Identification of Clinically Useful Cancer Prognostic Factors: What Are We Missing?

Lisa M. McShane, Douglas G. Altman, Willi Sauerbrei

The number of cancer prognostic markers that have been validated as clinically useful is pitifully small, despite decades of effort and money invested in marker research (1–3). For nearly all markers, the product has been a collection of studies that are difficult to interpret because of inconsistencies in conclusions or a lack of comparability. Small, underpowered studies; poor study design; varying and sometimes inappropriate statistical analyses; and differences in assay methods or endpoint definitions are but a few of the explanations that have been offered for this disappointing state of affairs (4–11). Researchers attempting to conduct meta-analyses of prognostic marker studies encounter many difficulties (12–14). In this issue of the Journal, a meta-analysis by Kyzas et al. (15) of the tumor suppressor protein TP53 as a prognostic factor in head and neck cancer provides compelling empirical evidence that selective reporting biases are a major impediment to conducting meaningful meta-analyses of prognostic marker studies. These biases have serious implications, not only for meta-analyses but also for interpretation of the cancer prognostic literature as a whole.

Kyzas et al. demonstrate bias in the estimated association between TP53 status and mortality relating to a number of factors. The prognostic importance of TP53 decreased as the pool of included studies increased from published and indexed studies to any published studies, and then to the full set of all published and unpublished studies with retrievable data. In addition, when marker positivity definition and choice of clinical endpoint were retrospectively partly standardized across studies, there was a decrease in the overall estimated association compared with that obtained when each study used its preferred marker positivity criterion and endpoint.

The authors are appropriately cautious in acknowledging that confidence intervals for the estimates across the different analysis sets do overlap and that one cannot determine what might have been the effect of additional nonretrievable data on the observed trends. Nonetheless, we believe that this study provides the most compelling evidence yet that the published prognostic
literature is a serious distortion of the truth. Even the most trusting and optimistic readers must feel pangs of suspicion when comparing, for example, the highly asymmetric scatterplot of study-specific association estimates observed for the published indexed studies [figure 3, A in Kyzas et al. (15)] to the nearly symmetric scatter of estimates resulting when unpublished studies are included [figure 3, C in Kyzas et al. (15)]. There is a clear message that studies showing stronger associations were more likely to be published in indexed journals, and there appeared to be a preference for reporting endpoints that produced the “best” results. Careful examination of the causes and discussion of potential remedies for these problems is warranted.

Most prognostic marker studies are conducted on retrospective collections of specimens. Archived specimen collections with good clinical and pathologic annotation and reliably collected clinical follow-up information are rare; many existing collections are not representative of well-defined populations and accompanying data may not be reliable. Frequently, study sample sizes are driven more by the number of specimens available than by any rigorous design criteria or statistical power considerations. With specimens in hand, a typical marker study is much easier to conduct than a randomized clinical trial. Often, institutional review requirements are minimal and formal protocols are rare; data collection may be informal and not stringently quality controlled. If a small marker study produces uninterpretable or equivocal results, an investigator may choose to abandon it and never publish the results. When a marker effect is found to be non-statistically significant in a small study, investigators, reviewers, and editors may choose to dismiss the study as underpowered. However, if the same study had produced a highly statistically significant result, it may be labeled as promising and be published. In contrast, a randomized clinical trial that had undergone institutional review, and for which considerable time and money had been invested for the study design, monitoring, and analysis, would be far more likely to be published, regardless of the findings. Yet, there is consistent evidence even for randomized clinical trials of selective reporting in relation to statistical significance (16,17).

Compounding the problem of selective publication is the tendency for many endpoints to be examined within the same prognostic marker study and, if assay values are dichotomized, for many cutpoints to be considered (18,19). The impact of this “overanalysis” is a further bias resulting in inflated estimates of association and increased chances of spurious findings. If marker studies were conducted according to prespecified protocols with clear designation of primary endpoints and assay scoring methods, these problems would be largely avoided. As in clinical trials, central parts of the analysis strategy should ideally be predefined. Additional explanatory analyses are highly welcome, but the character of the analyses should be stated in the paper and should influence the interpretation. We expect that few of the studies reviewed by Kyzas et al. were conducted according to a predefined protocol or had study sample sizes chosen to provide adequate power to detect a specified marker effect size. Most of the studies included fewer than 100 patients, a number that is obviously too small for a reasonable multivariable analysis. More complete and transparent reporting of marker studies would make it easier to distinguish carefully designed and analyzed studies from haphazardly designed and over-analyzed studies. The forthcoming REMARK guidelines (20) are an attempt to improve the reporting of tumor maker prognostic studies, but substantial improvement of the primary studies is a prerequisite.

The availability of individual patient data has been cited (14) as an important advantage for conducting meta-analyses of prognostic marker studies because it may permit reanalyses of data to accomplish standardization of marker values and endpoint definitions or to achieve comparability of patient populations. This approach leads to more interpretable overall estimates of association. It is shameful that, when primary investigators of 64 studies were contacted by Kyzas et al. in attempts to retrieve raw data, those for 22 studies were unable or unwilling to make their data available. Many journals, including the Journal of the National Cancer Institute, require that genomic data such as gene expression microarray data be made available as a condition of publication. Why should data availability be important only for high-dimensional marker data such as microarray data? The message is clear that, if data are not made available at the time of study completion and publication, it is unlikely that they will be retrievable later. Moreover, careful documentation should be viewed as essential for not only data variable definitions but, more generally, also for all aspects of the design, conduct, and analysis of the study. If the data are not properly documented at the time of collection, they may be uninterpretable several years later. Individual patient data from poorly designed, conducted, or documented studies will be of little value.

The tumor marker research community must come to the same realization that clinical trialists came to decades ago. If sound scientific principles of careful study design, adequate study size, scrupulous data collection and documentation, and appropriate analysis strategies are not adhered to, the field will flounder. Culture changes will be required. Stable and adequate funding will be required to have necessary personnel and infrastructure to collect, annotate, and maintain valuable specimen collections essential for high-quality retrospective studies. More importantly, the necessity of large, definitive prospective studies or prospectively planned meta-analyses for tumor marker research must be recognized.

References

(3) Schilsky RL, Taube SE. Introduction: tumor markers as clinical cancer tests—are we there yet? Semin Oncol 2002;29:211–2.


