Risk Prediction in Cardiovascular Medicine

Cardiovascular Risk Prediction
Basic Concepts, Current Status, and Future Directions

Donald M. Lloyd-Jones, MD, ScM

Few topics have received as much attention in the cardiovascular literature over the last 5 years as risk prediction. The assessment of risk has been a key element in efforts to define risk factors for cardiovascular disease (CVD), to identify novel markers of risk for CVD, to identify and assess potential targets of therapy, and to enhance the cost-effective implementation of therapies for both primary and secondary prevention of CVD. With the publication of the Third Report of the National Cholesterol Education Program’s Adult Treatment Panel (ATP-III) in 2001 and similar guidelines from other national and international bodies, risk prediction assumed a central role in the field of CVD prevention. Since then, numerous attempts have been made to refine and improve risk assessment methods. A basic review of the history and principles of risk prediction, their statistical underpinnings, and their clinical implications is thus required to help clinicians and researchers understand the increasingly complex approaches being undertaken to refine CVD risk prediction. Other articles in this series address the current adoption, utility, and effect of risk estimation algorithms in clinical practice.

Rationale for CVD Risk Prediction
Risk estimates can theoretically be used to raise population awareness of diseases (such as CVD) that cause a significant burden of morbidity and mortality, to communicate knowledge about that risk to individuals and subgroups, and to motivate adherence to recommended lifestyle changes or therapies. In clinical practice, risk prediction algorithms have been used most directly to identify individuals at high risk for developing CVD in the short term to select those individuals for more intensive preventive interventions. The prime example of this latter “high-risk” prevention strategy, which relies heavily on the quantitative prediction of CVD risk, is the approach promulgated by the ATP-III panel. In this algorithm, the stated prevention of CVD was achieved through the use of multivariable regression equations derived from the Framingham Heart Study. These models predicted the risk of developing CVD for individuals based on their age, sex, blood pressure, cholesterol levels, and other risk factors. The results of these models were then used to identify individuals at high risk, who were then targeted for preventive interventions.

Historical Perspective
Even with the most rudimentary understanding of CVD, it is clear that certain individuals appear to have a substantial likelihood of developing or dying of CVD, whereas others appear destined to live long lives free of CVD. However, translating this simple observation into a quantitative probability of disease requires a number of factors. First, reliable data must be available to indicate disease incidence rates. Second, a number of risk markers must be available to assess associations with future disease incidence. Statistical methods must be available to quantify the prospective association between risk markers and the occurrence of disease. Further refinement of risk prediction requires adequate statistical methods for quantifying the potential improvement that 1 risk prediction model may offer over another. Finally, and importantly, the utility of risk prediction algorithms must be assessed in the context of the clinical environment, including considerations of the burden and severity of the disease being predicted, the availability of safe and effective interventions to prevent disease, and cost-benefit considerations of applying those therapies to different segments of the population. Only in the last 5 to 10 years have all of these tools been available.

Early guidelines from the National High Blood Pressure Education Program and the National Cholesterol Education Program focused on those at the highest relative risks for developing CVD related to markedly elevated blood pressure and cholesterol levels. However, physicians and patients often have difficulty interpreting relative risk estimates. For example, a relative risk for disease of 10 might seem very high, but if the incidence rate in the referent group is close to 0, it will also be close to 0 in the group with the relative risk of 10. Conversely, a relative risk of 1.3 might be very important when a disease is common. Therefore, researchers and policy makers have focused on absolute risk estimation to provide absolute probabilities of developing CVD within a given time frame. These estimates may be more easily understood than relative risks, and they allow clinical recommendations for interventions in individuals who exceed unacceptable risk thresholds.

Current CVD Risk Prediction in Practice
Five- and 10-year risk estimates have been widely adopted by ATP-III and other guidelines and are most often based on multivariable regression equations derived from the Framingham Heart Study. These models are used to estimate the risk of developing CVD within a specific time frame, taking into account various risk factors such as age, sex, blood pressure, cholesterol levels, smoking status, and family history of CVD. The models provide an estimate of the likelihood of developing CVD within the next 5 or 10 years, and this information is used to guide preventive interventions. The models have been shown to be effective in identifying individuals at high risk for CVD and in targeting preventive interventions to those at highest risk.

From the Department of Preventive Medicine and Bluhm Cardiovascular Institute, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Ill.

Correspondence to Donald M. Lloyd-Jones, MD, ScM, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, 680 N Lake Shore Dr, Suite 1400, Chicago, IL 60611. E-mail dlj@northwestern.edu

(Circulation. 2010;121:1768-1777.)

© 2010 American Heart Association, Inc.

