New statistical methods for the evaluation of cardiovascular risk markers: what the clinician should know

Helge RØJSJØ* and Torbjørn OMLAND*†
*Division of Medicine, Akershus University Hospital, NO-1478 Lørenskog, Norway, and †Faculty Division, Akershus University Hospital, University of Oslo, Oslo, Norway

ABSTRACT

Calculation of odds ratios or hazard ratios by multivariate logistic regression or Cox regression analyses have traditionally been used to show that candidate risk markers provide prognostic information independently of conventional risk markers, but it has become increasingly clear that a statistically significant increase in risk in multivariate models may not represent a clinically meaningful improvement in overall prediction. This observation has prompted the development of more clinically relevant statistical methods, including tests of the ability to provide incremental discrimination compared with traditional risk markers by calculation of the C-statistic (corresponding to the area under the receiver operating curve) of the model and methods for evaluating improvement in risk classification by the use of event-specific re-classification tables or the use of the integrated discrimination improvement. In the present issue of Clinical Science, Khan and co-workers have evaluated the prognostic value of the GRACE (Global Registry of Acute Coronary Events) risk score, the cardiac biomarker NT-proBNP (N-terminal pro-B-type natriuretic peptide), and their combination, for the prediction of mortality after acute myocardial infarction, using a combination of statistical methods.

In the present issue of Clinical Science, Khan et al. [1] have compared the prognostic value of the GRACE (Global Registry of Acute Coronary Events) risk score, the cardiac biomarker NT-proBNP (N-terminal pro-B-type natriuretic peptide), and their combination, for the prediction of mortality at 30 days and 6 months in 1033 consecutive patients admitted to a single centre with acute myocardial infarction. The study [1] is of major interest not only because it compares the utility of a previously developed and widely used risk score [2] with that of what is probably the single strongest cardiovascular biomarker for prediction of death in patients with acute coronary syndromes [3,4], but also because the investigators have used multiple new and refined statistical methods for evaluating prognostic utility. These methods provide clinically relevant information beyond merely demonstrating that variables of interest are independently associated with outcome in multivariable Cox or logistic regression models.

Traditionally, logistic regression or Cox regression models have been used to demonstrate that a novel risk marker provides prognostic information independently of conventional risk markers. However, it has become increasingly clear that a statistically significant increase in risk in a multivariate model may not signify a clinically relevant change in overall prediction [5], prompting the

Key words: acute coronary syndrome, Global Registry of Acute Coronary Events (GRACE) score, myocardial infarction, N-terminal pro-B-type natriuretic peptide (NT-proBNP), prognosis.

Abbreviations: GRACE, Global Registry of Acute Coronary Events; IDI, integrated discrimination improvement; NRI, net reclassification improvement; NT-proBNP, N-terminal pro-B-type natriuretic peptide; ROC-AUC, area under the receiver operating curve.

Correspondence: Professor Torbjørn Omland (email torbjorn.omland@medisin.uio.no).
Commentary

existing categories of risk. This works fine in primary relevant parameter, it is predicated on knowledge of risk classification. Although the NRI may be a clinically (sensitivity) against the false positive rate (1—specificity), and indicates the probability across the spectrum of cut-off values that a case will have a higher value on the test than an individual not experiencing the outcome of interest. Ideally, all cases have higher values than non-cases, yielding an ROC-AUC of 1.0, whereas chance alone will result in an ROC-AUC of 0.50. For instance, a test with an ROC-AUC of 0.80 means that a subject with the disease (or outcome) of interest will have an 80% probability of having a higher value of the diagnostic or prognostic test than a subject without disease (or the outcome of interest). Likewise, the discriminatory value of a multivariable model can be assessed by calculating the C-statistic or C-index (a measure corresponding to the ROC-AUC) of multivariable Cox or logistic regression models.

Other investigators in the field have, however, argued that the ROC-AUC or C-index are imperfect for evaluating the clinical usefulness of a risk marker and thus, should not be the sole determinant of clinical utility. Moreover, these investigators have suggested that more emphasis should be put on the re-classification of individuals to their true risk group, a procedure commonly referred to as (re-)calibration of a risk model [6]. Recently, new statistical methods for prognostic assessment, including evaluating improvement in risk classification by the use of event-specific re-classification tables and the use of the IDI (integrated discrimination improvement), have therefore been proposed as alternative methods to ROC-AUC and C-index determination [7]. The IDI assesses improvement in risk discrimination, i.e. how well a model separates subjects with disease compared with those without disease, by estimating the change in the difference in the mean predicted probabilities of the outcome between those with and without the outcome in question after introducing the candidate biomarker to the model. A better model is reflected by a greater difference than observed in a poor prognostic model.

Alternatively, improvement in risk classification can be evaluated by the use of event-specific re-classification tables. Predicted probabilities for the outcome in question can be determined for individual patients and re-determined after adding the candidate biomarker to the model. The so-called NRI (net reclassification improvement) is calculated by assessing the net improvement in risk classification. Although the NRI may be a clinically relevant parameter, it is predicated on knowledge of existing categories of risk. This works fine in primary prevention, where well-accepted categories of risk exist, for instance as defined by the Framingham Risk score. Moreover, in the primary prevention setting a change between categories may have direct therapeutic implications, including initiation of preventive medication and assignment of treatment goals. The clinical implications are less clear in the setting of secondary prevention, where such a priori risk categories may not always exist, and the rationale for individualizing treatment based on estimated risk has not been shown as convincingly.

In their study, Khan et al. [1] not only demonstrate that the GRACE risk score and NT-proBNP provide statistically significant independent prognostic information, but use C-statistic calculations and NRI to assess the clinical utility of these risk markers. The GRACE risk score and NT-proBNP combined provided improved discrimination for mortality at 6 months, expressed as a significant increment in the C-statistic. Finally, the authors [1] demonstrated that classifying patients as low, intermediate or high risk based on a model that included both the GRACE risk score and NT-proBNP fitted better with the observed short- and long-term event rate than a model based on GRACE risk score alone. A potential problem with the NRI is that it does not weight improvement, a problem that is reflected by the re-classification data for 30-day mortality: a net positive re-classification by adding NT-proBNP to the GRACE model was driven by patients in the non-event group being moved down to a lower risk stratum. Evaluating patients experiencing events specifically, adding NT-proBNP to GRACE in fact provided inferior classification as compared with the use of the GRACE risk score alone. The clinical usefulness of adding NT-proBNP to the GRACE score for identifying subjects at high risk of early events thus appears limited. In contrast, the results of adding NT-proBNP to the GRACE risk score for predicting long-term mortality may have greater practical utility. The NRI again indicated an improved classification by the model combining NT-proBNP and GRACE than for GRACE alone, but, unlike the results for short-term mortality, this was largely driven by NT-proBNP improving risk stratification among patients who actually experienced an event during follow-up. Accordingly, as a ‘rule-in’ test, the clinical value of combining NT-proBNP and the GRACE risk score may be greater for long-term than short-term mortality.

Evaluation of the clinical utility of a cardiovascular risk marker is challenging. Although the use of the statistical methods described above represent obvious improvements in comparison with relying exclusively on odds ratios or hazard ratios obtained from logistic or Cox regression models, individually these new tests provide only part of the answer to the question of what represents a clinically useful test. To make a judgment concerning the prognostic value based on a battery of tests providing complementary information, i.e. the magnitude of the
relative and absolute risk, the discriminatory ability and its ability to reclassify patients, is probably the preferable approach.

REFERENCES


