The Process to Discover and Develop Biomarkers for Cancer: A Work in Progress

David F. Ransohoff

The process to discover and develop molecular biomarkers for cancer diagnosis (or prognosis) is a work in progress and is evolving. "Discovery" research—searching for markers by use of high-throughput technology without an a priori hypothesis—has produced few useful biomarkers over the last 10 years despite numerous claims in research publications and news reports about "95% sensitivity and 95% specificity" or better. The degree of disconnect between claims and products should make us ask why progress is so slow: Is it the normal stop-and-start of science? Or is there some systemic problem with the process that we currently use to discover and develop markers? After all, it took decades to evolve the process to discover and develop drug therapies. So, where does the evolution of the discovery and development of biomarkers stand and how can we improve the process?

Although the process is complicated, there are two major questions to address in research. First, can a marker discriminate well? That is, can it discriminate cleanly and reliably (eg, reproducibly and not due to artifact or bias) between persons with early-stage cancer vs those without cancer (or good vs bad prognosis)? Second, does that discrimination, when coupled with an intervention such as surgery or chemotherapy, lead to improved outcome? Answering the second question requires a randomized clinical trial, a design that is powerful and well understood but also expensive and time consuming. A randomized clinical trial is not even considered until a marker clearly shows reliable discrimination that warrants such an effort. Failures in current marker research occur not at the randomized clinical trial stage but earlier, when results that appear to have promising discrimination in early discovery turn out not even to be reproducible.

Pepe et al. (1) proposed a formal structure in 2001 "to guide the process of biomarker development" consisting of five "phases [that] are generally ordered according to the strength of evidence that each provides in favor of the biomarker, from weakest to strongest [and] the results of earlier phases are generally necessary to design later phases." The phase structure has been widely adopted for use in various research projects, including, as noted by Feng et al. (2), "by the EDRN [the National Cancer Institute's (NCI) Early Detection Research Network], a number of Specialized Program of Research Excellence (SPORE) consortiums, and other biomarker research projects."

With this background, the article in this issue of the Journal (3) from key authors of the phase proposal provides insight about the evolution of this process and about how to improve it. The article proposes a specific study design for the step just before a randomized clinical trial is done to demonstrate reliable discrimination in the kinds of persons who might then be studied in a randomized clinical trial. Reliability is achieved in design by prospective-specimen collection (by which the authors mean before the diagnosis is known) and

retrospective-blinded-evaluation (PRoBE) (3). The goal of the design is to avoid biases that compromise so much biomarker research: "Common biases that pervade the biomarker research literature would be eliminated if these rigorous standards were followed" (3).

The fact that a proposal is being made now about study design provides lessons about the state of evolution of marker research. Although using a design intended to avoid bias may seem natural and even taken for granted, it may surprise laboratory investigators unfamiliar with clinical research design that avoiding bias is not achieved at all by using phases. Phases provide a list of research questions and a rough idea of the kinds of subjects to study (4); they do not specify any details of research design that protect against the biases (4) that, as stated by Pepe et al. (3) "pervade the biomarker research literature." In the 6000-word article about phases (1), the word bias appears once; in the proposal for PRoBE (3), bias is the main focus. Design (and avoiding bias) and phases are simply different topics in the larger process of marker discovery and development.

The PRoBE design is already being used in a few settings and for exactly the reasons proposed. In a study of ovarian cancer diagnosis, serum specimens collected before diagnosis was known are being analyzed retrospectively in NCI's screening study of Prostate, Lung, Colon, and Ovarian cancer (PLCO) (5). In a study of breast cancer prognosis, specimens collected before prognosis was known were analyzed retrospectively in a study that used tissue collected by NCI's National Surgical Adjuvant Breast and Bowel Project clinical trial B-14 (6). In a study of colon adenoma diagnosis, specimens collected before diagnosis was known were analyzed retrospectively (7).

By putting research design and bias squarely on the table, the PRoBE proposal invites a larger discussion about other ways to improve the process of discovery and development of biomarkers for cancer. In particular, the PRoBE proposal focuses on what may be termed late-stage research, a pivotal evaluation after discovery shows discrimination promising enough to warrant a PRoBE study that, itself, may be expensive and time consuming. However, the main problem in current marker research is not at the randomized clinical trial stage but rather is weak discovery, when promising results are not even reproducible and so do not warrant a pivotal PRoBE study. Indeed, a PRoBE study of a weak marker will simply

Affiliation of author: Departments of Medicine and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Correspondence to: David F. Ransohoff, MD, Departments of Medicine and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC (e-mail: ransohof@med.unc.edu).

confirm that discovery did not work. How, then, can we improve the process of discovery and avoid inflated claims and disappointment? Although the topic is not developed, the PRoBE authors suggest using "key elements of the PRoBE design" in earlier phases (3). We can expect more discussion about this topic in the future.

Last, the PRoBE proposal challenges, if indirectly, the current conceptualization of phases. Although some kind of order is needed in marker discovery and development, is our current conceptualization of phases the best one or will it too evolve? Because the current five phases list more than 20 questions and require more than 20 research studies (1,4), an investigator may fairly ask, Do I need to do them all? In order? If not, where do I start? The PRoBE authors suggest that it may be appropriate to skip phases and even to do discovery and validation with the same set of specimens if there are enough specimen samples for independent evaluation (3). We can expect more discussion about these topics, too (4).

The PRoBE proposal (3) contributes to a useful discussion of the larger process by which we discover and develop biomarkers (4). The methods to evaluate any diagnostic test—molecular or other— have long been underdeveloped (8), compared with methods to evaluate drugs, and they will evolve in parallel with new knowledge about biology and technology. If we can understand and manage these evolutions well, then perhaps we can more efficiently discover and develop biomarkers useful for cancer diagnosis and prognosis.

References

1. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst.* 2001;93(14):1054–1061.
2. Feng Z, Prentice R, Srivastava S. Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. *Pharmacogenomics.* 2004;5(6):709–719.
3. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst.* 2008;100(20):1432–1438.
4. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. *J Clin Epidemiol.* 2007;60(12):1205–1219.
5. Winstead ER. Ovarian cancer study could speed early detection. *NCI Cancer Bulletin.* June 10, 2008;5(12):5–7.
6. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004; 351(27):2817–2826.
7. Ransohoff DF, Martin C, Wiggins WS, Hitt BA, Keku TO, Galanko JA, et al. Assessment of serum proteomics to detect large colon adenomas. *Cancer Epidemiol Biomarkers Prev.* 2008;17(8):2188–2193.
8. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. In: Knottnerus JA, ed. *The Evidence Base of Clinical Diagnosis*. London: BMJ Books; 2002:39–59.