Published in final edited form as:

Circulation. 2012 April 10; 125(14): 1748–1756. doi:10.1161/CIRCULATIONAHA.111.075929.

Comparison of the Framingham and Reynolds Risk Scores for Global Cardiovascular Risk Prediction in the Multiethnic Women's Health Initiative

Nancy R. Cook, ScD¹, Nina P. Paynter, PhD¹, Charles B. Eaton, MD², JoAnn E. Manson, MD, DrPH¹, Lisa W. Martin, MD³, Jennifer G. Robinson, MD, MPH⁴, Jacques E. Rossouw, MD⁵, Sylvia Wassertheil-Smoller, PhD⁶, and Paul M. Ridker, MD¹

¹Brigham & Women's Hospital, Harvard Med School, Boston, MA

²Brown University School of Med, Providence, RI

³George Washington University School of Med, Washington, DC

⁴University of Iowa, Iowa City, IA

⁵The National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

⁶Albert Einstein College of Med, Bronx, NY

Abstract

Background—Framingham-based and Reynolds risk scores for cardiovascular disease (CVD) prediction have not been directly compared in an independent validation cohort.

Methods and Results—We selected a case-cohort sample of the multi-ethnic Women's Health Initiative Observational Cohort, comprising 1722 cases of major CVD (752 MIs, 754 ischemic strokes, and 216 other CVD deaths) and a random subcohort of 1994 women without prior CVD. We estimated risk using the ATP-III score, the Reynolds risk score, and the Framingham CVD model, reweighting to reflect cohort frequencies. Predicted 10-year risk varied widely between models, with 10% or higher risk in 6%, 10%, and 41% of women using the ATP-III, Reynolds, and Framingham CVD models, respectively. Calibration was adequate for the Reynolds model, but the ATP-III and Framingham CVD models over-estimated risk for CHD and major CVD, respectively. After recalibration, the Reynolds model demonstrated improved discrimination over the ATP-III model through a higher c-statistic (0.765 vs. 0.757, $p=0.03$), positive net reclassification improvement (NRI) (4.9%, $p=0.02$) and positive integrated discrimination improvement (IDI) (4.1%, $p<0.0001$) overall, excluding diabetics (NRI=4.2%, $p=0.01$), and in white (NRI=4.3%, $p=0.04$) and black (NRI=11.4, $p=0.13$) women. The Reynolds (NRI=12.9, $p<0.0001$) and ATP-III (NRI=5.9%, $p=0.0001$) models demonstrated better discrimination than the Framingham CVD model.

Conclusions—The Reynolds Risk Score was better calibrated than the Framingham-based models in this large external validation cohort. The Reynolds score also showed improved discrimination overall and in black and white women. Large differences in risk estimates exist between models, with clinical implications for statin therapy.

Correspondence: Nancy R. Cook, ScD, Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston, MA 02215, Tel: 617-278-0796, Fax: 617-264-9194, ncook@rics.bwh.harvard.edu.

Conflict of Interest Disclosures: Dr. Ridker is listed as a co-inventor on patents held by the Brigham and Women's Hospital, Boston, MA, that relate to the use of inflammatory biomarkers in cardiovascular disease that have been licensed to AstraZeneca and Siemens. Drs. Ridker and Manson are listed as co-inventors on a patent held by Brigham and Women's Hospital, Boston, MA, that relates to the use of inflammatory biomarkers in diabetes that has been licensed to AstraZeneca.

Keywords

cardiovascular disease risk factors; models; prediction; risk score; statins

Introduction

Traditional Framingham risk factors of age, hypertension, smoking, diabetes, and total and HDL cholesterol form the basis for the Adult Treatment Panel III (ATP-III) coronary heart disease (CHD) risk prediction model.¹ Cardiovascular risk, however, also relates to family history, markers of inflammation such as high-sensitivity C-reactive protein (hsCRP), and hemoglobin A1c (HbA1c) among diabetics. These additional biomarkers are included in the Reynolds Risk Score, an alternative global risk algorithm developed in 2007 for women² and men.³

Both the Framingham ATP-III and the Reynolds scores have received class I recommendations from the American College of Cardiology and the American Heart Association,⁴ and both scores are endorsed as part of the national guidelines for cardiovascular disease prevention in Canada.⁵ However, to date there has been no direct comparison of these two risk scoring systems in an independent prospective cohort that was not used in the derivation of either score. In addition, a Framingham prediction model has recently been developed for total cardiovascular disease (CVD),⁶ but this has not yet been validated in an external population.

All of these risk models for CVD have been developed primarily among white men and women, with little validation in multi-ethnic populations.^{1,2,6} A Framingham risk model for hard CHD events was validated in subcohorts of black and Native American women, but these included very small numbers of events.⁷ Other studies did not have the same success in validating a Framingham model in various populations.^{8,9} How well these models fit in diverse populations remains to be determined.

