

COMMENTARY

Comments on ‘Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond’

by M. J. Pencina, R. B. D’Agostino Sr, R. B. D’Agostino Jr, R. S. Vasan, *Statistics in Medicine*

From: Philip Greenland*

*Feinberg School of Medicine, Departments of Preventive Medicine and Medicine,
Northwestern University, Chicago, IL, U.S.A.*

In 1998, I co-authored an editorial [1] that identified three major problems related to cardiovascular risk assessment and risk reduction: (a) measurement of risk factors and collection of clinical data relevant to patient risk; (b) interpretation of risk-related data with estimation of risk in absolute terms (e.g. risk of an event per year) as well as relative terms (i.e. low, intermediate, or high compared with others of the same age and sex); and (c) on the basis of risk estimation results, intervention to minimize disease risk or to prevent risk factor development in the future. We noted that problems existed in each of these critical steps, and called for further research in each of these areas to achieve better control of cardiovascular risk factors in the clinical setting. Since the editorial appeared, much work has focused on the second of the three issues above—improvements in the statistical assessment and interpretation of risk. The article by Pencina and colleagues in this issue [2] contributes to this area and is a definite step forward in the assessment of new risk factors. It remains to be seen, of course, whether it can have an impact on either of the other two steps, each of which remains highly important for the clinical application of the new statistical approach.

As Pencina *et al.* point out [2], the critical statistical aspects of any risk assessment tool are discrimination and calibration. Discrimination is probably most important to the clinician—how well can the practitioner distinguish a diseased (or a future diseased) patient from a person who is not (or will not be) affected? For screening purposes, the clinician would like to have a test (or assessment tool) that is highly sensitive, capable of detecting nearly all—or all—future cases. In the type of cases discussed by Pencina *et al.*, coronary heart disease, preventive treatments are available and are highly effective, but they must be applied in a cost-effective manner, so the clinician would also prefer to have an assessment tool that is also highly specific. As noted by Pencina and colleagues, the area under the ROC curve (AUC) is a common summary statistic that describes the discrimination of a diagnostic or prognostic test, and the number has actual meaning that a clinician can understand. The AUC is the probability that a randomly selected person from the affected (or soon to be affected) group will have a higher test score than a person randomly

*Correspondence to: Philip Greenland, 680 North Lake Shore Drive, Suite 1102, Chicago, IL 60611, U.S.A.

selected from the non-cases. This measure directly relates to the degree of overlap of scores for affected and non-affected and is a function of the sensitivity–specificity results for *all possible test results* for the assessment tool.

In practice, clinicians do not usually focus on all possible test cutoff values, however, as they tend to be most interested in results clustering near important treatment (or decision-making thresholds). As noted more than 30 years ago by Framingham investigators, in their very first paper on risk assessment in the clinical medicine literature [3], by focusing on people with the very highest risk scores, one was able to identify a very large proportion of future cases of coronary heart disease and other vascular complications of atherosclerosis. Nonetheless, if the scores for affected and unaffected overlap a great deal, a cutoff that detects most of the cases will still identify a very large number of ‘false positives’ who will require treatment without evidence of benefit. Thus, treatment of many false positives makes the treatment approach progressively less cost effective.

For virtually all single risk factors, the decision-making situation is dismal. We showed that more than 90 per cent of patients who will eventually experience coronary artery disease have at least one major risk factor, including unfavorable levels of cholesterol, blood pressure, cigarette smoking, or diabetes [4]. The rub here was that over 80 per cent of non-cases also had at least one of these risk factors present, so while sensitivity was high, specificity was poor. The same rule applies to new tests such as C-reactive protein, lipoprotein (a), and homocysteine [5]. As noted by Pencina *et al.* [2] and many others too numerous to cite, multiple risk factors can be combined to create a risk assessment method that is both highly sensitive and somewhat more specific than is the case for single risk factors. This is why risk assessment using tools such as the Framingham Risk Score (FRS) has generally superseded risk estimation using single measures of risk, no matter whether they be lipids, inflammatory markers, thrombotic factors, or genetic markers. All single measures perform less well than multi-marker tests [6].

The advantage of the new statistical approach of Pencina *et al.* [2] is that it focuses on the distribution of scores that are of greatest interest to clinicians, those at the critical decision thresholds. The new approach demonstrates that it is feasible for a statistician to focus on these critically important zones of interest and discover whether a new test ‘moves a patient’ from a zone of lower risk to a zone of higher risk, thus crossing a treatment threshold. However, with any test, there is the possibility of moving a patient both to a higher risk (into a more intensive treatment zone) and also to a lower risk (and out of the treatment zone). If the net result is that nearly equal numbers are moved in each direction, with the net result almost unchanged, then the test is of little overall value. This is presumably what the AUC tells us with most new tests—no real net improvement in discrimination. But, by focusing closely on the treatment thresholds, the new approach of Pencina *et al.* [2] may move this field closer to its goal of better discrimination of affected compared with unaffected patients. For this, Pencina *et al.* are to be congratulated for this new approach.

At the same time, before getting too excited over this new statistical approach, it remains to be seen also whether a new test can lead to improvements in the clinical control of cardiovascular risk. Caution is advised as the results to date are not encouraging. For example, a study in which coronary calcium measurement was added to standard cardiovascular risk failed to lead to improved cardiovascular risk control unless also combined with intensive risk factor management [7]. The message was that better testing strategies alone did not result in better clinical outcomes. Future studies should also assess whether new tests, or new testing algorithms, can also be effectively incorporated into clinical practices and ultimately lead to improved patient outcomes. Nevertheless, new statistical approaches are one of the key steps along this process, and Pencina *et al.* [2] have provided an important step forward in this regard.

COMMENTARY

REFERENCES

1. Greenland P, Grundy S, Pasternak RC, Lenfant C. Problems on the pathway from risk assessment to risk reduction. *Circulation* 1998; **97**(18):1761–1762.
2. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, in this issue.
3. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham study. *American Journal of Cardiology* 1976; **38**(1):46–51.
4. Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, Wilson PW. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *Journal of American Medical Association* 2003; **290**(7):891–897.
5. Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *The New England Journal of Medicine* 2004; **350**(14):1387–1397.
6. Ware JH. The limitations of risk factors as prognostic tools. *The New England Journal of Medicine* 2006; **355**(25):2615–2617.
7. O'Malley PG, Feuerstein IM, Taylor AJ. Impact of electron beam tomography, with or without case management, on motivation, behavioral change, and cardiovascular risk profile: a randomized controlled trial. *Journal of American Medical Association* 2003; **289**(17):2215–2223.