</td>
<td>3,523 (43.3)</td>
</tr>
<tr>
<td>Lipid-lowering therapy</td>
<td>531 (3.2)</td>
<td>257 (3.2)</td>
</tr>
<tr>
<td>Current multivitamin use</td>
<td>4,805 (29.7)</td>
<td>2,321 (28.8)</td>
</tr>
<tr>
<td>Cholesterol, median (IQR), mg/dL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>208 (183-235)</td>
<td>208 (184-236)</td>
</tr>
<tr>
<td>LDL-C</td>
<td>121.0 (100.1-144.1)</td>
<td>121.3 (100.9-143.8)</td>
</tr>
<tr>
<td>HDL-C</td>
<td>51.9 (43.1-62.5)</td>
<td>52.2 (43.4-62.5)</td>
</tr>
<tr>
<td>Non-HDL-C</td>
<td>153.9 (128.7-181.6)</td>
<td>153.8 (129.4-181.2)</td>
</tr>
<tr>
<td>Apolipoprotein A-I, median (IQR), mg/dL</td>
<td>148.9 (132.5-167.7)</td>
<td>149.6 (132.7-168.6)</td>
</tr>
<tr>
<td>Apolipoprotein B-100, median (IQR), mg/dL</td>
<td>99.7 (83.5-120.8)</td>
<td>100.1 (84.2-120.8)</td>
</tr>
<tr>
<td>Lipoprotein(a), median (IQR), mg/dL</td>
<td>10.5 (4.4-32.0)</td>
<td>10.7 (4.3-32.6)</td>
</tr>
<tr>
<td>hsCRP, median (IQR), mg/L</td>
<td>2.0 (0.8-4.3)</td>
<td>2.0 (0.8-4.4)</td>
</tr>
<tr>
<td>Fibrogen, median (IQR), mg/dL</td>
<td>349.8 (306.7-402.6)</td>
<td>351.7 (306.0-402.7)</td>
</tr>
<tr>
<td>sICAM-1, median (IQR), ng/mL</td>
<td>343.1 (301.1-394.2)</td>
<td>345.1 (300.0-394.8)</td>
</tr>
<tr>
<td>Homocysteine, median (IQR), µmol/L</td>
<td>10.4 (8.7-12.8)</td>
<td>10.5 (8.8-13.0)</td>
</tr>
<tr>
<td>Creatinine, median (IQR), mg/dL</td>
<td>0.71 (0.63-0.80)</td>
<td>0.71 (0.63-0.80)</td>
</tr>
<tr>
<td>HbA1c†</td>
<td>5.0 (4.8-5.2)</td>
<td>5.0 (4.8-5.2)</td>
</tr>
</tbody>
</table>

*Abbreviations: HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LDL-C, low-density lipoprotein cholesterol; MI, myocardial infarction; sICAM-1, soluble intercellular adhesion molecule 1. 
†Body mass index is calculated as weight in kilograms divided by height in meters squared.

©2007 American Medical Association. All rights reserved.

(Reprinted with Corrections) JAMA, February 14, 2007—Vol 297, No. 6 613
Model Derivation and Development

In the model derivation cohort, 35 potential variables (and all possible interactions between them) were evaluated for model inclusion. Of these, only 9 were included in model A, the best-fitting predictive model with the smallest BIC value; age, systolic blood pressure, current smoking, apolipoprotein B-100, hsCRP, apolipoprotein A-I, parental history of myocardial infarction before age 60 years, and 2 interaction terms, hemoglobin A$_{1c}$ if diabetes was present and lipoprotein(a) level if apolipoprotein B-100 was 100 mg/dL or higher. The β coefficients, standard errors, and P values for each of these covariates in best-fitting model A are shown in Table 2.

Given selection of these 9 variables, some markers, such as homocysteine and sICAM-1, appeared to predict risk, but did not satisfy the BIC criterion for model inclusion. Other notable variables that did not further minimize the BIC once the above variables were taken into account included body mass index, alcohol use, exercise frequency, menopausal status, hormone therapy, fibrinogen, and creatinine.