To address these issues, we directly examined the clinical performance of the Framingham and Reynolds scores in a case-cohort analysis conducted within the Women's Health Initiative Observational Study (WHI-OS), a multi-ethnic, prospective cohort of more than 90,000 initially healthy postmenopausal American women. Specifically, we directly compared model fit in this independent validation cohort of women for three prediction algorithms: the Framingham score currently used in the ATP-III guidelines,¹ the Reynolds Risk Score,² and the Framingham score for total CVD.⁶ As all three scores were derived in predominantly white populations, the WHI-OS provided the opportunity to address their performance in a multi-ethnic population, and separately in black and white subgroups.

Methods

Women were participants in the WHI-OS¹⁰ and its long term follow-up, the WHI Extension Study. The WHI-OS includes 93,676 ethnically diverse postmenopausal women aged 50 to 79 years recruited between 1994 and 1998 at 40 clinical centers targeting minority groups to obtain a cross-section of the US population.¹¹ Of these, 71,872 had no prior history of myocardial infarction (MI), stroke, revascularization procedures, pulmonary embolism, deep vein thrombosis, peripheral vascular disease, or cancer, and 60,890 additionally had baseline blood specimens and baseline risk factor information.

The WHI Clinical Coordinating Center collected baseline information on sociodemographic characteristics, lifestyle factors, health behaviors, and medical history, including blood

pressure measurements. Diabetes¹² and family history, defined here as MI before age 55 in men and 65 in women, were self-reported. Participants brought current medications to clinic visits to assess medication use.

Self-reported outcome data through September 2008 were confirmed through medical record review by centrally trained physicians.¹³ MI and coronary death were combined for the CHD outcome. Medical records, electrocardiogram readings, and cardiac enzyme and troponin determinations were used for confirmation. Strokes were defined as rapid onset of a persistent neurologic deficit attributed to an obstruction or rupture of the brain arterial system, lasting more than 24 hours and without evidence of other cause. These were classified as ischemic or hemorrhagic through review of brain imaging study reports. Underlying cause of death was classified on the basis of death certificates, medical records, and other records such as autopsy reports. The primary endpoint for this analysis is a combined endpoint of major CVD, including CHD, ischemic stroke, and death due to cardiovascular causes. This project has been approved by the Institutional Review Board at the Brigham and Women's Hospital, Boston, MA.

Sample selection

Because of the large size of the WHI, to reduce costs of biochemical assays a prospective case-cohort design¹⁴ was employed in this WHI substudy. To maximize efficiency for examining non-whites, selected cases included all eligible CVD cases from black (n=200), Hispanic (n=53), and Asian (n=55) women, and women with other/unknown ethnicity (n=55). For efficiency, the remaining 1637 of 2000 cases were randomly selected from 2370 cases among white women.

A subcohort of approximately 2000 women was selected using the same eligibility criteria and stratified to match cases by race/ethnicity and five-year age groups. Further exclusion for this analysis of those with other prior CVD conditions, including transient ischemic attack (TIA), CVD surgery, or congestive heart failure (CHF) led to a final sample size of 1,722 cases and a subcohort of size 1,994, of whom 121 were also cases. Among those in the subcohort who did not develop CVD, the median (25%, 75%) follow-up time was 9.9 (8.6, 11.8) years.

For women in the selected samples, blood specimens collected and stored at study entry were assayed centrally for total cholesterol, HDL cholesterol, hsCRP, and HbA1c (among diabetics) using standardized procedures. The core laboratory is certified by the National Heart, Lung, and Blood Institute/Centers for Disease Control and Prevention Lipid Standardization Program.

Statistical methods

The data were analyzed throughout as a case-cohort study^{14,15} using appropriate weighting of the observations.¹⁶ Because the numbers in the full sample were known, our stratified sampling enabled us to mimic or recapture the characteristics of the full WHI cohort using reweighting by the sampling frequency. Overall population characteristics were estimated using inverse probability weights in Proc Survey means in SAS 9.2.¹⁷ To first verify the associations of risk factors within the WHI sample, weighted Cox regression¹⁸ was used to estimate hazard ratios using Proc Phreg of SAS, and asymptotic variance estimates were computed following Langholz and Jiao.¹⁹ Continuous risk factors were treated in a continuous fashion as well as in clinical risk categories.

Predicted values for CHD and CVD were obtained using published equations from the Framingham risk scores for CHD¹ and CVD⁶ and from the Reynolds models for CVD.²

Framingham risk factors include age, blood pressure, antihypertensive treatment, smoking, diabetes, and total and HDL cholesterol. The ATP-III model is intended for those without diabetes, which is considered a risk equivalent. The Reynolds Risk Score additionally includes hsCRP, family history of premature MI (before age 60), and hemoglobin A1c among diabetics only. The fit of the models in the WHI data was examined using appropriate weighting. The c-statistic for survival data^{20,21} was computed and differences between models assessed using bootstrapping. Calibration plots were used to compare observed and predicted risk within deciles using inverse sampling weights. In addition, for completeness, we also considered the alternative Framingham simple model which uses body mass index (BMI) instead of lipids (available at <http://www.framinghamheartstudy.org/risk/gencardio.html#>).

