

## Individual Risk

Ralph H. Stern, MD, PhD

From the Divisions of Cardiovascular Medicine and Molecular Medicine and Genetics, Department of Internal Medicine, University of Michigan, Ann Arbor, MI

Patients don't have an "individual risk" or unique probability of an outcome. Outside Mendelian inheritance, risks are conditional probabilities and differ as the risk factors included differ, at times substantially. This lack of reliability is an inherent limitation and is not resolved by including additional risk factors. Groups of like individuals need to be assembled to measure the probability of an outcome. Many groups, like any individual, can be identified, eg, groups of the same age, sex, race, or any combination of these attributes (or any others). That each of these groups may have different risk means there is no such thing as

individual risk. This issue was identified by John Venn in 1866 and is known as the reference class problem. Models relate risk factors to outcomes in populations. The number calculated for an individual should not be reported as their individual or true risk, nor should it be used as the sole criterion for clinical decisions. Instead, Feinstein proposed relying on clinically important subgroups. An example would be utilizing an individual's blood pressure as the primary determinant of hypertension treatment decisions, not an unreliable individual risk estimate. *J Clin Hypertens (Greenwich)*. 2012;14:261–264. ©2012 Wiley Periodicals, Inc.

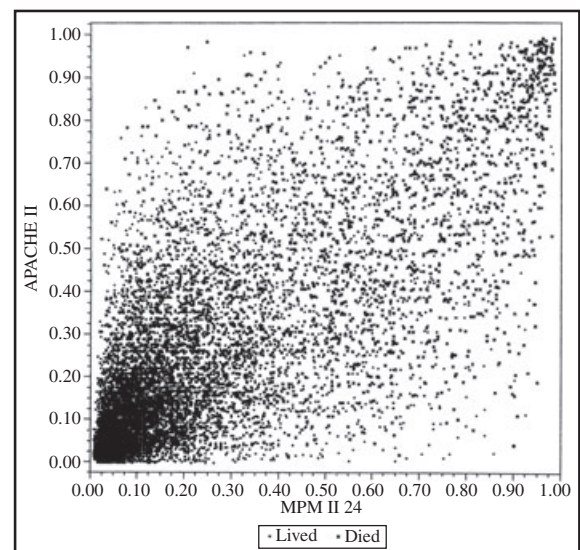
The Framingham Heart Study played a key role in identifying risk factors for coronary artery disease, and Framingham investigators coined the term factors of risk.<sup>1</sup> Logistic regression was developed to analyze the epidemiologic data from Framingham.<sup>2</sup> By 1973, the logistic regression model included sex, age, cigarette smoking, blood pressure, serum cholesterol, glucose tolerance, and electrocardiography (left ventricular hypertrophy). Based on this model, the American Heart Association published the *Coronary Risk Handbook: Estimating Risk of Coronary Heart Disease in Daily Practice*.<sup>3</sup> It stated: "The purpose of this Handbook is to provide the physician with a method for easily estimating risk of coronary heart disease in patients who have no clinical evidence of coronary heart disease, and for guiding his choice of preventive management." The handbook included tables that presented the 6-year risk of coronary heart disease for individuals based on their risk factors. It contained a caution: "The figures given in the tables should be taken only as guides to risk. They are accurate estimates of group experience but not necessarily the experience of any individual." The publication of this handbook represents an historic bench-to-bedside transition. An epidemiologic research method associating risk with specific risk factors in populations became a clinical method providing risk estimates for individuals. During the past 40 years, individual risk estimation has been broadly accepted in medicine.

**Address for correspondence:** Ralph H. Stern, MD, PhD, CVC  
Cardiovascular Medicine, 1500 East Medical Center Drive SPC5853, Ann Arbor, MI 48109-5853  
**E-mail:** stern@umich.edu

**Manuscript received:** November 9, 2011; **Revised:** December 30, 2011; **Accepted:** December 31, 2011  
**DOI:** 10.1111/j.1751-7176.2012.00592.x

### DISCORDANCE OF INDIVIDUAL RISK ESTIMATES

There is much evidence that comparable risk stratification methods give different individual risk estimates. In a classic paper that has been ignored, Lemeshow and colleagues<sup>4</sup> evaluated 3 different methods for predicting hospital mortality in a large cohort of intensive care unit patients. All 3 methods were comparable in terms of calibration and discrimination; however, a scatter plot (Figure) of the predicted risk for 2 of the methods showed an astounding amount of discordance between the individual risk estimates. The other 2



