EDITORIAL

On the Quantitative Assessment of Predictive Biomarkers

Jason P. Fine, Michael J. Pencina

Affiliations of authors: Department of Biostatistics, University of North Carolina Chapel Hill, Chapel Hill, NC (JPF); Department of Biostatistics and Bioinformatics, Duke University, Durham, NC (MP).

Correspondence to: Jason P. Fine, ScD, Department of Biostatistics, University of North Carolina Chapel Hill, Campus Box 7420, McGavran Greenberg Hall, Pittsburgh st, Chapel Hill, NC (e-mail: jfine@bios.unc.edu).

Predictive biomarkers are among the most direct steps towards converting precision medicine into a clinical reality. By their very definition they are fundamentally different from prognostic or diagnostic biomarkers—instead of helping identify individuals who will develop the disease, they focus on differentiating between those who will and will not “benefit” from treatment. Given this difference, a natural question is whether the performance metrics of sensitivity, specificity, positive predictive values (PPVs), and negative predictive values (NPVs) (“four metrics”) used in the prognostic/diagnostic context should be extended to the predictive setting. The two brief communications published in this issue of the Journal approach the question from two different perspectives.

Using methods and assumptions similar to those adopted by Zhang et al. [1], Simon [2] presents simple formulas for estimating the four metrics, where the goal is predicting “benefit” at the individual level. That is, the interest is identifying individuals who would have improved outcomes with treatment but not without treatment and conversely individuals who would not have improved outcomes with treatment but not without treatment. Such metrics provide a clear understanding of the impact of treatment at the individual level. However, they suffer from inherent limitations owing to their reliance on unverifiable assumptions, arising from the fact that only one of two potential outcomes (one with treatment, one without treatment) can be observed on an individual. Simon argues that such analyses may still be useful without such assumptions if one employs a second “pragmatic interpretation.” Specifically, Simon shows that they can be interpreted as a comparison based on two randomly selected individuals, one treated and one control/standard of care with the same measured covariates. This interpretation circumvents the main criticism about estimating these quantities raised in the other communication by Janes et al. [3]. This second interpretation may be of interest in evaluating the clinical utility of a biomarker at the population level and may in principle be estimated without any assumptions.

Given the substantive concerns in Janes et al., the question arises as to whether or not there is a role for the “probability of treatment benefit” in assessing the predictive value of biomarkers in treatment decision-making. As an individual-level quantity, it has a deeper causal interpretation than the pragmatic population-level interpretation of Simon. However, as noted in Janes et al., restrictive assumptions are needed. This tradeoff is common in causal inference, even in randomized clinical trials when treatment efficacy at the individual (as opposed to population) level is desired. For those interested in such endpoints, either in studying treatment efficacy or in assessing the predictive value of a biomarker in treatment selection, a conservative approach is to employ a sensitivity analysis. This approach explores the extent to which the estimated individual-level benefit varies as the assumptions underlying the analysis are varied. If the results are consistent across a wide range of scenarios, then such individual-level information might be used to develop biomarker policies at the population level, although alternative clinically useful quantities, as advocated in Janes et al., might also be appropriate.

Issues are raised in Janes et al. as to whether the four metrics described in Simon (sensitivity, specificity, PPV, NPV) are ideal in the predictive setting when employing the “pragmatic” interpretation. Such interpretation applies to two independent subjects and hence may not be an average of individual-level effects, as occurs in the usual analysis of treatment efficacy in a randomized clinical trial. An alternative approach presented by Janes et al. focuses on the concept of treatment effect defined as “given my biomarker value, what are the differences in the probabilities I will experience a bad outcome with treatment and without treatment.” In a randomized trial, this quantity can be estimated without assumption and provides a clear rationale for the development of treatment rules. While this quantity might be less desirable than “the probability of treatment benefit,” it is easily understood as an average of individual-level effects, that is, the differences in the individual probabilities of bad outcomes on the two treatments at a given level of the biomarker. Janes et al. argue that simplicity of interpretation and lack of sensitivity to unverifiable assumptions make it attractive for the assessment of predictive biomarkers. The authors suggest these average probability differences in bad outcomes,
and not “the probability of treatment benefit”, are of greater utility in cost benefit analyses.

An important point that is not directly addressed in either Simon or Janes et al. is that the evaluation of prognostic/diagnostic and predictive biomarkers should occur in the context of what other predictors might be available to assess the likelihood of response with and without treatment. When other markers or standard factors are already in use or easy to collect, the value of the “new” marker should be considered on an incremental basis. What is the added gain over existing markers? Can the methods be further developed along these lines?

Another important issue is that both articles focus on predicting treatment response as a way to improve the probability of a single efficacy outcome. However, as noted by Janes et al., many treatments, especially in oncology, have numerous side effects. Thus, it would be worthwhile to evaluate novel markers in at least a two-dimensional setting—to what extent it helps quantify the probability of an efficacious outcome with and without treatment but also to what extent it can help differentiate between the risks of side effects associated with treatment or standard of care. One might imagine tradeoffs, where efficacy and toxicity effects associated with a novel marker are in opposing directions. Such multidimensionality adds an additional complication to an already complex problem, but it is hard to ignore given the focus on selecting optimal therapies. The adaptation of the metrics discussed in Simon and Janes et al. to this setting requires careful consideration.

Note
The authors have no conflicts of interest to declare.

References