Circulation is available at http://circ.ahajournals.org

DOI: 10.1161/CIRCULATIONAHA.109.849166
Table 1. Examples of Currently Available CVD Risk Prediction Scores

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>End Point</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framingham, 1998</td>
<td>All CHD</td>
<td>Includes CHD death, MI, unstable angina, and angina pectoris</td>
</tr>
<tr>
<td>ATP-III risk estimator,</td>
<td>Hard CHD</td>
<td>Includes CHD death and nonfatal MI</td>
</tr>
<tr>
<td>2001 (Framingham)</td>
<td>Global CVD</td>
<td>Includes CHD death, all CHD, stroke, heart failure, and claudication</td>
</tr>
<tr>
<td>Framingham global CVD,</td>
<td>Hard CHD</td>
<td>Includes CHD death and nonfatal MI</td>
</tr>
<tr>
<td>2008</td>
<td>Global CVD</td>
<td>Includes CHD, stroke, and transient ischemic attack</td>
</tr>
<tr>
<td>PROCAM†</td>
<td>Global CVD</td>
<td>Includes CVD death, MI, stroke, and revascularization</td>
</tr>
<tr>
<td>QRISK20</td>
<td>CVD</td>
<td>Includes CVD death, MI, stroke, and revascularization</td>
</tr>
<tr>
<td>Reynolds risk score</td>
<td>Global CVD</td>
<td>Includes CVD death only; does not include nonfatal events; multiple region-specific (northern European, southern European) and country-specific versions available</td>
</tr>
<tr>
<td>(women)21</td>
<td>Global CVD</td>
<td></td>
</tr>
<tr>
<td>Reynolds risk score</td>
<td>Global CVD</td>
<td></td>
</tr>
<tr>
<td>(men)22</td>
<td>CVD death</td>
<td></td>
</tr>
</tbody>
</table>

Table data from Grundy et al.24

*CHD risk equivalents include clinical manifestations of noncoronary forms of atherosclerotic disease (peripheral arterial disease, abdominal aortic aneurysm, and carotid artery disease), ie, transient ischemic attacks, stroke of carotid origin, or >50% obstruction of a carotid artery or diabetes mellitus.
†Risk factors include cigarette smoking, hypertension (blood pressure ≥140/90 mm Hg or on antihypertensive medication), low-density lipoprotein cholesterol (<40 mg/dL), family history of premature CHD (CHD in male first-degree relative <55 years of age; CHD in female first-degree relative <65 years of age), and age (men ≥45 years; women ≥55 years).

Interpretation of Absolute Risk Estimates

Ten-year absolute risk estimates may serve not only as a basis for decision making about institution of lipid-lowering drug therapy but also as a useful means for risk communication with patients. However, interpretation of a 10-year risk estimate must be done appropriately. For a given patient with a 10-year risk estimate of 7%, it would be inappropriate to state that that individual’s risk is 7%. Rather, a correct interpretation would be to say that, given 100 similar individuals, we expect that 7 will experience an event in the next 10 years and 93 will not.

Table 2. Risk Classification Algorithm Used in the ATP-III 2004 Update

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>High risk</td>
<td>CHD or CHD risk equivalent* or ≥2 risk factors† and 10-y predicted risk of &gt;20%</td>
</tr>
<tr>
<td>Moderately high risk</td>
<td>≥2 Risk factors and 10-y predicted risk of 10% to 20%</td>
</tr>
<tr>
<td>Moderate risk</td>
<td>≥2 Risk factors and 10-y predicted risk of &lt;10%</td>
</tr>
<tr>
<td>Lower risk</td>
<td>0–1 Risk factor</td>
</tr>
</tbody>
</table>

Table data from Grundy et al.24

*CHD risk equivalents include clinical manifestations of noncoronary forms of atherosclerotic disease (peripheral arterial disease, abdominal aortic aneurysm, and carotid artery disease), ie, transient ischemic attacks, stroke of carotid origin, or >50% obstruction of a carotid artery or diabetes mellitus.
†Risk factors include cigarette smoking, hypertension (blood pressure ≥140/90 mm Hg or on antihypertensive medication), low-density lipoprotein cholesterol (<40 mg/dL), family history of premature CHD (CHD in male first-degree relative <55 years of age; CHD in female first-degree relative <65 years of age), and age (men ≥45 years; women ≥55 years).

In the current ATP-III treatment algorithm,24 all patients with existing CHD, diabetes mellitus, peripheral arterial disease, aortic aneurysm, or symptomatic or significant carotid artery disease are categorized as high risk and treated as if they require secondary prevention of CHD. For those without CHD or CHD risk equivalents, absolute risk estimates are categorized into discrete risk categories (Table 2) to guide primary prevention decision making. Individuals with 10-year risks ≥20% are recommended for immediate lipid-lowering drug therapy to reduce their risk. In patients with intermediate risk (10% to 20%), the recommendation is either to start drug therapy or to pursue other noninvasive testing for further risk stratification. Lower-risk subjects are typically not recommended for drug therapy, but for lifestyle modification as appropriate.