Because different endpoint definitions were used in development of each of the three models, recalibration was necessary to compare models using reclassification methods. The Framingham ATP-III score predicts 'hard' CHD, defined as MI and coronary death,¹ while the Reynolds Risk Score predicts a composite CVD outcome defined as incident MI, ischemic stroke, coronary revascularization, and cardiovascular death.² The Framingham CVD score predicts 'total' CVD defined as all coronary events (MI, coronary death, coronary insufficiency, and angina), cerebrovascular events (including ischemic stroke, hemorrhagic stroke, and TIA), peripheral artery disease (intermittent claudication), and heart failure.⁶ As described above, the major CVD endpoint used in these WHI data comprised CHD, ischemic stroke, and CVD death, and did not match any of these precisely. To correct for differences in endpoint definition, after initial evaluation of fit, models were recalibrated to the WHI cohort using logistic regression calibration for 10-year risk.^{22,23} This process does not change the coefficients for risk factors, but changes the intercept only, to alter the mean predicted risk. The average predicted risk for each model then approximately equaled the overall incidence of major CVD at ten years in the WHI cohort of eligible women.

We used plots to examine the calibration of original and recalibrated models. These plot the average predicted risk within deciles against the observed risk in that decile, adding a reference line for perfect calibration. To directly compare recalibrated models, we examined the integrated discrimination improvement (IDI)²⁴ and reclassification statistics,²⁵ including reclassification calibration chi-squares,²⁶ and net reclassification improvement (NRI).²⁴ For the reclassification tables, clinical-based cut points of 5, 10, and 20% were used.²⁷ Survival methods were used throughout,^{28,29} and measures were reweighted to reflect the distribution in the overall cohort. Statistical tests of reweighted measures were derived using bootstrap samples. Models were also examined among women without diabetes, eliminating those on statins or other cholesterol-lowering medications (7%), and separately among white and black women.

Results

Of 1,722 incident cardiovascular cases, 752 were MIs, 754 ischemic strokes, and 216 CVD deaths. Baseline characteristics for cases and the subcohort are shown in Table 1 both in the selected subsample and reweighted to the population distribution. In the subcohort sample the average age was 69 years with 5% current smokers and 5% diabetics. Reweighted population estimates were age 62 years, 6% current smokers, and 4% diabetics. Cases included more smokers and diabetics, and generally higher risk factor levels. Reweighted distributions in the subcohort are shown by race/ethnicity in Supplemental Table 1. Blacks were slightly younger than whites, but had higher proportions of smokers and diabetics.

Risk factor associations

Multivariable Cox regression models confirmed the association of risk factors with CVD, both overall and among whites and blacks (Table 2). Each risk factor had significant associations in the overall sample except for total cholesterol, which was the same after excluding women on cholesterol-lowering medications. When examined separately using CHD as an endpoint, total cholesterol was a significant predictor (hazard ratio (HR) for total cholesterol $\geq 240 = 1.30$, 95% confidence interval (CI) = 1.01–1.66). Estimated effects of each risk factor were generally consistent for whites and blacks, although results were more variable among blacks due to smaller numbers. The effect of measured blood pressure was weaker but anti-hypertensive medication stronger among blacks, and the effect of total cholesterol was stronger among black women. Age, diabetes, smoking, HDL-cholesterol, hsCRP, and family history all had significant independent effects on major CVD. Regressions using continuous versions of the risk factors showed similar results (Supplemental Table 2). Interactions of age and total cholesterol and of smoking and age included in the ATP-III model were not significant in these data, and are not included in these models.

Predicted risk and model fit

Predicted 10-year risk was estimated using published equations for each model. These varied widely between models, as shown in the distributions in Figure 1A. Average risk was 3.8%, 4.6%, and 10.9% for the ATP-III, Reynolds, and Framingham CVD models. Estimated risk was 10% or higher in 5.5%, 10.3%, and 41.1% of women, respectively (Figure 1B), and risk was 20% or higher in 0.5, 2.6, and 10.6% of women in the three models.

To assess calibration, predicted risk for the ATP-III score was compared to observed 10-year rates of CHD among non-diabetics only, to match its intended use (Figure 2A). As shown, the ATP-III model over-estimated risk of CHD, with predicted values higher than those observed. In contrast, the CHD score was actually better calibrated to the risk of major CVD among all women (Figure 2B). The Reynolds risk score appeared relatively well-calibrated for the endpoint of major CVD (Figure 2C). The Framingham CVD model, developed for a broader definition of CVD, greatly over-estimated risk of major CVD as anticipated (Figure 2D). Similar patterns of over-estimation of risk were seen for the ATP III model for CHD and the Framingham model for CVD among blacks (Figure 3) and whites (Supplemental Figure 1) separately, as the predicted risk was higher than the observed risk in each group.

