Commentary

Evaluating New Cardiovascular Risk Factors for Risk Stratification

Ralph H. Stern, PhD, MD

There is an active debate on the best approach to evaluate new cardiovascular risk factors. The divergent views have been most clearly laid out in the case of C-reactive protein (CRP) and its incremental value over conventional cardiovascular risk factors. Advocates of CRP argue that its addition improves global measures of fit,[1] while others point out that it does not improve measures of discrimination.[2,3] Discussions surrounding this debate have largely been technical, focusing on unique data sources and specific analytic approaches. The underlying principles, in contrast, have not been sufficiently highlighted. This paper presents a clinical perspective on the potential value of adding new cardiovascular risk factors to current methods for risk stratification. It proposes that novel risk factors should be evaluated by their effects on the population risk distribution curve, which presents the probability of occurrence of different levels of risk in a population as determined by a predictive model. Broader population risk distribution curves represent superior risk stratification. Last, it should be recognized that there are many correct ways to risk-stratify a population and that the best approach ultimately depends on the clinical goals.

THE USE OF RISK MODELS IS TO STRATIFY RISK

The goal of risk stratification is to identify subpopulations of differing cardiovascular risk within a larger general population to allow interventions to be employed selectively. Mathematic risk models must be applied to populations to determine their clinical utility. New cardiovascular risk factors that do not improve risk stratification within these models are not of clinical value, although they may increase scientific understanding or reveal new opportunities to develop preventive measures.

Applying a predictive model to an entire population generates a population risk distribution curve. These are rarely presented in the cardiovascular literature but have important implications for assessing the value of risk factors. We have been unable to locate a graph depicting the population distribution of cardiovascular risk determined by a Framingham risk score, despite the widespread use of this risk score in the literature. An example of this is depicted in Figure 1. This population risk distribution curve shows how frequently patients with different levels of cardiovascular risk occur in the population. As expected, the higher their cardiovascular risk, the less often these patients are encountered. It is uncommon to find healthy patients with a 10-year risk >20%, the range expected for patients with coronary artery disease. Graphs of this type represent a valuable way to evaluate the clinical importance of risk factors. Neither global measures of fit nor the receiver operating characteristic (ROC) curve so directly communicate the results of the risk stratification process.

BROADER RISK DISTRIBUTION CURVES REPRESENT SUPERIOR RISK STRATIFICATION

Absent any information about risk factors, all members of the population would be assigned the same risk. The risk distribution "curve" would be a spike at the mean population risk. As risk factors are added to a mathematic risk model, the

From the Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI
Address for correspondence: Ralph H. Stern, PhD, MD, CVC Cardiovascular Medicine, 1500 E Medical Center Drive SPC5853, Ann Arbor, MI 48109-5853
E-mail: stern@umich.edu

www.lejacq.com ID: 7814

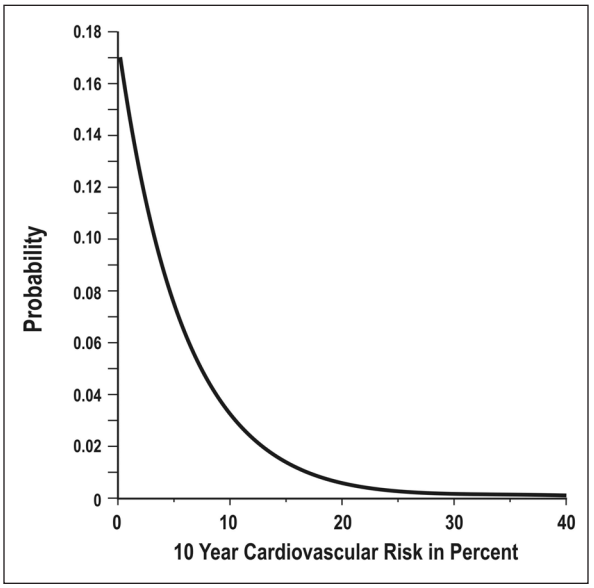

Figure 1. Approximation to the frequency distribution of 10-year cardiovascular risk in US adults without cardiovascular disease or cardiovascular disease risk equivalents. This distribution (an exponential distribution with a lambda of 17) assigns patients to low, intermediate, and high risk in the same proportions reported in the literature.[12]

risk distribution curve should progressively widen, so long as the risk factors contribute additional information. Figure 2 provides a hypothetical example of risk distribution curves of increasing width, demonstrating improved risk stratification. The increasing width shows an increasing ability to separate the population into subgroups of higher and lower risk.

If two mathematic risk models produce identical risk distribution curves when applied to the same population, then the risk stratification of the two models is equivalent. Real-world data are unlikely to generate mathematically identical risk distribution curves; however, similar risk distribution curves should be considered clinically equivalent, even if there are mathematic differences.