The interval of 10 years for risk estimation was chosen as the focus for a variety of reasons. Most notably, consideration of 10-year risk identifies those patients most likely to benefit from drug therapy in the near term, thus improving cost-effectiveness and safety of therapy. In addition, robust data for estimation of current CVD risks associated with risk factors in a contemporary environment require a focus on shorter-term follow-up. The risk for CHD associated with traditional risk factors is continuous across the population, and there are no obvious natural thresholds. Nonetheless, the thresholds used by ATP-III for clinical decision making were determined from population data and cost-effectiveness estimates in an era when statin medications cost substantially more than they do currently.1 In addition, optimal prevention necessitates broader application of lipid-lowering drugs because the majority of events occur in those with average or only modestly adverse levels of blood lipids (because this is where the vast majority of the population at risk is found). Therefore, it seems likely that future guidelines may choose lower thresholds for therapy in light of demonstrated benefit of statins in populations at predicted risk <20%, the availability of inexpensive statins, and longer-term safety data.
The clinician’s job, therefore, is to understand the probabilistic nature of the estimate and to weigh additional factors, such as obesity or the presence of a family history of premature CVD, or perhaps to perform further testing that might modify the assessment and assist in discerning whether the individual patient is more likely to be among the diseased or the disease-free group at the end of 10 years. These additional factors have not been incorporated into many risk estimation equations because they do not achieve statistical significance or contribute meaningfully beyond the traditional risk factors in predicting 10-year risk across the entire population. Nonetheless, they can assist in modifying the base risk estimate, particularly in intermediate-risk individuals, as noted below. It is also important to recognize that a “low-risk” risk estimate (eg, <5%) does not mean “no risk,” particularly in the context of longer time horizons than 10 years. If the patient is expected to live >10 years, his or her 10-year risk estimate will perforce increase over time, and any adverse levels of risk factors present at younger ages may cause marked elevations in lifetime risks for CVD.

A common criticism of risk prediction models is that they provide risk estimates for populations, not individuals. However, this naïve criticism ignores the fact that much of currently recommended medical practice and essentially all evidence-based practice require the application of population mean effects to individuals. This will continue to be the case for the foreseeable future (at least for CVD risk estimation) until the potential promise of “personalized medicine” is reached, perhaps through refined genetic profiling.

**Metrics for Assessing the Utility of Absolute Risk Prediction Equations**

Demonstration that a risk score has a significant statistical association with the outcome of interest is necessary but grossly insufficient for evaluating its utility. A number of metrics are available to assist in the evaluation of the performance and utility of risk estimation models. In general, these metrics assess parameters of the risk prediction model (similar to a diagnostic test), its ability to discriminate future cases from noncases, the calibration of the model, the model fit, and the informativeness of the model for the outcome of interest. Consideration of all of these factors is important to understanding the utility of a risk score. Newer methods of assessment involving analysis of risk reclassification also allow comparison of different risk stratification algorithms using novel markers or risk scores. Knowledge of a few of these metrics and concepts will suffice for most clinicians to interpret the utility of risk prediction models.

**The C Statistic**

The most widely reported measure of model discrimination for CVD risk prediction models is the C statistic. The C statistic is a function of both the sensitivity and specificity of the model across all of its values, and it represents the ability of the score to discriminate (future) cases from noncases. Simply put, the C statistic indicates the probability that a randomly selected patient who develops the disease (a “case”) will have a higher risk score than a randomly selected noncase. Thus, a C statistic of 0.75 for a given model would indicate that a randomly selected case has a higher score than a randomly selected noncase 75% of the time. The C statistic can vary from 1.0 (perfect discrimination) to 0.5 (random chance, indicating that the score being applied is no better than flipping a coin). C statistics <0.70 are thought to indicate inadequate discrimination by current convention, whereas those between 0.70 and 0.80 are considered acceptable and those between 0.80 and 0.90, excellent.

The C statistic has received some criticism as a metric for assessment of CVD risk prediction models. It is fair to say that reliance on the C statistic alone as a measure of test performance has significant limitations. For example, the C statistic simply indicates whether a risk score is providing appropriate rank ordering of risk for cases and noncases, not whether the estimated and observed risks are similar (which is a function of calibration) or how much greater the estimated risk is between selected cases and noncases.

A notable property of the C statistic is that extremely large odds ratios (or relative risks) are required to achieve clinically meaningful levels or increases in the C statistic. For example, as demonstrated by Pepe et al, for a binary risk marker considered in isolation, a univariate odds ratio of ≥9.0 would be required for a C statistic that provides excellent discrimination of cases from noncases and similarly for a continuous risk marker in which the distribution of risk scores differs by ≥2 SDs. These magnitudes of relative risk are very rarely observed in clinical practice. However, the combination of multiple, independent risk markers, as in the FRS and similar scores, does provide these magnitudes of relative risk; hence, these scores typically have C statistics in the range of 0.75 to 0.80.