**FIGURE.** Scatterplot of probability of hospital mortality from APACHEII and MPMII<sub>24</sub>. Reproduced with kind permission from Lemeshow et al.<sup>4</sup>

scatter plots were described as similar. Other examples of the discordance of individual risk estimates can be found in a recent review<sup>5</sup> and analysis.<sup>6</sup>

More recently, understanding the discordance of individual risk estimates has become important for interpreting reclassification analysis. In this approach, 2 predictive models are compared after separating patients into risk categories considered clinically important and cross-tabulating the results for the 2 models. When the predictions are discordant, individuals will be assigned to different categories by the different methods. The different assignments are often interpreted as evidence that one model is superior, rather than that different predictors give different predictions.

Discordance can be produced by differences in accuracy or discrimination. Thus, accuracy and discrimination need to be assessed to interpret discordance. It makes little sense to evaluate discordance when models are inaccurate. The discordance depicted by Leshow and colleagues did not reflect differences in accuracy or discrimination.

## THE CONCEPT OF INDIVIDUAL RISK

Risk is the probability of an undesirable outcome. Individual risk is the probability an individual will experience an undesirable outcome. However, an individual either does or does not experience an undesirable outcome, so risk can never be determined for an individual. In spite of this, it is assumed that a unique probability, a true risk, of an undesirable outcome for an individual exists. From this perspective, different models used in the clinic to calculate individual risks are estimating this unknown true risk with some error. Pepe<sup>7</sup> has dismissed the concept of true risk as having “major scientific problems” and as “interesting” but “nebulous.”<sup>8</sup>

Scientifically there is no reason to believe that a unique probability or true risk of an undesirable outcome in an individual exists. That is only plausible when some simple physical process is involved, such as flipping a coin. The probabilities of Mendelian inheritance, resulting from the segregation of chromosomes, are likely the only such probabilities encountered in medicine.<sup>8</sup>

Since risk cannot be measured in an individual, there is no way to experimentally verify any of the individual predictions provided by a model. This can only be achieved by assembling a group of patients like the individual. But there are many groups like an individual that can be identified, eg, groups of the same age, sex, race, or any combination of these attributes (or any others). That each of these groups can have different probabilities of an outcome means a unique individual risk cannot be defined. This issue was identified by John Venn as early as 1866 and is known as the reference class problem.<sup>5</sup> Von Mises<sup>9</sup> gave an example for the probability of death. From experience, life insurance companies knew that 0.011 of 40-year-old men who passed a medical examination

and were issued insurance would die in the next year. But he described it as “utter nonsense” to say any individual had a 0.011 probability of dying. For a group of 40-year-old men and women, a lower probability would be expected and any 40-year-old man belongs in this combined sex group as much as in the single sex group. And he could be included in a large number of other groups that would have yet other probabilities of death.

It is important to recognize that even if all known (and unknown) risk factors were to be included in a risk stratification model, it would be more discriminating (and perhaps more useful clinically), but the resulting individual risk estimates would not become true risks.

## MODELS

Models were developed for research on disease in populations. Essentially they assign risks to subgroups defined by the included predictors. This risk stratification of a population may be useful for efficiently allocating resources within a population. For example, the 2 methods of predicting intensive care unit (ICU) mortality discussed previously provide equivalent risk stratification of the patient population. That is, each assigns a similar fraction of the patient population to different risk strata. Thus, if economic considerations supported treating only patients above a given level of risk, either method could be used. Even though each method assigns the same fraction of the patient population to the high-risk subgroup, the compositions of the high-risk subgroups differ. As a result, some individuals assigned treatment by one method would be denied treatment by the other.

But it is a mistake to use terms such as “individual risk” or “true risk” for the number we calculate from a risk model. Calling them event frequencies would be preferable.

Mathematically inferior models produce subgroups of near-average risk, while mathematically superior models produce subgroups with widely varying risks. This discrimination is a function of the number and potency of included predictors. The commonly used metrics of discrimination, the c-statistic and receiver operating characteristic curve area under the curve, reflect this dispersion. More discriminating models producing broader risk distributions may be advantageous for clinical use.