RISK DISTRIBUTION CURVES AND ROC CURVES CONTAIN IDENTICAL INFORMATION

The ROC curve, which depicts the relationship between sensitivity and specificity, has been an important tool for assessing diagnostic methods and predictive models.[4] Diamond[5] showed that the ROC curve for a risk model can be calculated from the population risk distribution. Since the ROC curve can be derived from the risk distribution curve (and vice versa), these curves contain identical information. This means that the mathematic analyses of the ROC curve in the risk factor research literature can be better understood simply by evaluating the corresponding risk distribution curve.

Because the ROC curve and the risk distribution curve contain identical information, properties of one curve must correspond to properties of the other. The area under the ROC curve (the ROC curve AUC), used to compare both diagnostic methods and predictive models, is an example. Although the ROC curve AUC is commonly understood to measure the overlap of the risk distribution curves of patients with and without events, also known as discrimination, it is equally valid to consider the ROC curve AUC as a measure of the width of the population risk distribution curve. For example, the 3 population risk distribution curves with increasing width in Figure 2 have ROC curve AUCs of 0.55, 0.65, and 0.75. Cook[6] provides a similar example using the β distribution.

THERE ARE MANY "CORRECT" WAYS TO RISK-STRATIFY A POPULATION

For diagnostic methods, there is only one way to be correct: perfectly discriminate between patients with and without a disease. This is the definition of a gold standard or perfect test. For a predictive model to match this performance, it would need to identify patients who will or will not have an event in the future. Unfortunately, this is not possible because the occurrence of events is random or stochastic.

An alternative and more realistic criterion is that a predictive model is correct if it accurately assigns risk to different subpopulations. This property is referred to as calibration. It is commonly evaluated by comparing the observed to predicted risk for each decile of risk.[7] For a given population, there will be a multitude of risk stratification methods, all of which are correct or calibrated. However, because they utilize different risk factors, these methods will assign different risks to the same individual. Consider risk-stratifying the same population twice, once by systolic pressure and once by low-density lipoprotein cholesterol. Assume the corresponding cardiovascular risk for each decile was the same for the 2 risk factors. Then the risk stratification of the population by the 2 risk factors is identical. But patients with low blood pressure may have high low-density lipoprotein cholesterol, and vice versa, so individuals may be assigned very different risks depending on which risk factor is used to risk stratify the population.

DISCUSSION

Because the goal of risk stratification is to identify subpopulations of differing cardiovascular risk, the population risk distribution curve is the most useful way to assess a predictive method's risk stratification in a given population. Neither global measures of fit nor ROC curves show this directly.

Framingham risk models are currently able to identify adult patients without cardiovascular disease with 10-year risks ranging from <1% to >30%. Figure 1 (an approximation) illustrates the corresponding population risk distribution.

It is unknown whether this risk stratification can be improved. Newer risk factors[8] or even measures of carotid atherosclerosis combined with CRP[9] have not improved risk stratification using conventional risk factors. Diamond[5] showed that risk models have limited discriminatory ability. As discussed above, perfect discrimination would mean that a predictive method could identify patients who will or will not have an event in the future. But making accurate predictions over a 10-year period is only possible in deterministic systems (eg, prediction of eclipses in the solar system). The occurrence of clinical events in low-risk individuals is often thought to indicate that improved predictive methods are possible, but as long as the predicted number of events match the observed number of events in a low-risk subpopulation, this is not the case.

In comparing risk models, one should recognize that the predicted risk for individual patients can vary between calibrated models. An example of this for univariate risk models using systolic blood pressure and low-density lipoprotein cholesterol was given previously. If additional risk factors were added to such univariate models, patients with high levels of the additional risk factor would be assigned higher risks, while those with low levels of the additional risk factor would be assigned lower risks. This would occur throughout the risk distribution curve, ensuring extensive scrambling or reassortment of individual patients with the addition of each new risk factor. This would also occur even if the population risk distribution curve is not changed by the additional risk factors. Ridker and colleagues[10] described 2 calibrated multivariate models that differ in the number of risk factors. In spite of having the same discrimination, these models classified many patients into different risk strata. In this situation, the risk distribution curves should be the same, and thus the number of patients moved out of a risk stratum will be balanced by the number of patients moved into that risk stratum when additional risk factors are added.

