Biomarkers for Prediction of Cardiovascular Events

TO THE EDITOR: Wang et al. (Dec. 21 issue)[1] suggest that novel biomarkers in aggregate have little influence on the prediction of first cardiovascular events or death. We wonder whether there were sufficient analyses to warrant this conclusion. As described, there were only 68 "major" cardiovascular events in women and 101 in men. In the crucial multimarker analyses, a quarter of these events were eliminated owing to the obligatory inclusion of the urinary albumin-to-creatinine ratio in all the multimarker scores. Would these scores have been more useful if urine biomarkers had been excluded and 100% of the events included?

Second, it is surprising that a study of risk prediction would include prevalent cardiovascular disease, yet analyses that include stroke would not consider atrial fibrillation. Finally, the inclusion of heart failure and coronary insufficiency as "major" cardiovascular events is problematic, since they are not included in the more restrictive definition of "hard" events (myocardial infarction and death from coronary causes) outlined by the third report of the Adult Treatment Panel (ATP III) of the National Cholesterol Education Program. The ATP III definition is currently recommended to determine lipid treatment goals and aspirin use.[2] The study by Wang et al. would be more useful for clinical practice if analyses were also performed according to the ATP III definition.

Kiran Musunuru, M.D., Ph.D.
Roger S. Blumenthal, M.D.
Ciccarone Center
Baltimore, MD 21287
rblument@jhmi.edu

1. Wang TJ, Gona P, Larson MG, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. N Engl J Med 2006;355:2631-9.
2. 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;106:3143-421.

TO THE EDITOR: Wang et al. use the C statistic to determine the usefulness of biomarkers in cardiovascular risk assessment. However, the C statistic has at best a marginal role in selecting variables for prediction models in which the task is to assess the risk of future disease in a currently healthy population. For example, had the authors applied the C statistic individually to low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and blood pressure, they would have been forced to conclude that none of these commonly accepted risk factors have clinical usefulness either. A more principled analysis would have sought evidence on the basis of the proportion of patients correctly reclassified as being at higher or lower risk when novel biomarkers were included.[1-3] This approach is crucial for patients with a 10-year risk of 5 to 10% or 10 to 20% — the "intermediate" groups in which an increase in risk by a factor of 3 (say, from 8 to 24%) would dramatically shift clinical decision making, despite a minimal change in the C statistic. Although never mentioned by Wang et al., it is already known that the risk profile for 25% of such patients in the Framingham cohort would be reclassified on the basis of measurement of high-sensitivity C-reactive protein.[4]

Paul M Ridker, M.D.
Nancy R. Cook, Sc.D.
Brigham and Women's Hospital
Boston, MA 02115
pridker@partners.org

Dr. Ridker is listed as a coinventor on patents held by Brigham and Women's Hospital that relate to the use of inflammatory biomarkers in cardiovascular disease and diabetes.

1. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med 2006;145:21-9.
2. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928-35.
3. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA 2007;297:611-9.
4. Wilson PW, D'Agostino RB, Sullivan L, O'Donnell CJ. Increased CRP and long term risk for cardiovascular events in middle age men and women. Circulation 2006;114:Suppl II:II-877–II-878. abstract.

TO THE EDITOR: The reason given by Wang et al. for not subjecting LDL cholesterol, HDL cholesterol, blood pressure, and smoking status to the same scrutiny by the C statistic that they used for evaluating novel biomarkers is that these "conventional" risk factors are already widely accepted and part of the Framingham risk score and ATP III algorithms. However, without explanation, Wang et al. further include the body-mass index and serum creatinine level in their base model, precluding any comparisons with previous prediction algorithms. We are not aware of any previous Framingham reports showing that either the body-mass index or the creatinine level has an effect on the C statistic. Until an analysis is presented that places all putative risk factors on an equal footing and uses more appropriate methods to evaluate these risk factors, few conclusions can be drawn.

Diane M. Becker, M.P.H., Sc.D.
Johns Hopkins University School of Medicine
Baltimore, MD 21287
Samia Mora, M.D., M.H.S.
Brigham and Women's Hospital
Boston, MA 02115
David C. Goff, Jr., M.D., Ph.D.
Wake Forest University School of Medicine
Winston-Salem, NC 27157

TO THE EDITOR: In the Perspective article accompanying the report by Wang et al., Ware[1] estimates that a relative risk would have to be more than 200 to have sufficient sensitivity and specificity to be a useful predictor of cardiovascular risk in individual patients. This is so even though much smaller relative risks do define populations at increased risk. Other observers have made this case on theoretical[2] and empirical[3] grounds, yet the concept has not entered into clinical thinking. The implications for evidence-based medicine, and for what clinicians communicate to patients, are large. Risk stratification is advocated as a way of increasing the effectiveness and efficiency of health care.[4] Clinical practice guidelines commonly include stratification based on relative risks that are much smaller than 200. For example, family history is associated with an increase in the risk of breast and colorectal cancers by a factor of 2 to 6 and is the basis for recommending aggressive screening in some patients. Few clinicians, including those who prepare guidelines, understand how little such advice is likely to help a given patient. It is time for a paradigm shift in the application of risk stratification to the care of individual patients.