**Measures of Calibration**

As opposed to measures of discrimination, which indicate the ability of a risk model to rank order individuals’ risks, measures of calibration assess the ability of a risk prediction model to predict accurately the absolute level of risk that is subsequently observed. For example, demonstrating that a risk prediction model is well calibrated would require an observed event rate of close to 7% if the model estimates that the risk for a certain subgroup of individuals is 7% over 10 years. Calibration is often assessed visually by dividing the population at risk into quantiles (eg, deciles) of predicted risk and plotting the predicted risk versus the observed event rate for each quantile (Figure 1). The statistical metric often used to test for the calibration of a risk model is the Hosmer-Lemeshow χ² test. A value of P<0.05 for such a test would indicate poor calibration of the model for the population. Note that a test may have poor calibration while still having reasonable discrimination ability, as in Figure 1B, in which the rank ordering of risk by the model is hierarchical but calibration is visibly poor.

**Measures of Model Fit and Informativeness**

Other measures such as likelihood ratio tests and the Bayes information criterion are now commonly used to assess the utility of risk prediction models. Simply put, these tests, when statistically significant, can indicate whether a risk model is predicting disease incidence better than chance alone. They
can further indicate whether the addition of new factors to a base model provides better risk prediction than the base model alone, provided that all of the same individuals are being assessed by both models. The Bayes information criterion adds a penalty for using more variables in the model, so a price is paid for the addition of a variable to the model unless it substantially improves risk prediction and overcomes the penalty.

Risk Reclassification Analysis

One of the newest paradigms for assessing the utility of risk prediction models involves measurement of the proportion of individuals who are reclassified from 1 risk stratum (based on estimated risk provided from a first model) to a different risk stratum (based on estimated risk from a different model, or a model that has additional variables compared with the first model). Some of these risk reclassifications would be appropriate (based on subsequent observed events), meaning that some individuals who have events are reclassified to higher predicted-risk strata, and some who do not have events would be reclassified to lower predicted-risk strata. However, some reclassifications would be incorrect or inappropriate (eg, by moving future cases to lower predicted-risk strata). This is demonstrated in Table 3.

Pencina et al30 have proposed 2 indexes, the net reclassification improvement and the integrative discrimination index, to attempt to quantify the appropriateness and amount of overall reclassification. In general, the net reclassification improvement indicates how much more frequently appropriate reclassification occurs than inappropriate reclassification with use of the new model; for this test, a value of \( P<0.05 \) suggests that a significantly greater number are being reclassified appropriately than are being reclassified inappropriately. The integrative discrimination index (which is equivalent to the difference in \( R^2 \) between the 2 models being compared)31 can be thought of as indicating how far individuals are moving on average along the continuum of predicted risk. In many cases, models may have a significantly better net reclassification improvement, but if the integrative discrimination index is small (even if significant), then a given individual’s change in predicted risk with the new model will be small on average (those at higher baseline risk will likely change more than those at lower baseline risk). As an example, consider a new risk prediction model that is being compared with the FRS for stratifying a population into ATP-III risk categories. The new model might have a significant net reclassification improvement, reclassifying a net of 10% of people more appropriately, meaning that people who subsequently suffer events are upstaged in their predicted risk and people who do not suffer events are downstaged more often than cases are downstaged or noncases are.

\[ \text{Table 3. Theoretical Example of Appropriate* and Inappropriate† Risk Reclassification By a New Compared With an Existing CVD Risk Prediction Model} \]

<table>
<thead>
<tr>
<th>New Model Predicted Risk Stratum, %</th>
<th>0–5</th>
<th>6–&lt;10</th>
<th>10–20</th>
<th>&gt;20</th>
</tr>
</thead>
<tbody>
<tr>
<td>People who develop events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–5</td>
<td>No change</td>
<td>Reclassified up, appropriate</td>
<td>Reclassified up, appropriate</td>
<td>Reclassified up, appropriate</td>
</tr>
<tr>
<td>6–&lt;10</td>
<td>Reclassified down, inappropriate</td>
<td>No change</td>
<td>Reclassified, up, appropriate</td>
<td>Reclassified up, appropriate</td>
</tr>
<tr>
<td>10–20</td>
<td>Reclassified down, inappropriate</td>
<td>Reclassified, down, inappropriate</td>
<td>No change</td>
<td>Reclassified up, appropriate</td>
</tr>
<tr>
<td>&gt;20</td>
<td>Reclassified down, inappropriate</td>
<td>Reclassified, down, inappropriate</td>
<td>Reclassified down, inappropriate</td>
<td>No change</td>
</tr>
<tr>
<td>People who do not develop events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–5</td>
<td>No change</td>
<td>Reclassified up, inappropriate</td>
<td>Reclassified up, inappropriate</td>
<td>Reclassified up, inappropriate</td>
</tr>
<tr>
<td>6–&lt;10</td>
<td>Reclassified down appropriate</td>
<td>No change</td>
<td>Reclassified up inappropriate</td>
<td>Reclassified up inappropriate</td>
</tr>
<tr>
<td>10–20</td>
<td>Reclassified down, appropriate</td>
<td>Reclassified down, appropriate</td>
<td>No change</td>
<td>Reclassified up, inappropriate</td>
</tr>
<tr>
<td>&gt;20</td>
<td>Reclassified down, appropriate</td>
<td>Reclassified down, appropriate</td>
<td>Reclassified down, appropriate</td>
<td>No change</td>
</tr>
</tbody>
</table>

*Appropriate reclassification indicates that the new model reclassified some individuals who ultimately developed events into a higher predicted-risk group or reclassified some individuals who ultimately did not have events into a lower predicted-risk group.