Table 2 also presents β coefficients, standard errors, and P values for simplified model B, which was otherwise identical to model A, but substituted total and HDL-C for apolipoproteins B100 and A-I, and eliminated the interaction term requiring measurement of lipoprotein(a) if apolipoprotein B-100 was 100 mg/dL or higher.

In the derivation data set, the BIC value for model B (BIC = 9067.5) was not as small as that of the best-fitting model A (BIC = 9039.4), suggesting some loss of predictive ability with clinical simplification. However, model B nevertheless was associated with smaller BIC values than were models based on covariates used in the ATP-III prediction model (BIC = 9098.5) or those based on covariates used in the Framingham Risk Score (BIC = 9161.2). Thus, in the model derivation set, both model A and model B appeared to improve risk prediction over that achieved with currently measured covariates Box.

Model Testing and Validation

Table 3 presents summary statistics regarding the performance of models A and B in terms of predicting risk among the 8158 women in the prospective validation data set. For each prespecified global summary statistic (Entropy, Yates Slope, Brier Score, and C statistic), models A and B provided improvement over prediction models based on covariates used in the ATP-III or Framingham models or when the published ATP-III or Framingham Scores were directly applied. With regard to comparisons of predicted and observed risk, P values for the Hosmer-Lemeshow statistics for model A and B indicated good calibration. Although calibration was suboptimal for the 3 published score models, part of this effect was due to a difference in end-point definition.

Reclassification and Clinical Application

Although formal statistical testing provides a method of evaluating model superiority, we believe the critical issue for clinical application is the proportion of patients reclassified using a new risk algorithm and whether the magnitude of this reclassification is large enough to alter physician behavior with regard to prevention.

To address this issue, Table 4 presents the proportion of women in the validation cohort initially classified as having a 10-year risk of less than 5%, 5% to less than 10%, 10% to less than 20%, and 20% or higher based on ATP-III covariates (with coefficients reestimated in the WHS data) who would be reclassified higher or lower-risk categories by model A and model B. As shown for model A, the proportion of women reclassified was small for those with a 10-year risk of less than 5% (2.5%). However, 43% of all women estimated to be at 5% to less 10% risk or at 10% to less than 20% risk using ATP-III covariates were reclassified to higher or lower clinical risk categories when model A was used instead. Table 4 also shows that actual event rates for model A matched well
with predicted rates in nearly all groups; of the 681 participants reclassified by model A, all but 93 were placed into more accurate risk categories.

Table 5 presents similar analyses for women who did not have diabetes with direct application of the published ATP-III risk score. As shown, about 50% of all women with an estimated 10-year risk for coronary heart disease of 5% to less than 10% or 10% to less than 20% according to ATP-III were reclassified to higher or lower risk categories when model A was used instead. Again, there was excellent matching of actual and predicted rates for model A; of the 722 participants without diabetes who were reclassified by model A, all but 2 were placed into more accurate risk categories.

As also presented in Table 4 and Table 5, similar effects were achieved for clinically simplified model B limited to age, systolic blood pressure, hemoglobin A1c, if diabetic, current smoking, total and HDL-C, hsCRP, and parental history of myocardial infarction before age 60 years. Although the proportion of individuals at intermediate-risk reclassified by model B (30%-45%) was smaller than that of model A (43%-50%), there was still excellent matching of actual to predicted event rates in nearly all groups. For example, of the 647 participants without diabetes in Table 5 who were reclassified by model B, all but 2 were placed into more accurate risk categories. Neither new algorithm added substantive information for women at very low initial risk (<5% 10-year risk based on published ATP-III risk scores).

Examples for Outpatient Clinical Practice: The Reynolds Risk Score

As a practical example, Table 6 provides estimated 10-year risks based on variables in our most parsimonious model (model B, the Reynolds Risk Score) for a 50-year-old women smoker without diabetes with an ATP-III estimated risk of 11.5%. As shown, 10-year risk estimates based on model B range from a low of 4.9% to a high of 18.4% for this hypothetical patient.

