Determining Decision Limits for New Biomarkers: Clinical and Statistical Considerations

David A. Morrow* and Nancy R. Cook

The pathway from discovery to clinical adoption of new biomarkers is marked by a series of significant benchmarks that relatively few cardiovascular biomarkers have successfully navigated (1). For those biomarkers that do emerge as viable candidates for clinical use, determining appropriate decision limits takes on substantial importance. In our experience, this task is demanding and is regularly executed poorly. In this issue of Clinical Chemistry, Tang and colleagues assess the prognostic performance of myeloperoxidase (MPO) as a biomarker for stable ischemic heart disease and perform internal validation of a cutpoint (2). This example provides an opportunity to address the clinical and statistical issues that affect the determination of clinical decision limits for new biomarkers.

Clinical Considerations

Ambiguity concerning decision limits frustrates clinicians and negatively affects the clinical adoption of a biomarker. At the time of clinical introduction, a new biomarker ideally should have well-characterized decision limits that (a) are pragmatic to apply, (b) have undergone validation in multiple studies, (c) have been evaluated in the relevant population(s) and application(s), and (d) have achieved synergy between available scientific data and regulatory labeling.

Relevance of Pragmatism

Although it is not always possible in clinical medicine, practitioners crave certainty in interpreting test results and much prefer to think in dichotomous terms. Therefore, a tension sometimes exists between the desire to keep cutpoints simple (so that they are easily remembered and convenient to apply) and the limitations imposed by the complexity of best capturing the diagnostic, prognostic, or therapeutic information offered by the biomarker. This tension can create controversy in the modeling of a new biomarker as reasonable, often opposing, arguments arise from the clinical and statistical perspectives. We believe that both points of view should be taken into consideration in the evaluation of a new biomarker; however, we assert that a strong motivation exists in clinical medicine for the development of dichotomous or categorical cutpoints to facilitate decision-making.

In some cases, the biology of a biomarker is such that a clear threshold exists in the relationship between its concentration and a diagnosis or outcome. In this situation, a dichotomous cutpoint is preferable from all perspectives, although the absolute value may confer some additional information. For example, because cardiac troponin was not detectable in the blood of healthy individuals until recently, any detectable increase in concentration was indicative of myocardial injury, diagnostic of myocardial infarction in the appropriate setting, and associated with an adverse prognosis (3). More often, however, when a predictor variable is continuous, risk increases throughout the range of the predictor concentrations (e.g., cholesterol). Nevertheless, despite the near-linear relationship, clinical guidelines that use cholesterol are based on specific thresholds rather than on a continuous model (4). Discrete decision limits are needed for clinicians to remember them, to develop guidelines that are possible to transmit and implement, and to create benchmarks for quality of care. Therefore, as for cholesterol, clinicians have embraced the convenience of defined cutpoints for most cardiovascular biomarkers to support decision-making (5) and have sometimes used categorical approaches (e.g., low risk or “rule-out,” intermediate or gray zone, high risk or “rule-in”) to partially account for the graded nature of the prognostic and diagnostic relationships (6).

Clinical Requirements for Validation

The aim of clinical pragmatism does not supersede the need for decision limits that adequately reflect the information offered by the new biomarker and that have undergone sufficient exploration and validation. Initial reports typically estimate the risk relationship between a biomarker and outcomes with cutpoints that have been optimized for the data set in which they were de-
Commonly, the goal is to determine a decision limit (or cutoff) for a biomarker that provides accurate predictions of disease state or clinical outcomes (9). The cutoff should be established so that the risk of clinical events for patients above and below the cutoff are significantly different (10). The ideal cutoff for a biomarker needs to be based on the shape of the relationship of a predictor with the disease. If the shape exhibits a threshold, then that point is a natural choice for a cutoff. If the relationship is monotonic, however, choosing a single cutoff can be arbitrary, and from a statistical perspective it is preferable in such cases to use a continuous measure, either alone or in a comprehensive model, to estimate risk. Risk strata for clinical use can then be established from the predicted risk. Alternatively, thresholds can be determined at the level of the individual patient via net-benefit curves, which display the benefit of a particular treatment strategy across a range of thresholds (9).