†Inappropriate reclassification indicates that the new model reclassified some individuals who ultimately developed events into a lower predicted-risk group or reclassified some individuals who ultimately did not have events into a higher predicted risk group.
upstaged. But, if the integrative discrimination index is small (say, <1%), then most of the net reclassification is occurring immediately adjacent to the thresholds such as a change from a predicted risk of 19.6% with an old model to a predicted risk of 20.3% with a new model. This change would cross the ATP-III decision threshold for immediate lipid-lowering therapy but have no real impact in understanding or forecasting the patient’s risk (especially in light of the relatively arbitrary clinical decision thresholds imposed by ATP-III). The clinical significance of such small movements is heavily dependent on the threshold selected. Indeed, this scenario is what is often observed in current studies comparing older and newer CVD risk prediction scores.32

### Adding New Markers to Risk Prediction Models and Comparing Risk Prediction Models

Given the overall acceptable to excellent risk discrimination provided by existing multivariable risk prediction models, they remain the logical standard to which new risk markers must be added to demonstrate improvement or against which newer models must be compared. The methods that should be used to evaluate novel markers of cardiovascular risk were recently summarized and recommended by a special panel of the American Heart Association.33 All of the metrics discussed above should be examined and assessed for statistically significant improvement, and then clinical judgment should be applied to determine whether there is also clinically meaningful improvement that would affect decision making in a reasonable number of patients. That said, if the change in the C statistic is small for a new compared with an existing risk score or when a novel risk marker is added to an existing model, the addition of new risk markers can help reclassify some individuals, and this is often the group for which the addition of information from a new test is most clinically useful. In the Women’s Health Study,34 when C-reactive protein (CRP) was added to a model with traditional risk factors, the C statistic did not improve measurably for the whole population (the reported change was from 0.81 to 0.81, although this was statistically significant given the large sample size). However, women in the middle of the predicted risk spectrum in this cohort (FRS predicted risk of 5% to 9%) with a CRP >3.0 had an observed event rate that was equal to or greater than that of some women with an FRS-predicted risk of ≥10% (Figure 2). It should be noted, however, that the traditional risk model (FRS) determines a far greater magnitude of risk than does CRP level, indicating the important context provided by the traditional risk model. Similar results have been observed with the addition of parental history information to traditional risk scores,41 with minimal change in overall discrimination but potential benefit in intermediate-risk groups. More recent data suggest a larger change in the C statistic associated with the addition of coronary artery calcium scores to the FRS, with increments in the C statistic of 0.02 to 0.11 in different race/ethnic groups.38

Scores of novel biomarkers have undergone similar assessments in recent years. In essentially every case, the incremental value in discrimination and in calibration has been very small. For example, in 1 study, Folsom et al32 examined 19 novel putative risk markers for CHD in the Atherosclerosis Risk in Communities cohort. Considered individually, the vast majority of the markers added between 0.000 and 0.005 to the C statistic for the traditional risk model, and several markers actually decreased the C statistic slightly compared with that for the traditional model. Calibration was not improved significantly by the addition of these markers.

Because single risk markers have not added substantially to risk prediction, a number of investigators have examined the addition of multiple markers simultaneously to traditional risk equations. In the Cardiovascular Health Study, Shlipak et

### Assessments of the FRS

In most cohorts studied to date, the FRS provides very good discrimination, as evidenced by C statistics that typically range from 0.75 to 0.80.34–38 Invariably, the FRS and similar risk scores discriminate risk better for women than they do for men, often having C statistics in excess of 0.80 for women. In addition, the FRS has good discrimination in most populations (including those from outside the United States) in which it has been studied because it contains age as a covariate and because of the fairly universal associations between CVD and the major traditional risk factors the FRS also incorporates. The calibration of the FRS in diverse populations does differ more substantially, however, because of variable “background” CVD incidence rates in different settings. Studies to date indicate that the FRS discriminates CHD risk very well and is well calibrated for a wide range of white and black populations in the United States. For other populations such as Asian Americans, American Indians, Hispanic Americans, and native Chinese, discrimination remains acceptable, but the FRS tends to overestimate risk.17,39 However, with simple steps to recalibrate the FRS model, it performs quite well in both discrimination and calibration.17 One of the benefits of this approach is that recalibration of the FRS to a local population currently can help avoid the effort and expense of developing novel cohorts and awaiting information on risk factor relationships.