In order to directly compare discrimination of the three models for the endpoint of CVD, the models were recalibrated such that the average predicted risk equaled the overall reweighted population estimate of 4%. The percent of women with estimated risk of 10% or higher was then 6.6%, 7.7%, and 6.2% for the ATP-III, Reynolds, and Framingham CVD model. After recalibration of the mean, all three models showed reasonable calibration to the major CVD endpoint (Supplemental Figure 2). C-statistics for major CVD (unaffected by recalibration) were 0.757 for the ATP-III model, 0.765 for the Reynolds model, and 0.750 for the Framingham CVD model. Differences in the c-statistics, though small, were all statistically significant (Table 3). When the simple non-lab based Framingham CVD model was considered, the calibration was very similar to that of the lab-based model. However, the discrimination was worse, with a c-statistic of 0.747, which was significantly lower than that for the Reynolds model ($p=0.0004$), but more similar to the lab-based Framingham CVD model ($p=0.57$) and the ATP-III model ($P=0.055$).

Risk reclassification

Fit of the recalibrated models was directly compared using reclassification measures (Table 3). Reclassification tables comparing the Reynolds to both Framingham models using both the crude and recalibrated predicted values are shown in Supplemental Tables 3A–D. These show the numbers of women in the total WHI cohort who would be reclassified into the various risk groups. For example, using the original scores in Table 3A, 664 (23%) of the 2,943 women at ATP III risk of 10–<20% would be reclassified as $\geq 20\%$ risk and 544 (18%) as less than 10% risk. In addition, of the 10,620 women originally at 5–<10% risk, 576 (5%) would be reclassified above 20% and 3,028 (29%) above 10%. Compared to the recalibrated ATP-III model, the Reynolds model showed improvement in discrimination based on the NRI (4.9%, 95% CI=1.2–8.7%, $p=0.010$), with improvement of 4.0% ($p=0.02$) in cases, and 1.0% ($p=0.30$) in non-cases. Improvement was also seen using the IDI (4.1%, 95% CI=2.7–5.7%, $p<0.0001$; relative IDI=117.3%), the continuous NRI (32.1%, 95% CI=24.2–39.5%, $p<0.0001$), and reclassification calibration chi-squares (273.5 vs. 185.7), indicating that calibration improved with the Reynolds model. When comparing the Framingham CVD model to the ATP-III model, the former provided a worse fit even though it was developed for the endpoint of CVD rather than CHD. The NRI and IDI were negative, calibration worse, and c-statistic lower (Table 3). Comparison of the Reynolds to the Framingham CVD model indicated improved fit on all measures, with an NRI of 12.9% (95% CI=9.4–16.3%, $p<0.0001$). Comparison of the Reynolds model to the simple non-lab based Framingham model was very similar with an NRI of 11.9% ($p<0.0001$).

Model fit statistics for recalibrated models were similar among whites only (Table 3). Although not always significant due to smaller numbers, the direction of effects was the same and often stronger among black women, with the Reynolds model showing improved fit over the ATP-III score, and both providing better fit than the Framingham CVD score.

Since the ATP-III model was intended for those without diabetes, we also recalibrated the models separately including only women without diabetes at baseline (Table 3). Although the c-statistics and reclassification calibration chi-squares were similar for the ATP-III and Reynolds models, the NRI showed a significant improvement favoring the Reynolds model (4.2%, 95% CI=1.1–7.5%, $p=0.01$), as did the continuous NRI and the IDI (both $p<0.0001$). Finally, results were very similar after excluding women on cholesterol-lowering medications (Supplemental Table 4).

Discussion

This validation study prospectively examined the fit of three cardiovascular risk prediction models in a large-scale external cohort, the WHI-OS, a national, geographically representative, racially and ethnically diverse sample of US women. Models included the Framingham-based ATP-III model,¹ the Reynolds Risk Score, and the Framingham CVD model.⁶ The published ATP-III model was poorly calibrated for CHD, and the Reynolds Risk Score was better calibrated for major CVD than the Framingham CVD model. Moreover, after recalibration, the Reynolds Risk Score continued to show statistically significant but modest improvement in fit when compared to either of the Framingham-based models.

We know of no prior work directly comparing the Framingham and Reynolds scores in an independent prospective cohort. However, the findings here that family history, inflammation, and HbA1c among diabetics improve global risk prediction are consistent with observations made in other settings considering each factor in isolation.^{30–32} Further,

the lack of effect modification in our data by ethnicity suggests that these findings are clinically relevant in ethnically diverse populations.

Care must be used when interpreting these data, as the endpoints used to generate the three scores differ from each other and from the primary endpoint of the WHI-OS, which could lead to lack of calibration. However, the ATP-III score over-estimated risk of its intended endpoint, 'hard CHD' among non-diabetics, a finding replicated in the Women's Health Study (data available on request). The WHI-OS sample is older than that of Framingham^{6,7} and traditional risk markers are generally less predictive in older age ranges,^{33,34} although the ATP-III score includes interactions of age with smoking and total cholesterol. One Framingham model has previously been validated in six prospective, ethnically diverse cohorts⁷ and was shown to have reasonable calibration overall, but prior data specifically for the ATP-III model are sparse. Other, primarily European, investigators have also reported that various versions of Framingham models over-predict CHD incidence,^{8,9,35,36} Since clinicians in practice cannot re-calibrate a score for an individual patient, population calibration is essential.