Robert H. Fletcher, M.D.
Suzanne W. Fletcher, M.D.
Harvard Medical School
Boston, MA 02115
robert_fletcher@hms.harvard.edu

1. Ware JH. The limitations of risk factors as prognostic tools. N Engl J Med 2006;355:2615-7.
2. Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? BMJ 1999;319:1562-5.
3. Elmore JG, Fletcher SW. The risk of cancer risk prediction: "What is my risk of getting breast cancer?" J Natl Cancer Inst 2006;98:1673-5.
4. Joy JE, Penhoet EE, Petitti DB, eds. Saving women's lives: strategies for improving breast cancer detection and diagnosis. Washington, DC: National Academies Press, 2005.

TO THE EDITOR: With regard to the discussion by Ware, another reason there is confusion about the clinical usefulness of prognostic studies is that a Bayesian approach — widely used in diagnosis — is not routinely used in prognosis. Likelihood ratios are summary statistics of diagnostic tests. They indicate the degree to which the chance of having a condition changes relative to baseline as a result of the test, thus relating the risk after a certain test result to the baseline risk. In prognostic studies, the hazard ratio is frequently used as a summary statistic. The hazard ratio usually relates the risk in a group with one test result to the risk in a group with a different test result, not to the baseline risk. If the entire cohort of patients were used as a reference, the hazard ratio would have a meaning similar to that of the likelihood ratio. When overall event rates are small, the hazard ratio is equal to the likelihood ratio. This would allow for a more clinically meaningful interpre-

tation of a study's results and the estimation of risk for individual patients.

Gregory Mints, M.D.
Nirav R. Shah, M.D., M.P.H.
New York University School of Medicine
New York, NY 10016
mintsg01@popmail.med.nyu.edu

TO THE EDITOR: The limitations of using risk factors to predict outcomes, as discussed by Ware, is a particular problem for acute care when the mere fact that the patient presents with a specific symptom results in such a high pretest probability that the presence or absence of conventional risk factors is essentially irrelevant. Residents find it difficult to understand that inquiring about smoking status is not helpful when doing a workup of a middle-aged patient with sudden chest pain. Unfortunately, such confusion is not limited to medical trainees. Medicare rules require an extensive inquiry into the patient's history (including family and social factors) to justify higher levels of billing for evaluation and management. One worries about the atypical myocardial infarctions that must have been missed because the patient was considered to be at low risk after a negative history.

Mark Hauswald, M.D.
University of New Mexico
Albuquerque, NM 87131
mhauswald@salud.unm.edu

DR. WANG AND COLLEAGUES REPLY: In response to Musunuru and Blumenthal: we observed no difference in the usefulness of multimarker scores in analyses that included the entire sample (i.e., that included persons without urine specimens). As stated in the article, patients with prevalent cardiovascular disease were excluded from analyses for incident cardiovascular disease. Further adjustment for atrial fibrillation did not alter our findings. We had too few fatal or nonfatal myocardial infarctions to examine this end point separately with adequate statistical power. The inclusion of heart failure and coronary insufficiency (unstable angina with documented changes on electrocardiography) as part of the composite end point is consistent with the definition of "hard" cardiovascular disease used in previous Framingham reports and based on standardized criteria in use since the inception of the cohort.[1] Given the morbidity associated with these conditions, we believe that it is valid to consider them "major" events.

Ridker and Cook object to the use of the C statistic to evaluate novel biomarkers. Although several measures of model performance exist, the C statistic remains one of the standard tools for assessing screening tests because it is easy to interpret, objective, and applicable to different populations.[2] Nonetheless, we also evaluated the ability of biomarkers to "reclassify" the risk of cardiovascular disease, the approach advocated by Ridker and Cook. The majority of reclassified patients were moved into a lower risk category, rather than a higher one. Specifically, the multimarker score reclassified only 1% of patients who were at low or intermediate risk into the high-risk group (predicted risk, ≥20% on the basis of our 8-year follow-up), the only shift likely to affect pharmacologic decision making.[3] Even if we were to restrict biomarker assessment to the intermediate-risk groups, only 2.7% would be reclassified as being at high risk. Notably, the use of C-reactive protein measures in the Women's Health Study reclassified an even smaller proportion of women (0.2%) who were at low or intermediate risk as being at high risk.[4] The measurement of C-reactive protein did not enter our multimarker score for cardiovascular disease because it was less informative than several other biomarkers we studied.

