Patient Safety and Clinical Effectiveness as Imperatives for Achieving Harmonization Inside and Outside the Clinical Laboratory

Ronald W. McLawhon1*

For clinical laboratory, the case for standardization and harmonization has been evident for more than 4 decades, since Radin first proposed using traceable reference standards for calibration as a means to harmonize laboratory results produced with different in vitro diagnostic methods (1). Over the intervening period, the subject has been reviewed (2) and debated extensively—both inside and outside the laboratory community—yet a striking majority of our physician and surgeon colleagues still fail to grasp or understand the limitations of current laboratory measurements, the lack of interchangeability of results obtained by different analytical methods, and the resulting effects on interpretation, clinical decision-making, and patient management.

Some incremental progress has been made in addressing these issues through professional organizations, multidisciplinary practice guidelines developed by national and international committees, peer-review publications, and the global in vitro diagnostic manufacturing industry. Notable successes include the National Glycohemoglobin Standardization Program (including the recommendations of the Diabetes Control and Complications Trial), the National Cholesterol Education Program, and the National Kidney Disease Education Program, which have helped drive improvements in laboratory methods for hemoglobin $A_1c$, total cholesterol, and creatinine, respectively, and helped establish clinical practice guidelines based on laboratory measurements that meet defined performance criteria. As recently as October 2010, the AACC hosted an international conference for a diverse group of stakeholders to improve harmonization of laboratory results and to make recommendations for the future.

Regrettably, personnel in the clinical laboratory must continue to cope on a day-to-day basis with the blissful ignorance or blatant denial of these multifac-
rent tumor marker assays, the different molecular variants and isoforms detected by these assays, the lack of interchangeability of results obtained with the different methods used to monitor patients with known disease, and the need to reestablish new baselines when making conversions. These and other identified limitations confound test interpretation and impede the ability to create specific, defined, and uniformly accepted cutoff values for clinical decision-making.

In this issue of Clinical Chemistry, Stephan et al. present findings (7) from a study of a large cohort of patients that compared the effects of different prostate-specific antigen (PSA) methods (for total PSA and the percentage of free PSA), as well as calibration changes, on the results of logistic regression–based nomograms for predicting the risk of prostate cancer (PCa). Multivariate models, including artificial neural networks and logistic regression–based nomograms, have been proposed since the late 1990s (8, 9) as an adjunct for PCa risk prediction, for monitoring disease progression, and for guiding therapeutic intervention, to help address the low specificity of measurements of total and the percentage of free PSA alone for diagnostic and prognostic assessments. Recently, such combined approaches, which can account for other factors such as age, digital rectal examination findings, and prostate volume, have been advocated widely (10) and embraced by the medical community; however, an important assumption of these approaches has been that variation in the PSA measurement methodology used does not affect assessment outcomes. The NACB guidelines, in discussing the promise of nomograms for predicting PCa risk, suggested that these tools may be “the most accurate means of individualizing therapy and predicting outcome, and reflect the most recent advances in patient management” (6). The expert panel cautioned, however, that it might be difficult to select the best nomogram when several competing versions apply to the same clinical decision.

A critical and comprehensive comparison of the effects of assay-dependent variation in measurements of total PSA and the percentage of free PSA on commonly used PCa risk-prediction nomograms has not previously been well documented. To clinical chemists and pathologists familiar with the issue of the lack of PSA assay harmonization, it may seem intuitively obvious that some differences in the performance and outcomes of these nomograms could be observed with different assays and calibration methods. Despite the advances in the availability of WHO reference materials and assay improvements from all manufacturers, we are still far from achieving the desired harmonization and interchangeability of PSA results across all available methods (11, 12). Stephan and his colleagues, who were early proponents of the value of PCa nomograms, have documented the obvious (and not so obvious) limitations of these nomograms, the use of which may alter clinical decisions, depending on the PSA methods used. In this large retrospective study of nearly 800 patients (which included some assays that are no longer commercially available), Stephan et al. have documented very important PSA test method–dependent differences between 5 commonly used assays when results are applied to 5 of the most widely used and readily available (“plug and play”) regression-based predictive nomogram models. Not only did results of the 5 assays lead to different PCa probabilities with the same nomogram, but the various nomograms (which differ in their inclusion of other factors, such as the percentage of free PSA, sampling density, and prostate volume) also produced different PCa probabilities when the same PSA assay was used. Although the authors’ ROC curve analyses yielded comparable areas under the curves, there were significant differences between the 5 assays in the diagnostic sensitivities and specificities at various PCa probability cutoff values for the majority of the evaluated nomograms.

Although this study may not have examined the effects on all currently used nomograms, such as those described by Finne et al. (13), and did not take into account other PSA assays, its findings raise some well-founded concerns that merit the attention of both clinicians and the laboratory community. Yes, the conclusions of Stephan et al. may state the obvious—namely, the accuracy of predicted PCa probabilities produced with different nomograms is affected and can be compromised by the lack of harmonization of PSA assays. Moreover, the variations in PCa risk prediction and discrimination can be substantial and unacceptable, depending on the model used and differences in calibration, even with the same assay. Nonetheless, this study reinforces the point that until harmonization of results is truly achieved in the laboratory, caution is still warranted. Few tools outside the laboratory can substitute for harmonization for accurately informing and appropriately guiding clinical decision-making and ensuring safe and effective patient care.

In the US, PCa remains the most common cancer in men and the second-leading cause of cancer deaths. Although the incidence of PCa and PCa deaths has declined steadily over the last 2 decades because of early detection and treatment (6), we must remain vigilant in raising awareness that current laboratory results are neither standardized nor harmonized and may lead to erroneous decisions with serious consequences (i.e., just plugging the numbers into a risk-calculation nomogram can give misleading results). The decade-old Institute of Medicine report (14) on building a safer health system states that “to err is human,” but no-
where in the report does it say that “to forgive is divine,” especially now that we have knowledge of factors that can lead to the wrong call and potential harm.

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: No authors declared any potential conflicts of interest.

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