The Framingham CVD model was also poorly calibrated for the endpoint of major CVD. However, it was developed for the broader endpoint of total CVD including several other conditions, namely angina, coronary insufficiency, TIA, peripheral artery disease, and CHF. As statins reduce stroke as well as CHD, scoring systems including stroke have recently been favored,³⁷ whereas the inclusion of other endpoints such as CHF are more controversial. In this research setting, re-calibration allowed us to minimize differences in outcomes and directly compare risk scores for a comparable endpoint. As shown in the re-calibrated analyses, the Framingham CVD score, developed using an endpoint closer to that used here, was not superior to the ATP-III score for CHD only, and both of these discriminated less well than the Reynolds score. It is possible that the lower levels of discrimination seen with the Framingham CVD model may reflect reduced effects of the traditional risk factors on these secondary endpoints.

More importantly, the number of women potentially eligible for statin therapy can vary greatly depending on the equation and endpoint used. As shown in Figure 1, the estimated distributions of predicted risks in the population vary widely across models. Using the WHI data, in a hypothetical population of 100,000 women the number who would be classified at 10% or higher risk would be 5,549 with the published ATP-III model, 10,304 with the Reynolds score, and 41,074 with the Framingham CVD model (Supplemental Tables 3AC). Even if all models were perfectly calibrated for their respective outcomes, the choice of endpoints needs to be addressed. Whether statins are equally effective for the various CVD endpoints, and whether the risk-benefit equation is the same for all endpoints, are important criteria in developing guidelines for therapy.

We believe these data may have clinical implications for CVD prevention in otherwise healthy middle-aged women. Risk prediction algorithms have been widely used to better target cardiovascular preventive therapies, in particular the use of statins. A recent meta-analysis including 13,154 women in primary prevention statin trials reported a 37% reduction in cardiovascular event rates.³⁸ However, the great majority of women destined to suffer a cardiovascular event have ATP-III scores less than 10 percent. These women would not qualify for treatment under current guidelines yet could benefit from statin therapy.³⁹ Using data from the WHI-OS, among women with 10-year ATP-III risks of 5 to 10 percent, the Reynolds score would reclassify 15% to a lower risk category (< 5%) and over 28% to a higher risk category (> 10%), including 5% with estimated risk exceeding 20% (Supplemental Table 3A).

Limitations of our analysis merit consideration. First, the WHI-OS is composed exclusively of women, so these results cannot be generalized to men. However, prior work in the Physicians Health Study has shown that the Reynolds Risk Score for men improves fit and reclassification in that setting.³ Second, there may be remaining differences in endpoint definition. Confirmation procedures for Framingham, WHS, and the WHI-OS may have differed somewhat even for common endpoints such as CHD. Third, we were not able to fully assess the calibration of the Framingham CVD model since we did not have data available on the broader definition used for that model. Fourth, estimates from this case-cohort sample may offer less precision than those based on the full WHI cohort.

In sum, in this large scale comparison of risk prediction models commonly used in North America, the Reynolds Risk Score significantly improved fit as compared to either the Framingham-based ATP-III CHD risk score or the newer Framingham CVD score. Within the WHI-OS, the greatest impact of the Reynolds Risk Score appeared to be among those women with 5 to 10 percent 10-year estimated risk according to ATP-III, a group including a large number of women destined to suffer MI or stroke, and in whom trial data indicate efficacy of statin therapy in reducing cardiovascular events.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Women's Health Initiative Investigators: A full listing of Women's Health Initiative investigators can be found at http://whiscience.org/publications/WHI_investigators_longlist. **Additional Contributions:** We thank the Women's Health Initiative investigators, staff, and study participants for their outstanding dedication and commitment. We also thank Dr. Bryan Langholz for advice concerning inference for the case-cohort sample.

Funding Sources: This project was supported by National Heart, Lung, and Blood Institute's Broad Agency Announcement contract number HHSN268200960011C. The Women's Health Initiative program is funded by the National Heart, Lung, and Blood Institute, the National Institutes of Health, and the US Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-9, 32122, 42107-26, 42129-32, and 44221.

References

1. Executive Summary of the 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). *JAMA*. 2001; 285:2486–2497. [PubMed: 11368702]
2. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women. *JAMA*. 2007; 297:611–619. [PubMed: 17299196]
3. Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. *Circulation*. 2008; 118:2243–2251. [PubMed: 18997194]
4. Greenland P, Alpert JS, Beller GA, Benjamin EJ, Budoff MJ, Fayad ZA, Foster E, Hlatky MA, Hodgson JM, Kushner FG, Lauer MS, Shaw LJ, Smith SC Jr, Taylor AJ, Weintraub WS, Wenger NK, Jacobs AK. for the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circ*. 2010; 122:e584–636.
5. Genest J, McPherson R, Frohlich J, Anderson T, Campbell N, Carpentier A, Couture P, Dufour R, Fodor G, Francis GA, Grover S, Gupta M, Hegele RA, Lau DC, Leiter L, Lewis GF, Lonn E, Mancini GB, Ng D, Pearson GJ, Sniderman A, Stone JA, Ur E. 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of