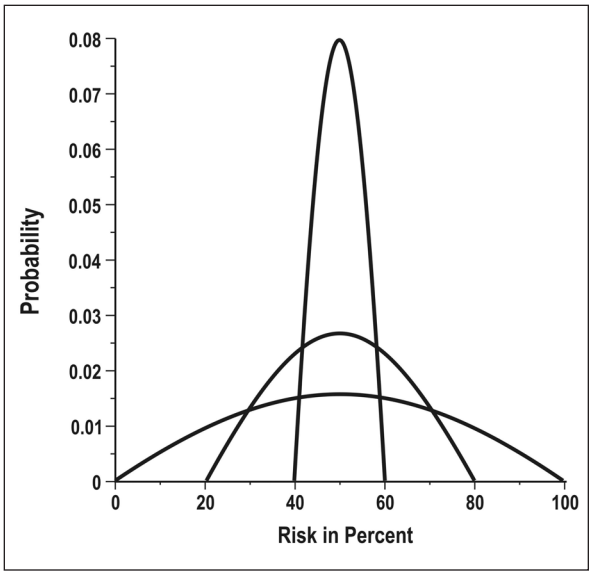

Figure 2. Sinusoidal population risk distribution curves with the same mean risk but increasing width, indicating improved risk stratification. The broader the risk distribution curve, the better the separation of patients of differing risks.

Reynolds and associates[11] compared 3 cardiovascular risk models that predict similar numbers of patients above the same level of risk with a simulated population and showed limited concordance in the individuals identified. Lack of concordance when multiple cardiovascular risk models are used to risk stratify the same population has not been extensively studied. However, it should be expected that there will be a large number of cardiovascular risk models based on different risk factors that are calibrated and produce similar population risk distribution curves, but these models may classify individuals differently.

This should be contrasted with the situation with diagnostic testing. In that case, there is only one correct answer for an individual, "intermediate" probability indicates a poor assessment, and further testing is indicated. But once a population has been risk-stratified by a calibrated model, further testing may be counterproductive, even for "intermediate-risk" patients. Further testing may simply replace one reasonable population risk stratification with another equivalent one, while the reassignment of individuals leads to confusion and wasted effort. It is tempting to use multiple risk estimation methods in the same patient to avoid missing something. This will lead to erroneous risk stratification of the population and overtreatment, as the number of individuals identified as high-risk by any one of a number of risk models

will exceed the number of high-risk individuals in the population.

Whether a risk factor contributes to risk stratification is a separate issue from whether it is causative or should be treated even if a therapy is available. Cook[6] has shown that systolic blood pressure, total cholesterol, high-density lipoprotein cholesterol, or smoking status can be removed from models including all the Framingham risk factors with minimal change in discrimination. It appears that only a few cardiovascular risk factors are required to adequately risk-stratify the population. Treatment decisions need to be based on clinical trials, not on whether addition of a risk factor improves risk stratification.

How well a model risk-stratifies a population should be the primary criterion for choosing a method for clinical use, but there are additional criteria that could be considered. These would include cost, precision, reproducibility, availability, safety of the test, and whether modification of the risk factor reduced risk.

There is a bewildering array of new risk factors under investigation. Examples include CRP, coronary artery calcium score, and measures of arterial stiffness. From a clinical perspective, if risk factors do not improve risk stratification as assessed by the population risk distribution curve, they are not improving patient care. Risk factor evaluation by regression methods only measures association and provides no information on the population risk distribution curve. In contrast, the standard measure of discrimination, the ROC curve AUC, does. Few cardiovascular risk factors have been appropriately evaluated for clinical utility. CRP has been appropriately evaluated and has failed to demonstrate clinical utility.[1–3] Therefore, at present it does not appear that CRP needs to be added to the Framingham risk factors for cardiovascular risk calculation. Although diagnostic methods arrive at unique assessments for an individual patient, predictive methods do not, even if they are calibrated and have similar population risk distribution curves. In this case, restratification of a population may be unhelpful.

References

1 Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein cardiovascular risk prediction models for women. *Ann Intern Med.* 2006;145:21–29.

2 Lloyd-Jones DM, Liu K, Tian L, et al. Narrative review: assessment of C-reactive protein in risk prediction for cardiovascular disease. *Ann Intern Med.* 2006;145:35–42.

3 Wilson PWF, Nam B-H, Pencina M, et al. C-reactive protein and risk of cardiovascular disease in men and women from the Framingham Study. *Arch Intern Med.* 2005;165:2473–2478.

4 Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation.* 2007;115:654–657.

5 Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol.* 1992;45:85–89.

6 Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;115:928–935.

7 Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons; 1989.

8 Wang TJ, Gona P, Larson MG. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med.* 2006;355:2631–2639.

9 Cao JJ, Arnold AM, Manolio TA. Association of carotid artery intima-media thickness, plaques, and C-reactive protein with future cardiovascular disease and all-cause mortality: the Cardiovascular Health Study. *Circulation.* 2007;116:32–38.

10 Ridker PM, Buring JE, Rifai N, et al. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA.* 2007;297:611–619.

11 Reynolds TM, Twomey PJ, Wierzbicki AS. Concordance evaluation of coronary risk scores: implications for cardiovascular risk screening. *Curr Med Res Opin.* 2004;20:811–818.

12 Ford ES, Giles WH, Mokdad AH. The distribution of 10-year risk for coronary heart disease among US adults: findings from the National Health and Nutrition Examination Survey III. *J Am Coll Cardiol.* 2004;43:1791–1796.