Robust validation of decision limits for new biomarkers typically requires (a) examination in separate studies, (b) a sample size sufficient to reliably discriminate differences in performance between proposed cutoffs, (c) relevance of the studied population to the clinical application, and (d) evaluation for the specific intended indications. These requirements are intensive with respect to resources. Consequently, this work frequently is completed after the initial clinical introduction. It is not unusual that pertinent influences of patient characteristics on the performance of accepted cutoffs are recognized only after commercial availability of a biomarker test, as exemplified by studies of natriuretic peptides in women and individuals with renal dysfunction (5) and studies of C-reactive protein across race and ethnicity (7). In addition, one cannot assume that a single cutoff will be useful across a variety of applications (e.g., for both diagnosis and risk stratification) and conditions (e.g., stable and unstable ischemic heart disease). Moreover, the appropriateness of these cutoffs for selecting specific therapies is a separate question. Because the risk relationships may vary, it is necessary to study the biomarker for each of the intended applications and populations to completely assess the proposed decision limits.

Statistical Considerations

In their study (2), Tang et al. sought to select the best single cutoff value for MPO. The authors chose a cutoff that maximized the area under the ROC curve (AUC). When there is a single dichotomous predictor, this exercise is the same as maximizing the Youden index (sensitivity + specificity − 1), which is considered “optimal” in the sense that it minimizes the overall misclassification rate (8). What is optimal in any situation, however, depends on other factors, such as the relative costs of false-negative and false-positive classifications. The best cutoff to use for treatment decisions should take into account both the prevalence of the disease and the relative importance of sensitivity vs specificity, which can be assessed through alternative-weighting studies (8).

Other possibilities for choosing cutoffs can be based on the shape of the relationship of a predictor with the disease. If the shape exhibits a threshold, then...
ods, more potential for overfitting than the usual Cox modeling because it incorporates multiway interactions that fit the observed data much more closely. Therefore, the use of this single cutoff, determined from a single data set without adequate internal validation, must be regarded as needing additional investigation.

Summary and Recommendations

The availability of well-validated decision limits is vital to optimal integration of a new biomarker into clinical practice. Approaches to internal validation and data-mining methods, such as those used by Tang et al. (2), lead to overfitting and overestimation of risk relationships and are generally not sufficient for selecting final clinical cutoffs. Such methods, when applied correctly, can be reasonable for suggesting cutoffs for external validation. Biomarkers that have monotonic linear relationships with risk are best handled as continuous variables when incorporated into comprehensive risk models. As consistently demonstrated in clinical practice and professional society guidelines, however, practitioners will almost always seek thresholds to provide structure for clinical decision-making, such as those existing for cholesterol. Therefore, such cutoffs warrant development and validation. Although the approach is demanding, we recommend assessment of clinical decision limits by external validation in 2 or more data sets that are appropriate to each of the proposed clinical application(s), with attention paid to the possibility of differences in risk relationships in clinically relevant subpopulations.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: None declared.

Consultant or Advisory Role: D.A. Morrow, Beckman Coulter, Instrumentation Laboratories, Ortho Clinical Diagnostics, Roche Diagnostics, and Siemens.

Stock Ownership: None declared.

Honoraria: None declared.

Research Funding: D.A. Morrow, Beckman Coulter, BG Medicine, Ortho Clinical Diagnostics, Roche Diagnostics, Siemens, and Singulex.

Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

7. Albert MA, Glynn RJ, Buring J, Ridker PM. C-reactive protein levels among women of various ethnic groups living in the United States (from the Women’s Health Study). Am J Cardiol 2004;93:1238–42.