### Examples of Attempts to Improve Risk Prediction Through the Addition of Novel Markers to Existing Risk Scores

A substantial body of literature over the last 5 years has been devoted to examining the addition of newer risk markers to traditional risk scores. Essentially all of the single additional markers studied to date have yielded little additional clinical benefit when considered as screening tests across the entire spectrum of risk. Given that most novel markers are correlated with traditional risk factors and therefore do not have excessively high independent odds ratios (or relative risks), it is difficult for them to change the C statistic substantially when added to traditional risk models.30 However, within subgroups predicted to be at intermediate risk by traditional models, the addition of new risk markers can help reclassify some individuals, and this is often the group for which the addition of information from a new test is most clinically useful. In the Women’s Health Study,34 when C-reactive protein (CRP) was added to a model with traditional risk factors, the C statistic did not improve measurably for the whole population (the reported change was from 0.81 to 0.81, although this was statistically significant given the large sample size). However, women in the middle of the predicted risk spectrum in this cohort (FRS predicted risk of 5% to 9%) with a CRP >3.0 had an observed event rate that was equal to or greater than that of some women with an FRS-predicted risk of ≥10% (Figure 2). It should be noted, however, that the traditional risk model (FRS) determines a far greater magnitude of risk than does CRP level, indicating the important context provided by the traditional risk model. Similar results have been observed with the addition of parental history information to traditional risk scores,41 with minimal change in overall discrimination but potential benefit in intermediate-risk groups. More recent data suggest a larger change in the C statistic associated with the addition of coronary artery calcium scores to the FRS, with increments in the C statistic of 0.02 to 0.11 in different race/ethnic groups.38

Scores of novel biomarkers have undergone similar assessments in recent years. In essentially every case, the incremental value in discrimination and in calibration has been very small. For example, in 1 study, Folsom et al32 examined 19 novel putative risk markers for CHD in the Atherosclerosis Risk in Communities cohort. Considered individually, the vast majority of the markers added between 0.000 and 0.005 to the C statistic for the traditional risk model, and several markers actually decreased the C statistic slightly compared with that for the traditional model. Calibration was not improved significantly by the addition of these markers.

Because single risk markers have not added substantially to risk prediction, a number of investigators have examined the addition of multiple markers simultaneously to traditional risk equations. In the Cardiovascular Health Study, Shlipak et
al reported that the addition of 6 novel biomarkers [including interleukin-6, CRP, fibrinogen, lipoprotein(a) and factor VIII levels, plus presence of anemia] to a traditional risk model improved the C statistic from 0.73 to 0.74 in older individuals with chronic kidney disease but decreased the C statistic from 0.73 to 0.72 in those with normal renal function. In the Framingham study, the addition of brain natriuretic peptide levels plus microalbuminuria to a traditional risk model for CVD events yielded an increase in the C statistic from 0.76 to 0.77.44

More recent studies have extended these types of analyses to assess reclassification as well. In a cohort of 5067 Swedish individuals free of CVD at baseline who were followed up for a median of 12.8 years, models with traditional risk factors had C statistics of 0.758 and 0.760 for CVD and CHD events, respectively. When 2 independently significant novel biomarkers were added to the traditional models, the C statistics improved by 0.007 and 0.009 for CVD and CHD events, respectively, indicating minimal clinical utility. Likewise, the proportion of participants reclassified was modest (8% overall for CVD risk and 5% for CHD risk). The net reclassification improvement was 0% for CVD events and 4.7% for CHD events, indicating no or only minimal net appropriate improvement in risk classification. Greater improvements in reclassification were observed in analyses restricted to intermediate-risk individuals, but any correct reclassification was due almost entirely to downward reclassification of individuals who did not experience events (yet). Greater amounts of net reclassification have been observed with the addition of some novel biomarkers in other cohorts, especially when age has been effectively removed from the risk model (either through matching in case-control studies or the use of cohorts with very narrow age ranges).45

Examples of Newer Risk Prediction Models
As noted above, a surfeit of published risk prediction scores for various CVD or CHD end points exist that have been derived from a variety of different populations. As expected, CVD risk scores tend to perform best in populations similar to those from which they were derived. One recent example, the Reynolds risk score for women, used state-of-the-art statistical methodologies to derive and compare a new risk score in women health professionals enrolled in the Women’s Health Study clinical trial. Although an unbiased approach was used to select variables for this model, it is noteworthy that the traditional risk factors were still the strongest predictor variables for a broad CVD end point. In addition, CRP levels and parental history also were selected for inclusion in the model on the basis of improvement in the Bayes information criterion. When compared with a Framingham-like model recalibrated for the end point of this study, overall discrimination and calibration of the 2 models were similar. Internal validation of the Reynolds risk score revealed that 5.8% of 8149 women were reclassified when the Reynolds risk score was applied (compared with the FRS), with about half of those being reclassified upward and half downward. The clinical implications of downward reclassification are unclear because withholding therapy from those potentially at higher risk based on traditional risk factors does not seem warranted given available clinical trial data. Likewise, statin medications now have a long track record of safety and have been estimated to be very cost-effective or even cost-saving to levels of predicted risk as low as 10%.46,47 Thus, the generalizability and utility of this score and many others are still being assessed.