With regard to reclassification, as shown in the Figure for a representative population of 100,000 US women without diabetes at intermediate risk (80,000 at 5% to less than 10% and 20,000 at 10% to less than 20% 10-year risk by ATP-III), use of the clinically simplified Reynolds Risk Score would place 13,500 of these women at low risk, 48,500 at low to moderate risk, 32,500 at moderate to high risk, and 5,400 at high risk.

Comment

In this study of 24,558 initially healthy US women followed up for a median of 10.2 years, we developed and validated risk prediction algorithms that reclassified 40% to 50% of women currently predicted to be at intermediate risk into higher- or lower-risk categories and did so with greatly improved accuracy when compared with models based on current ATP-III prediction scores. This effect was present not only for our best-fitting model (model A) but also for a simplified clinical model limited to age, systolic blood pressure, hemoglobin A1c, if diabetic, current smoking, total and HDL-C, hsCRP, and parental history of myocardial infarction before age 60 years (model B, the Reynolds Risk Score).

In addition to providing opportunity for improved risk stratification, we believe these data have clinical implications for the targeting of preventive therapies. In these analyses, large proportions of women with 10-year risk estimates of 5% to less than 10% or of 10% to less than 20% based on current ATP-III risk scores were reclassified at either higher or lower risk of total cardiovascular disease when either of the new algorithms was used. In current US treatment guidelines that take into account the benefits, risk, and cost of lipid-lowering therapy, statins are considered an option for those with 10-year risk estimates of 10% or greater, a more conservative approach taken in Europe typically limits statin therapy to those with 10-year risks of 20% or more. In both settings, application of the models described herein should allow more accurate targeting of statin prescriptions to those patients with the most appropriate level of risk so as to minimize toxicity and maximize benefit and cost efficacy.

We also believe these data provide optimism regarding novel cardiovascular risk factors. In our best-fitting model, hemoglobin A1cs, hsCRP, lipoprotein(a), apolipoproteins A-I and B-100, and parental history were included because each contributed to minimization of the BIC. However, homocysteine, fibrinogen, sICAM-1, and creatinine were not included in our par-
simonious models despite univariate risk associations. Similarly, neither body mass index nor exercise frequency added further prognostic information on overall global risk. By contrast, we observed that glucose control as evaluated by hemoglobin A1c was an effective biomarker in these women that modified the risk associated with diabetes.

Our findings might appear to conflict with a recent report from the Framingham Heart Study in which only marginal utility for novel risk factors was described. Equally important, that analysis relied on data from 1712 women who experienced only 68 vascular events, many of which were coded as heart

| Table 3. Summary Statistics Comparing 2 Novel Risk Prediction Algorithms to Prediction Based on Covariates in the ATP-III and Framingham Scores, and to Direct Application of These Latter Global Risk Algorithms, Based on Data From the Validation Cohort (n = 8158)* |
|-----------------|-------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Global measures | Entropy | 779.8 | 778.0 | 784.2 | 793.0 | 791.4 | 823.3 | 936.9 | 919.6 |
|                | Yates slope, % | 5.74 | 5.49 | 5.13 | 5.13 | 5.13 | 2.58 | 4.75 | 4.65 |
|                | Brier score  | 0.02246 | 0.02243 | 0.02249 | 0.02254 | 0.02253 | 0.02308 | 0.02418 | 0.02396 |
| Discrimination | C statistic | 0.809 | 0.808 | 0.805 | 0.791 | 0.791 | 0.787 | 0.752 | 0.751 |
| Calibration     | Hosmer-Lemeshow P value‡ | .38 | .62 | .46 | .18 | .16 | <.001 | <.001 | <.001 |

Abbreviations: ATP, Adult Treatment Panel; LDL-C, low-density lipoprotein cholesterol.
* Lower values of entropy and Brier score and higher values of Yates Slope and C statistic indicate better fit.
† Including history of diabetes.
‡ A significant value for the Hosmer-Lemeshow statistic indicates a significant deviation between predicted and observed outcomes.