We disagree with Becker et al. that conventional risk factors and novel biomarkers should be given equal opportunity to enter risk models. Novel biomarkers should be judged on their ability to add to existing risk factors that constitute the current standard of care. Lipid levels, blood pressure, and smoking status are easily ascertained, modifiable with lifestyle changes or drug therapy, and linked to cardiovascular disease in a causal manner. Also, large, randomized trials have firmly established the benefits of modifying these risk factors.

The body-mass index and serum creatinine level were included in our models because they are routinely assessed and may confound biomarker levels.[5] Excluding these factors from the models resulted in unchanged biomarker results.

Thomas J. Wang, M.D.
Massachusetts General Hospital
Boston, MA 02114
tjwang@partners.org

Martin G. Larson, Sc.D.
Ramachandran S. Vasan, M.D.
Boston University
Boston, MA 02118

1. Kannel WB, Wolf PA, Garrison RJ, eds. The Framingham Study: an epidemiological investigation of cardiovascular disease. Section 34. Some risk factors related to the annual incidence of cardiovascular disease and death in pooled repeated biennial measurements: Framingham Heart Study, 30 year follow-up. Bethesda, MD: National Heart, Lung, and Blood Institute, 1987. (NIH publication no. 87-2703.)
2. D'Agostino RB, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180-7.
3. Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation 2004; 110:227-39. [Erratum, Circulation 2004;110:763.]
4. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med 2006;145:21-9.
5. Wang TJ, Larson MG, Levy D, et al. Impact of obesity on plasma natriuretic peptide levels. Circulation 2004;109:594-600.

DR. WARE REPLIES: The C statistic is one of several global measures of the performance of risk-prediction rules. It measures the ability of a prediction rule to classify an entire study population into groups that will and those that will not have the predicted event. Though methods for evaluation of risk-prediction models are still evolving, maximization of the area under the receiver-operating-characteristic curve is competitive with likelihood maximization as a method for deriving risk-prediction models on the basis of linear combinations of biomarkers.[1]

Ridker and Cook note that changes in the C statistic are small even when recognized predictors are added to prediction models. They have shown in studies of the Women's Health Study population[2] that after the inclusion of age, systolic blood pressure, and smoking status in a model for prediction of cardiovascular risk, other individual predictors increase the C statistic by 0.01 or less. However, the data show a close correspondence between the changes in the C statistic and the likelihood ratio with the addition of single covariates to the prediction model, though with substantial differences in scale.

Global measures have limitations as measures of risk prediction. One limitation is conceptual. The outcome is inherently stochastic; we cannot hope to achieve perfect prediction. The other is that global measures of risk prediction may be insensitive to improvements that predominantly affect subgroups of the study population.

Ridker and Cook compare nested prediction models by assessing the extent to which a more complex model reassigns patients to more appropriate risk strata. In a recent study,[3] they compared risk prediction in 7911 women without diabetes in the Women's Health Study using a model based on traditional risk factors and an enhanced model that included parental history of myocardial infarction and levels of high-sensitivity C-reactive protein. Among 767 women with a predicted 10-year risk of more than 5% according to the traditional model, 339 women (44%) were reclassified by the refined model. The observed risks in the women who were reclassified were consistent with the revised predicted risks. In the entire sample, 647 women (8%) were reclassified.

In this population, improvement in risk prediction occurred predominantly in women at elevated risk according to the traditional model. The fact that the high-risk subgroup represents about 10% of the Women's Health Study population is one explanation for the small change in the C statistic resulting from this refinement. Of 10 novel biomarkers considered in their study, only high-sensitivity C-reactive protein was included in the final, "clinically simplified" model.

The comments of Fletcher and Fletcher, Mints and Shah, and Hauswald are appreciated and require no elaboration.

James H. Ware, Ph.D.
Harvard School of Public Health
Boston, MA 02115
ware@hsph.harvard.edu

1. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med 2006;145:21-9.
2. Pepe MS, Cai T, Longton G. Combining predictors for classification using the area under the receiver operating characteristic curve. Biometrics 2006;62:221-9.
3. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA 2007;297:611-9.