The potential benefit of adding new risk factors (eg, coronary artery calcium) to an existing model (eg, Framingham risk model) is a model producing a broader risk distribution.<sup>10</sup> But assuming both models correctly assign risk, the rank order of individuals in the population cannot be maintained. This is because a group of individuals that was correctly assigned a 12% risk, for example, cannot subsequently be correctly assigned a 15% risk. The benefit of improving the population risk stratification is achieved at the cost of shuffling the rank order of the individuals in the population.

This redistribution is even seen when discrimination is not improved. Mihaescu and associates<sup>10</sup> characterize this as “an updated model, compared with the additional model, simply makes different errors, not fewer errors.”

## DISCUSSION

The key points that there is no such issue as individual risk and that different predictors give different predictions have been made previously by the Framingham investigators: “It must be understood that there is no such thing as an unconditional probability of cardiovascular disease developing, nor any conditional probability that may not alter if other factors are entered into consideration.”<sup>11</sup>

It is said that a man with one watch knows what time it is, while a man with two watches is never sure. If models are understood to generate individual risk estimates, discordance presents a dilemma for the clinician. Assuming both models are accurate, then their estimates are equally valid. The mathematic problem of risk stratifying a population does not have a unique solution.

This may become especially troublesome when additional information is obtained sequentially as individual risk estimates may rise or fall, producing contradictory information and therapeutic recommendations.<sup>12</sup>

It is important to recognize the origin of these differences and not misinterpret their importance. For example, the observation that some individuals in intermediate-risk categories, as estimated by Framingham risk models, may be in low- or high-risk categories, as estimated by other methods, is expected and by itself has no importance. Yet this reclassification has been interpreted as meaningful and was used as a criterion in a recent assessment of the clinical utility of emerging risk factors.<sup>13</sup>

A number of approaches to the problem of discordant risk estimates might be or have been proposed. One option would be to only use a single method, avoiding the generation of discordant individual risk estimates. The perceived inadequacies of current methods do not provide a justification for obtaining additional estimates. Although current methods may omit one or more risk factors, they may be perfectly adequate for risk-stratifying a population. If newer methods including additional risk factors significantly improve discrimination at reasonable cost, then they could replace current methods. If a current method is not accurate (calibrated) for some demographic or population, it should be refitted (remodeled) or recalibrated. A new method is not required to deal with this problem. Finally, current methods may not identify patients at low risk who will have events, but this is an inherent limitation of probabilistic methods.

A second option would be to only add a new risk marker “after adequate counseling of the patient of the uncertain benefits and risks of reclassification, and

only if the patient and physician understand, discuss, and are amenable to the treatment implications of risk reclassification.”<sup>14</sup>

A third option would be to average the discordant individual risk estimates. However, the discordance does not represent statistical variation around a “true” individual risk value and averaging will reduce the discrimination provided by each of the models.

A fourth option is to provide the patient with the discordant estimates.<sup>15</sup> Clearly this would be confusing to both patient and physician. An advantage of this approach would be that patients and clinicians would become familiar with the discordance of individual risk estimates. If the discordant estimates were considered equally valid and lead to different preventive measures, the patient could choose the preventive measure they preferred. These could range from taking a statin, to taking tamoxifen, to having risk-reducing breast surgery.

A fifth option for discordant estimates is to utilize the highest probability estimate, as this “assigns a final risk level based on the model that best accounts for a client’s specific risk factor history.”<sup>16</sup> A concern with this approach is that each model may be calibrated or accurate, but a policy of using the maximum estimate may not. If enough models are considered, too many individuals may end up with above-average estimates, analogous to the fictional Lake Wobegon, where all the children are above average.

A sixth option would be to measure an additional risk factor selectively in patients at intermediate risk and, if the risk factor level is high, assign the individual to a higher-risk category.<sup>17</sup> However, once a continuous risk estimate is developed, categorization destroys information. Prior to the categorization, individuals at the upper and lower boundaries of a category were readily differentiated, while individuals on either side of a category boundary were understood to be similar. Thus, if there were a simple method for revising risk estimates based on the additional risk factor, it should be applied to the uncategorized risk estimates. By only increasing estimates, this approach will lead to a loss of accuracy via the Lake Wobegone effect. Finally, it is difficult to justify revising only the risk estimates of those at intermediate risk. A better approach would be to use a multivariate method that includes the additional risk factor. Kooter and colleagues<sup>18</sup> have raised these points and importantly demonstrated that there is no straightforward way to update risk estimates.