| Table 4. Cardiovascular Risk Reclassification in Validation Cohort Comparing Models A and B to Models Based on Current Adult Treatment Panel III Covariates* |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| ATP-III 10-Year Risk Categories, % | <5% | 5% to <10% | 10% to <20% | ≥20% | Total No. (%) Reclassified |
| No. (%) of participants | 6778 (97.5) | 168 (2.4) | 8 (0.1) | 0 | 6954 176 (2.5) |
| Actual event rate | 1.4 | 9.8 | 0 | 0 |

5% to <10%
| No. (%) of participants | 232 (29.1) | 455 (57.0) | 103 (12.9) | 8 (1.0) | 798 343 (43.0) |
| Actual event rate | 4.3 | 6.2 | 13.6 | 30.2 |

10% to <20%
| No. (%) of participants | 3 (1.0) | 85 (27.2) | 178 (56.9) | 47 (15.0) | 313 135 (43.1) |
| Actual event rate | 0 | 16.7 | 17.7 | 22.2 |

≥20%
| No. (%) of participants | 0 | 0 | 27 (32.1) | 57 (67.9) | 84 27 (32.1) |
| Actual event rate | 0 | 0 | 18.2 | 42.0 |

| 10-Year Risk Simplified Model B† | <5% | 5% to <10% | 10% to <20% | ≥20% | Total No. (%) Reclassified |
| No. (%) of participants | 6837 (98.3) | 117 (1.7) | 0 | 0 | 6954 117 (1.7) |
| Actual event rate | 1.4 | 10.7 | 0 | 0 |

5% to <10%
| No. (%) of participants | 158 (19.8) | 559 (70.0) | 81 (10.2) | 0 | 798 239 (30.0) |
| Actual event rate | 4.9 | 6.4 | 14.0 | 0 |

10% to <20%
| No. (%) of participants | 0 | 54 (17.3) | 221 (70.6) | 38 (12.1) | 313 92 (29.4) |
| Actual event rate | 0 | 20.8 | 15.4 | 28.5 |

≥20%
| No. (%) of participants | 0 | 0 | 21 (25.0) | 63 (75.0) | 84 21 (25.0) |
| Actual event rate | 0 | 0 | 18.4 | 39.1 |

Abbreviation: ATP, Adult Treatment Panel.
*All estimated and observed risks have been extrapolated to 10-year rates (number of events per 100 people per 10 years of observation). Nine missing values are for treatment of hypertension, a variable in the Adult Treatment Panel III model.
†Percentages may not sum to 100 due to rounding.
failure or coronary insufficiency. By contrast, the risk algorithms described herein rely on data from 24,558 women who experienced 766 hard cardiovascular end points. We also note that in a separate Framingham Heart Study analysis addressing the additive value of hsCRP, use of this biomarker alone reclassified 25% of those with ATP-III risks between 5% and 20%, data fully consistent with those presented herein.29

Despite advantages of sample size and power, limitations of our analysis merit discussion. First, because our data are limited to women and our cohort is largely white with a relatively narrow socioeconomic range, care should be taken before generalizing to other populations. We note, however, that all components of models A and B have previously been found to predict cardiovascular risk in men30-34 and that both hsCRP and parental history of vascular disease have previously been shown to predict risk within the Framingham cohort itself.29,35,36

Table 5. Cardiovascular Risk Reclassification in the Validation Cohort of Women Without Diabetes Comparing Models A and B to Models Based on the Published Adult Treatment Panel III Algorithm