Limitations of 10-Year Risk Prediction Models
Ten-year risk estimation represents a substantial improvement over clinician judgment alone for appropriate risk stratification.16,49 However, it has some acknowledged limitations. For example, because age is the most heavily weighted variable in 10-year risk models derived from populations that span the adult age spectrum, in younger adults (men <45 years of age and women <65 years of age), modest elevations in risk factors have little effect on 10-year risk.50,51 Even younger adults with substantial risk factor burden may still have 10-year risk estimates well below 10%,
although their remaining lifetime risks may exceed 50% on the basis of these risk factors. This is not a problem of the 10-year risk score per se nor incorrect risk prediction but rather a function of the decision thresholds that are imposed on the risk estimates and the short time horizon imposed by 10-year risk estimation.

The magnitude of risk factor levels needed to reach moderately high (>10% 10-year predicted risk) and high-risk (>20%) thresholds in the ATP III risk assessment algorithm has recently been investigated. Cavanaugh-Hussey et al entered ranges of risk factor levels into the ATP III risk assessment tool for men and women from 30 to 75 years of age. For almost all combinations of risk factors, even with extreme values, nonsmoking men <45 years of age and essentially all women <65 years of age have 10-year predicted risks below 10%. Thus, many younger patients with significant risk factor burden do not reach treatment thresholds based on current ATP-III recommendations. A similar study has recently evaluated the updated Framingham risk profile23 for global CVD (not just CHD) risk with somewhat similar findings, although the use of an expanded CVD end point allows more men and particularly more women in their 50s to exceed 10% and 20% predicted 10-year risk.

Several alternatives exist to address some of the limitations of 10-year risk estimates, particularly for younger adults. One possible solution would be to lower treatment thresholds for younger adults (eg, treat those <50 years of age with 10-year risk >5%). However, 10-year risk equations do not stratify or identify those at greater long-term risk well, particularly for men. A second alternative that has been suggested is to perform subclinical disease imaging in essentially all middle-aged adults to identify those with premature atherosclerosis. Whereas this option is attractive at first glance, because one is identifying the actual disease of interest, issues of cost, radiation exposure (for coronary calcium assessment), technician and reader reliability (for some modalities such as carotid intima-media thickness scanning), and uncertainty about the need for repeat imaging make this strategy a matter for continued active investigation. Recent statements from the American Heart Association and the American College of Cardiology recommend reserving imaging for those at intermediate risk when there is uncertainty about the need or desirability of starting lipid-lowering therapy. A final alternative is to change the horizon for risk estimation to provide long-term or lifetime risk estimates for CVD end points.

### Long-Term and Lifetime Risk Estimation as an Adjunct to 10-Year Risk Estimation

Because of the limitations of a sole focus on 10-year risk, various guidelines1,55 have recommended consideration of lifetime risks when making decisions about preventive efforts for CVD. More recently, Canadian2 and AHA3 guidelines suggest that clinicians consider risk factor burden in the context of patients’ lifetime risks for CVD. Thus, there is now enhanced interest in lifetime CVD risk prediction, but the methods for producing robust lifetime risk estimates and the available data have been limited until recently.

#### Rationale for Lifetime Risk Estimation

Traditional statistical and epidemiological methods for estimating risks such as those used in currently available 10-year CVD risk scores typically use Kaplan-Meier56 and Cox proportional-hazards57 methods. However, these methods do not account for competing risks (eg, death from non-CVD causes) that become increasingly relevant over longer-term follow-up. Methods of estimating lifetime risk do account for competing risks and therefore can provide more “real-life” risk prediction over the long term.59 Lifetime risk estimates are useful in assessing the burden of disease in a population, predicting the future burden of disease, and directly comparing lifetime risks between common diseases.58,59

#### Recent Studies

A number of recent studies25,52,60–66 have used lifetime risk estimation or competing-risk methods to examine long-term and lifetime risks for CHD and CVD. In the Framingham Heart Study, lifetime risks for atherosclerotic CVD were estimated by overall risk factor burden at 50 years of age.25 Participants were stratified into 5 mutually exclusive categories, as shown in Figure 3. Compared with participants with