Because the estimates depend on the model chosen, Lemeshow and associates<sup>4</sup> concluded that they should not be used to make patient care decisions, eg, withdrawing support from an ICU patient. A high-risk subgroup identified by one model is just one possible high-risk subgroup, not the only high-risk subgroup. Thus, any such subgroup is not uniquely entitled to treatment, such as a ventilator in a pandemic, ICU

support, or an organ transplant. Clinicians who use these models to make patient care decisions need to be aware of their limitations.

Models relate included risk factors to outcomes in populations. They are best understood to be providing one of many possible risk stratifications of a population, any of which may be useful for allocating resources efficiently. However, the number calculated for an individual should not be reported as their individual or true risk, nor should it be used as the sole criterion for clinical decisions.

Feinstein<sup>19</sup> noted that individual risk estimates could have “striking differences” and thus “that few clinical prognosticators would want to make predictions” using them. “Instead, clinicians would want the greater predictive “security” that is possible when the individual forecasts are made from results in a pertinent “resemblance” subgroup.” Such an approach would be simpler to implement and would have the appeal of aligning preventive treatments with risk factors. An example would be utilizing an individual’s BP as the primary determinant of hypertension treatment decisions, not an unreliable individual risk estimate.

## References

1. Wilson P, Greenberg H. Interview of William Kannel, MD. *Prog Cardiovasc Dis*. 2010;53:4–9.
2. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in framingham. *J Chronic Dis*. 1967; 20:511–524.
3. American Heart Association. *Coronary Risk Handbook: Estimating Risk of Coronary Heart Disease in Daily Practice*. New York, NY: American Heart Association; 1973.
4. Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med*. 1995;21:770–776.
5. Stern RH. The discordance of individual risk estimates and the reference class problem. arXiv:1001.2499v1[q-bio.QM]. <http://www.arXiv.org>. Accessed January 30, 2012.
6. Gurm HS, Kaufman SR, Smith DE, et al. Different risk models predict markedly different probability of death for the same patient: implications of using risk models for individual patients. *Eur Heart J*. 2011;32(suppl 1):742.
7. Pepe M. Rejoinder to Nancy Cook’s comment on “measures to summarize and compare the predictive capacity of markers.” *Int J Biostat*. 2010; 6: Article 25. <http://www.bepress.com/ijb/vol6/iss1/25>. Accessed January 30, 2012.
8. Pepe MS. Problems with reclassification methods for evaluating prediction models. *Am J Epidemiol*. 2011;173:1327–1335.
9. Von Mises R. *Probability, Statistics and Truth*. New York, NY: Dover Publications, Inc; 1981: 17.
10. Mihaescu M, van Zitteren M, van Hoek M, et al. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. *Am J Epidemiol*. 2010;172:353–361.
11. Kannel WB, D’Agostino RB, Sullivan L, Wilson PWF. Concept and usefulness of cardiovascular risk profiles. *Am Heart J*. 2004;148:16–26.
12. Mihaescu R, van Hoek M, Sijbrands EJG, et al. Evaluation of risk prediction updates from commercial genome-wide scans. *Genet Med*. 2009;11:588–594.
13. Helfand M, Buckley DI, Freeman M, et al. Emerging risk factors for coronary heart disease: a summary of systematic reviews conducted for the U.S. Preventive services task force. *Ann Intern Med*. 2009;151:496–507.
14. O’Malley PG, Redberg RF. Risk refinement, reclassification, and treatment thresholds in primary prevention of cardiovascular disease incremental progress but significant gaps remain. *Arch Intern Med*. 2010;170:1602–1603.
15. McTiernan A, Kuniyuki A, Yasui Y, et al. Comparisons of two breast cancer risk estimates in women with a family history of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 1997;50:547–556.
16. Euhus DM, Leitch AM, Huth JF, Peters GN. Limitations of the Gail model in the specialized breast cancer risk assessment clinic. *Breast J*. 2002;8:23–27.
17. Pearson TA, Mensah GA, Alexander RW, et al. Markers of inflammation and cardiovascular disease, application to clinical and public health practice, a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*. 2003;107:499–511.
18. Kooter AJ, Kostense PJ, Groenewold J, et al. Integrating information from novel risk factors with calculated risks the critical impact of risk factor prevalence. *Circulation*. 2011;124:741–745.
19. Feinstein AR. *Multivariable Analysis an Introduction*. New Haven, CT: Yale University Press; 1996:566.