<table>
<thead>
<tr>
<th>ATP-III 10-Year Risk Categories</th>
<th>10-Year Risk Best-Fitting Model A†</th>
<th>No. (%) Reclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5%</td>
<td>6803 (95.2)</td>
<td>314 (4.4)</td>
</tr>
<tr>
<td>Actual event rate</td>
<td>1.4</td>
<td>7.0</td>
</tr>
<tr>
<td>5% to &lt;10%</td>
<td>133 (22.1)</td>
<td>303 (50.2)</td>
</tr>
<tr>
<td>Actual event rate</td>
<td>2.9</td>
<td>8.4</td>
</tr>
<tr>
<td>10% to &lt;20%</td>
<td>8 (5.1)</td>
<td>36 (23.1)</td>
</tr>
<tr>
<td>Actual event rate</td>
<td>0</td>
<td>3.7</td>
</tr>
<tr>
<td>≥20%</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATP-III 10-Year Risk Categories</th>
<th>10-Year Risk Simplified Model B†</th>
<th>No. (%) Reclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5%</td>
<td>6836 (95.7)</td>
<td>297 (4.2)</td>
</tr>
<tr>
<td>Actual event rate</td>
<td>1.4</td>
<td>6.9</td>
</tr>
<tr>
<td>5% to &lt;10%</td>
<td>96 (15.9)</td>
<td>336 (55.7)</td>
</tr>
<tr>
<td>Actual event rate</td>
<td>4.0</td>
<td>8.3</td>
</tr>
<tr>
<td>10% to &lt;20%</td>
<td>6 (3.8)</td>
<td>31 (19.9)</td>
</tr>
<tr>
<td>Actual event rate</td>
<td>2.3</td>
<td>4.1</td>
</tr>
<tr>
<td>≥20%</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Abbreviation: ATP, Adult Treatment Panel.
*All estimated and observed risks have been extrapolated to 10-year rates (number of events per 100 people per 10 years of observation).
†Percentages may not sum to 100 due to rounding.

Table 6. Clinical Example: Estimated 10-Year Risk for a 50-Year-Old Smoking Woman Without Diabetes, According to ATP-III or to Clinically Simplified Model B (the Reynolds Risk Score)

<table>
<thead>
<tr>
<th>Blood Pressure, mm Hg</th>
<th>Cholesterol, mg/dL</th>
<th>hsCRP, mg/L</th>
<th>Parental History</th>
<th>ATP-III Model</th>
<th>Simplified Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>0.1</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>0.5</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>1.0</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>3.0</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>5.0</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>8.0</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>10.0</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>20.0</td>
<td>No</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>0.1</td>
<td>Yes</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>0.5</td>
<td>Yes</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>1.0</td>
<td>Yes</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>3.0</td>
<td>Yes</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>5.0</td>
<td>Yes</td>
</tr>
<tr>
<td>155/85</td>
<td>240</td>
<td>35</td>
<td>205</td>
<td>8.0</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Abbreviations: ATP, Adult Treatment Panel; HDL, high-density lipoprotein; hsCRP, high-sensitivity C-reactive protein.
SI conversion factor: To convert cholesterol from mg/dL to mmol/L, multiply by 0.0589.
*Parental myocardial infarction event before age 60 years.

©2007 American Medical Association. All rights reserved.
Second, our data on blood pressure, obesity, and family history were based on self-report. However, the WHS is composed of female health professionals who are known to provide accurate reports of lifestyle factors and health status, including blood pressure and weight.37,38 In addition, self-reported blood pressure, body mass index, and family history have previously been shown in the WHS to be strong predictors of cardiovascular risk, with odds ratios consistent in magnitude with those observed in other major studies.39-41 Regarding parental history, we used a conservative cut point of age younger than 60 years to be consistent with prior findings.42 The inclusion of family history in these algorithms underscores the importance of genetic influences on risk among women; in a recent study of women with low Framingham risk who had premature coronary disease in a first-degree relative, nearly a third had significant subclinical atherosclerosis and 17% had atherosclerotic burden exceeding the 90th percentile.43

Third, following recent recommendations,44 we elected in our analysis to use a combined end point of myocardial infarction, ischemic stroke, coronary revascularization, and cardiovascular mortality. We believe this is an appropriate choice because this end point has typically been used in major cardiovascular clinical trials evaluating interventions for primary prevention, including recent trials of aspirin and statin therapy.

Finally, we limited our analysis to blood-based biomarkers and traditional epidemiological risk factors, in part to ensure a cost-effective approach for primary prevention that could be directly compared with the ATP-III algorithm. These data thus do not examine the potential for atherosclerotic imaging tests to serve as an alternative method for evaluating risk. However, we believe the methods developed herein—variable selection in a derivation data set to minimize the BIC followed by prospective testing in a second validation cohort—should provide a structure for the formal evaluation of emerging risk predictors, including potential imaging tests.