- cardiovascular disease in the adult - 2009 recommendations. *Can J Cardiol.* 2009; 25:567–569. [PubMed: 19812802]
6. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circ.* 2008; 117:743–753.
 7. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA.* 2001; 286:180–187. [PubMed: 11448281]
 8. Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, Ebrahim S. Predictive accuracy of the Framingham risk score in British men: prospective cohort study. *BMJ.* 2003; 327:1–6. [PubMed: 12842922]
 9. Diverse Populations Collaborative Group. Prediction of mortality from coronary heart disease among diverse populations: is there a common predictive function? *Heart.* 2002; 88:222–228. [PubMed: 12181209]
 10. Langer RD, White E, Lewis CE, Kotchen JM, Hendrix SL, Trevisan M. The Women's Health Initiative Observational Study: Baseline characteristics of participants and reliability of baseline measures. *Ann Epidemiol.* 2003; 13:S107–121. [PubMed: 14575943]
 11. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Contr Clin Trials.* 1998; 19:61–109.
 12. Margolis KL, Qi L, Brzyski R, Bonds DE, Howard BV, Kempainen S, Liu S, Robinson JG, Safford MM, Tinker LT, Phillips LS. for the Women's Health Initiative Investigators. Validity of diabetes self-reports in the Women's Health Initiative: comparison with medication inventories and fasting glucose measurements. *Clin Trials.* 2008; 5:240–247. [PubMed: 18559413]
 13. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, Johnson KC, Proulx-Burns L, Pastore L, Criqui M, Daugherty S. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol.* 2003; 13:S122–128. [PubMed: 14575944]
 14. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biomka.* 1986; 73:1–11.
 15. Barlow WE. Robust variance estimation for the case-cohort design. *Biom.* 1994; 50:1064–1072.
 16. Borgan Ø, Langholz B, Samuelson SO, Goldstein L, Pogoda J. Exposure stratified case-cohort designs. *Lifetime Data Analy.* 2000; 6:39–58.
 17. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Amer Statist Assoc.* 1952; 47:663–685.
 18. Therneau TM, Li H. Computing the Cox model for case cohort designs. *Lifetime Data Analy.* 1999; 5:99–112.
 19. Langholz B, Jiao J. Computational methods for case-cohort studies. *Comp Stats Data Analysis.* 2007; 51:3737–3748.
 20. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* 1982; 247:1543–1546.
 21. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med.* 2004; 23:2109–2123. [PubMed: 15211606]
 22. Janssen KJM, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KGM. A simple method to adjust clinical prediction models to local circumstances. *Can J Anesth.* 2009; 56:194–201. [PubMed: 19247740]
 23. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* 2004; 23:2567–2586. [PubMed: 15287085]
 24. Pencina MJ, D'Agostino RBS, D'Agostino RBJ, Vasan RS. Evaluating the added predictive ability of a new biomarker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008; 27:157–172. [PubMed: 17569110]
 25. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circ.* 2007; 115:928–935.

26. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med.* 2009; 150:795–802. [PubMed: 19487714]
27. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med.* 2006; 145:21–29. [PubMed: 16818925]
28. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: Extension to survival analysis. *Stat Med.* 2010; 30:22–38. [PubMed: 20827726]
29. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011; 30:11–21. [PubMed: 21204120]
30. American Diabetes Association. Diagnosis and classification of Diabetes Mellitus. *Diab Care.* 2010; 33:S62–S69.
31. Chow CK, Islam S, Bautista L, Rumboldt Z, Yusufali A, Xie C, Anand SS, Engert JC, Rangarajan S, Yusuf S. Parental history and myocardial infarction risk across the world: the INTERHEART Study. *J Am Coll Cardiol.* 2011; 57:619–627. [PubMed: 21272754]
32. The Emerging Risk Factors Collaboration. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet.* 2010; 375:132–140. [PubMed: 20031199]
33. Beer C, Alfonso H, Flicker L, Norman PE, Hankey GJ, Almeida OP. Traditional risk factors for incident cardiovascular events have limited importance in later life compared with the Health in Men Study Cardiovascular Risk Score. *Stroke.* 2011; 42:952–959. [PubMed: 21311063]
34. Koller MT, Steyerberg EW, Wolbers M, Stijnen T, Bucher HC, Hunink MGM, Witteman JCM. Validity of the Framingham point scores in the elderly: results from the Rotterdam study. *Am Heart J.* 2007; 154:87–93. [PubMed: 17584559]
35. Barroso LC, Muro EC, Herrera ND, Ochoa GF, Hueros JI, Buitrago F. Performance of the Framingham and SCORE cardiovascular risk prediction functions in a non-diabetic population of a Spanish health care centre: a validation study. *Scand J Prim Health Care.* 2010; 28:242–248. [PubMed: 20873973]
36. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. *BMJ.* 2010; 340:c2442. [PubMed: 20466793]
37. Mosca M, Benjamin EJ, Berra K, Bezanson JL, Dolor RJ, Lloyd-Jones DM, Newby LK, Piña IL, Roger VL, Shaw LJ, Zhao D. Effectiveness-based guidelines for the prevention of cardiovascular disease in women 2011 update: a guideline from the American Heart Association. *Circ.* 2011; 123:1243–1262.
38. Mora S, Glynn RJ, Hsia J, MacFadyen J, Genest J, Ridker PM. Statins for the primary prevention of cardiovascular events in women with elevated high-sensitivity C-reactive protein or dyslipidemia. *Circ.* 2010; 121:1069–1077.
39. Ridker PM, Macfadyen JG, Nordestgaard BG, Koenig W, Kastelein JJ, Genest J, Glynn RJ. Rosuvastatin for primary prevention among individuals with elevated high-sensitivity c-reactive protein and 5% to 10% and 10% to 20% 10-year risk. Implications of the Justification for Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial for “intermediate risk”. *Circ Cardiovasc Qual Outcomes.* 2010; 3:447–452. [PubMed: 20736443]