---

**Figure 3.** Remaining lifetime risks for atherosclerotic CVD in men and women at 50 years of age. Optimal risk factors (RFs) are defined as untreated total cholesterol <4.65 mmol/L (<180 mg/dL), untreated blood pressure <120/80 mm Hg, nonsmoker, and nondiabetic. Not optimal risk factors are defined as untreated total cholesterol 4.65 to 5.15 mmol/L (180 to 199 mg/dL), untreated systolic blood pressure 120 to 139 mm Hg or diastolic blood pressure 80 to 89 mm Hg, nonsmoker, and nondiabetic. Elevated risk factors are defined as untreated total cholesterol 5.16 to 6.19 mmol/L (200 to 239 mg/dL), untreated systolic blood pressure 140 to 159 mm Hg or diastolic blood pressure 90 to 99 mm Hg, nonsmoker, and nondiabetic. Major risk factors are defined as total cholesterol ≥6.20 mmol/L (≥240 mg/dL) or treated; systolic blood pressure ≥160 mm Hg, diastolic blood pressure ≥100 mm Hg, or treated; smoker; or diabetic. Reprinted from Lloyd-Jones et al25 with permission from the publisher. Copyright © 2006, the American Heart Association.
≥2 major risk factors, those with optimal levels at 50 years of age had substantially lower lifetime risks (5.2% versus 68.9% in men; 8.2% versus 50.2% in women) and markedly longer median survivals (>39 versus 28 years in men; >39 versus 31 years in women). In fact, for those with any risk factor levels above optimal, 10-year risks were generally <10%, whereas remaining lifetime risks were substantial.29 On the basis of current ATP-III recommendations, such patients may not receive pharmacological lipid-lowering therapy despite the high likelihood that they will develop end-organ damage from atherosclerotic disease during their remaining lifespan. Even if they do receive lipid-lowering therapy, the message that they have a low 10-year risk may be insufficient to motivate them to adhere to lifestyle recommendations or to comply with therapy.

The concept of identifying younger individuals at low short-term but high lifetime risk was recently validated by comparison with subclinical atherosclerosis imaging data. In the Coronary Artery Risk Development in Young Adults Study,67 >90% of participants 32 to 47 years of age had a 10-year predicted risk <10%. However, among this group, approximately half had high predicted lifetime risk (≥39%). Even at these younger ages, participants with low 10-year but high lifetime risk estimates had significantly greater burden of carotid artery intima-media thickness and coronary artery calcification and greater progression of subclinical atherosclerosis compared with participants with low 10-year and low lifetime predicted risk.67 Data from the National Health and Nutrition Examination Survey 2003 to 2006 reveal that a majority (56%) of US adults, or 87 000 000 people, have such a low 10-year but high lifetime predicted risk for CVD.68

Pencina et al69 recently published a quantitative method for estimating long-term (30-year) risks for CVD that also accounts for competing risks. Validation of these equations in other settings is needed, but they may prove to be a powerful tool for adjunctive risk assessment over and above current 10-year risk estimates.

Thus, long-term and lifetime risk estimation may represent a potentially important new tool in our ability to identify patients at risk for CVD and to provide more complete information for risk communication. They should be considered for inclusion in future clinical prevention guidelines as an adjunct to current 10-year risk estimates. However, further study is needed to understand the clinical utility and impact of both short-and long-term risk estimation and communication strategies.

New Directions
Research in CVD risk prediction is actively ongoing. One area of interest that is being pursued is the derivation of age-specific risk equations for groups in narrow windows of age rather than the current models derived from populations with a wide range of ages in which age is included as a predictor variable. Without age in the risk model and with a focus on a specific age group, other risk markers may prove to be more potent predictors of events, as was recently shown in a cohort of Swedish men all of the same age.45 Other recent developments include the general consensus that risk models should focus on predicting global CVD events, not just CHD events. Such a focus will serve to identify more individuals at risk in the short term, overcoming some of the limitations of 10-year CHD risk prediction models noted above. This will be true for women especially, given their propensity to be at risk for stroke and heart failure earlier than CHD, in contrast with men.51

The Future of Risk Prediction in CVD Prevention
The field of cardiovascular risk prediction has now arrived at a crossroads. It is currently unclear whether further efforts to refine risk estimation models and use more complex, and individualized, risk-based prevention schemes will yield substantial improvements toward “personalized risk assessment.” The extreme alternative to personalized medicine may be to adopt population-level approaches in which all individuals above a certain age are treated to prevent CVD because almost all individuals are at substantial lifetime risk.27 Active investigations of polypill strategies based on age alone70 or based on estimated risk are ongoing. In these paradigms, risk assessment could still be a useful tool to improve communication and motivate physician and patient efforts in CVD prevention, to promote understanding of the need for adherence to preventive medication strategies, and to highlight and sustain efforts at therapeutic lifestyle modification.

Ultimately, it is clear that both population-level approaches and high-risk approaches, which depend on robust and accurate risk prediction models, are needed to curb the ongoing epidemics of CVD and CHD, which may be worsening for the first time in 4 decades.71 Substantial efforts are needed to optimize risk prediction models and prevention algorithms for risk communication, patient motivation, and clinical decision making. Whether and how the use of such models influences physician and patient behavior and improves patient outcomes is the subject of a future article in this series.

Disclosures
None.

References


**Key Words:** epidemiology prevention risk factors
Cardiovascular Risk Prediction: Basic Concepts, Current Status, and Future Directions
Donald M. Lloyd-Jones

Circulation. 2010;121:1768-1777
doi: 10.1161/CIRCULATIONAHA.109.849166

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/121/15/1768

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/