As 8 to 10 million US women have an ATP-III estimated 10-year risk between 5% and 20%, application of these data could have an immediate effect on cardiovascular prevention.45 A user-friendly calculator for the Reynolds Risk Score can be freely accessed at http://www.reynoldsriskscore.org.

**Figure.** Reclassification of Risk Using the Reynolds Risk Score for a Representative Population of 100,000 Intermediate-Risk US Women Without Diabetes

Percentages shown in ovals indicate the proportion of women distributed to risk categories based on Adult Treatment Panel III covariates. Reclassification using the Reynolds Risk Score is based on data shown in Table 5, Model B. CVD indicates cardiovascular disease.
Role of the Sponsor: The funding agencies had no involvement in the design and conduct of the study, the collection, management, analysis, and interpretation of the data, or in the drafting of the manuscript.

REFERENCES
aggerated hypoxemia in HAPE-susceptible participants in our study; they propose the comet-tail technique of chest ultrasonography as a means to test this hypothesis in future studies. Subclinical pulmonary edema in climbers remains controversial and relies on the assumption that an increased closing volume at high altitude indicates increased pulmonary extravascular fluid rather than a non-specific alteration related to exercise or subclinical bronchoconstriction. More important, pulmonary extravascular fluid accumulation may be present in the vast majority of healthy recreational climbers at our study site. It therefore appears unlikely that differences in arterial oxygenation between HAPE-susceptible and HAPE-resistant participants in our study were related to extravascular fluid accumulation, since this phenomenon, if existent, would be expected to occur with similar frequency in both groups. The suggestion to use ultrasound lung comets for the diagnosis and quantification of subclinical extravascular fluid accumulation at high altitude is interesting. However, this method, while potentially promising and easy to perform under field conditions, needs rigorous clinical validation before it can be proposed for this purpose.

Dr Dehnert and colleagues suggest that in the case report we refer to, preventive intake of a calcium channel blocker, rather than surgical correction of the atrial septal defect, may have prevented HAPE on subsequent visits to high altitude, but this is equally speculative. To definitively answer the question of whether PFO is a cause of HAPE would require a study in which HAPE-susceptible participants are exposed to high altitude before and after closure of their PFO.

Dehnert et al also hypothesize that in HAPE-susceptible individuals, an abnormal pulmonary pressure response during normoxic exercise might be more relevant to the patency of the foramen ovale than the pressure increase associated with occasional hypoxic exposure. This is an interesting speculation, but an exaggerated pressure response to normoxic exercise has not been a universal finding, and invasive studies have reported normal rather than exaggerated pulmonary artery pressure responses to this form of exercise in HAPE-susceptible individuals.

Moreover, the magnitude of the increase in pulmonary artery pressure is greater during hypoxic than during normoxic exercise, even at submaximal exercise levels. This suggests that the mechanical forces acting on the foramen ovale are probably higher during hypoxic exercise (eg, climbing at high altitude) than during normoxic exercise, even if maximal; these forces may thus be more relevant for causing its reopening. In line with this concept, we are not aware of any data showing an increased frequency of PFO in athletes performing strenuous normoxic exercise, such as long-distance runners or weightlifters.

Finally, as stated in our article, there is the alternative possibility that, in addition to exaggerated hypoxic pulmonary hypertension and a defective alveolar fluid clearance, a PFO may represent a constitutional anomaly associated with HAPE susceptibility.

Yves Allemann, MD
yves.allemann@insel.ch

Christian Seiler, MD
University Hospital
Swiss Cardiovascular Center Bern
Bern, Switzerland

Urs Scherrer, MD
Claudio Sartori, MD
Department of Medicine and Botnar Center for Extreme Medicine
Centre Hospitalier Universitaire Vaudois
Lausanne, Switzerland

Financial Disclosures: None reported.


CORRECTION