Figure 1. Distribution of risk estimated from the Framingham ATP-III score, the Reynolds risk score, and the Framingham CVD score among women in the WHI (A), and the estimated percent of women with risk of 10% or greater and 20% or greater using the published scores (B).

Figure 2.

Calibration plots for the original published risk prediction scores, including the ATP-III score and the coronary heart disease (CHD) outcome among only those without diabetes (**A**), and the ATP-III score and the cardiovascular disease (CVD) outcome (**B**), the Reynolds risk score and CVD (**C**), and the Framingham CVD score and CVD (**D**) among all women.

Figure 3.

Calibration plots for the original published risk prediction scores among black women only, including the ATP-III score and the coronary heart disease (CHD) outcome among those without diabetes (**A**), and the ATP-III score and the cardiovascular disease (CVD) outcome (**B**), the Reynolds risk score and CVD (**C**), and the Framingham CVD score and CVD (**D**) among all black women.

Table 1

Baseline characteristics in ischemic CVD cases and subcohort members, among those with no prior CVD (including TIA, CHF, CVD surgery), crude and weighted to the population distribution.*

Risk Factor	Subcohort	All Cases	MI	Stroke	CVD Death
N	1994	1722	752	754	216
Crude					
Age (yrs)	69 (63, 73)	69 (64, 73)	68 (63, 73)	69 (64, 73)	70 (64, 74)
Current Smoking (%)	4.9	8.9	9.0	7.8	12.0
Diabetes (%)	4.7	10.7	13.0	9.4	7.4
HbA1c among diabetics (%)	7.0 (6.2, 7.9)	7.5 (6.8, 8.9)	7.4 (6.5, 8.6)	7.7 (6.8, 9.6)	7.5 (7.0, 8.4)
Systolic blood pressure (mmHg)	128.0 (117, 140)	134 (122, 148)	133 (121, 147)	135 (123, 149)	133 (120, 142)
Hypertension medication (%)	26.6	38.3	38.3	38.6	37.0
Total Cholesterol (mg/dl)	225 (200, 256)	226 (198, 253)	229 (202, 258)	221 (196, 248)	227 (202, 252)
HDL Cholesterol (mg/dl)	54.4 (44.6, 66.6)	48.6 (39.8, 59.8)	48.8 (39.8, 60.1)	47.9 (38.8, 58.6)	50.5 (42.3, 62.0)
C-reactive protein (mg/L)	2.3 (1.0, 5.0)	3.1 (1.4, 6.2)	3.2 (1.4, 6.2)	2.9 (1.3, 6.3)	3.1 (1.5, 6.0)
Family history of MI before age 65 (%)	17.6	22.5	25.1	20.8	19.4
Weighted by Sample Frequencies					
Age (yrs)	62.0 (56.2, 67.8)	68.2 (63.2, 72.6)	67.8 (62.4, 72.3)	68.3 (63.7, 72.6)	69.2 (64.0, 73.6)
Current Smoking (%)	5.7	8.5	8.6	7.4	11.7
Diabetes (%)	3.7	10.1	12.2	8.8	7.0
HbA1c if diabetic (%)	7.0 (6.2, 8.0)	7.5 (6.7, 8.8)	7.4 (6.5, 8.6)	7.6 (6.8, 9.2)	7.3 (7.0, 8.4)
Systolic blood pressure (mmHg)	123.8 (113.9, 136.0)	133.1 (121.5, 146.9)	132.5 (120.6, 146.6)	134.2 (122.6, 148.4)	132.3 (120.4, 142.0)
Hypertension medication (%)	22.5	37.8	37.8	38.0	36.9
Total Cholesterol (mg/dl)	223.8 (199.5, 255.7)	225.2 (198.8, 252.7)	228.6 (202.0, 259.1)	220.2 (195.8, 247.5)	227.0 (202.1, 252.9)
HDL Cholesterol (mg/dl)	54.5 (44.6, 66.5)	48.7 (39.8, 59.8)	48.9 (39.7, 60.1)	48.0 (39.0, 58.5)	50.4 (42.1, 62.2)
C-reactive protein (mg/L)	2.4 (1.0, 5.2)	3.1 (1.4, 6.1)	3.1 (1.4, 6.1)	2.9 (1.3, 6.2)	3.0 (1.4, 5.8)
Family history of MI before age 65 (%)	19.6	23.1	25.6	21.5	20.0

* Values represent percent or median (25%, 75%).

Table 2

Results of multivariable Cox regression analysis, including all variables in the model, in case-cohort sample in ischemic CVD cases and subcohort members with no prior CVD – effects of risk factor categories.

Risk factor	Overall	Whites	Blacks
	HR (95% CI)	HR (95% CI)	HR (95% CI)
Age (yrs)			
60–69	3.05 (2.53–3.69)	3.39 (2.72–4.21)	1.58 (0.92–2.71)
70+	6.66 (5.48–8.09)	7.12 (5.70–8.90)	7.27 (3.99–13.25)
Diabetes	1.94 (1.40–2.67)	1.52 (1.03–2.23)	3.57 (1.75–7.31)
Current smoking	2.39 (1.76–3.26)	2.06 (1.43–2.95)	3.65 (1.63–8.18)
SBP (mmHg)			
120–<140	1.38 (1.15–1.65)	1.47 (1.21–1.79)	0.67 (0.35–1.30)
>=140	2.02 (1.65–2.49)	2.10 (1.67–2.62)	1.09 (0.54–2.20)
Hypertension medication	1.50 (1.26–1.77)	1.49 (1.23–1.79)	1.80 (1.08–3.01)
Total cholesterol (mg/dl)			
200–<240	0.93 (0.76–1.13)	0.93 (0.75–1.15)	1.50 (0.75–3.00)
>=240	1.06 (0.87–1.30)	1.00 (0.81–1.25)	2.18 (1.08–4.41)
HDL cholesterol (mg/dl)			
40–<60	0.66 (0.54–0.82)	0.66 (0.53–0.83)	0.34 (0.17–0.71)
>=60	0.42 (0.34–0.53)	0.42 (0.33–0.54)	0.32 (0.14–0.72)
C-reactive protein (mg/L)			
1–<3	1.22 (0.99–1.50)	1.21 (0.97–1.51)	1.83 (0.79–4.22)
>=3	1.45 (1.19–1.77)	1.43 (1.15–1.78)	1.75 (0.81–3.78)
Family history of premature MI	1.25 (1.03–1.52)	1.25 (1.02–1.53)	1.47 (0.70–3.07)

Table 3

Comparison of recalibrated models for CVD events in the WHI in ischemic CVD cases and subcohort members with no prior CVD, based on weighted survival estimates for case-cohort studies.

Model Comparison	Change in C-index	RC X ² _{old} *	RC X ² _{new} *	NRI (%)	Contm NRI (%)	IDI (%)
All Women						
RRS vs. ATP-III	0.008	273.5	185.7	4.9	32.1	4.1
(p)	0.032			0.010	<0.0001	<0.0001
Framingham CVD vs. ATP-III	-0.008	85.2	246.8	-5.9	-51.5	0.2
(p)	0.020			0.0002	<0.0001	0.42
RRS vs. Framingham CVD	0.016	283.5	153.2	12.9	60.0	3.9
(p)	<0.0001			<0.0001	<0.0001	<0.0001
Whites						
RRS vs. ATP-III	0.007	132.0	198.8	4.3	30.8	4.3
(p)	0.068			0.043	<0.0001	<0.0001
Framingham CVD vs. ATP-III	-0.008	80.2	227.1	-6.9	-52.8	0.3
(p)	0.015			<0.0001	<0.0001	0.34
RRS vs. Framingham CVD	0.016	275.6	186.4	13.0	58.6	4.1
(p)	<0.0001			<0.0001	<0.0001	<0.0001
Blacks						
RRS vs. ATP-III	0.007	15.2	8.0	11.4	32.9	3.5
(p)	0.56			0.13	0.004	0.010
Framingham CVD vs. ATP-III	-0.013	6.8	15.7	-2.1	-30.1	-0.1
(p)	0.16			0.76	0.001	0.80
RRS vs. Framingham CVD	0.020	41.2	14.0	23.7	76.7	3.6
(p)	0.018			0.004	<0.0001	0.0003
All Non-Diabetics						
RRS vs. ATP-III	0.0008	114.2	103.5	4.2	28.1	1.5
(p)	0.82			0.010	<0.0001	<0.0001
Framingham CVD vs. ATP-III	-0.016	70.3	225.8	-9.0	-61.4	-0.6
(p)	<0.0001			<0.0001	<0.0001	<0.0001

Model Comparison	Change in C-index	RC X ² [*] _{old}	RC X ² [*] _{new}	NRI (%)	Cont'n NRI (%)	IDI (%)
RRS vs. Framingham CVD	0.016	246.0	39.0	13.4	59.7	2.2
(p)	<0.0001			<0.0001	<0.0001	<0.0001

* Under population weighting. Only computed for cells with at least 20 observations (small number of cells in blacks).