Issues in Immunoassay Standardization: The ARCHITECT Folate Model for Intermethod Harmonization

Extensive variability among folate assay methods has been well described. A FLAIR round-robin study of serum and whole-blood folate results across 11 laboratories in seven European countries reported overall CVs of 18–41% (1). A CDC round-robin survey of 20 laboratories reported two- to ninefold differences in folate values among methods and laboratories (2). With increasing awareness of the nutritional importance of folate, both studies concluded that there was an acute need for improved standardization of folate testing. Recently, a comparison of five automated serum and whole-blood folate assays was reported (3). In that study, the methods were compared with the Bio-Rad Quantaphase II radioassay. Between the methods, differences as high as 40% for serum folate and 250% for whole-blood folate were noted. The authors made reference to differences in standardization as likely contributors to the variation in serum folate results, and/or differences in preparation of the whole-blood hemolysates as contributing to the variability in whole-blood folate results. The report served to underscore that little has changed with respect to between-method variability in folate testing since the round-robin surveys 8–10 years ago.

Heterogeneous assays for folate measurement represent an interesting and challenging problem from a standardization standpoint. Technically, folate methods are not immunoassays because they all rely on a high-affinity folate-binding protein (FBP) as the specific capture entity rather than anti-folate antibodies. In addition, folate (as pteroylglutamic acid used for food fortification) is available as a stable, highly purified powder, which would seem to be ideal for gravimetric preparation of standards. With both analyte and analyte binders seemingly normalized across methods, why the large differences?

The answer, which we discuss later, highlights once again the general difficulty in standardizing immuno- and bioreceptor-based assays in an absolute sense. Stemman (4) recently wrote a thoughtful review of the scientific and practical difficulties associated with standardizing immunoassays. In his review, he discussed standardization between available commercial assays in lieu of an accepted primary reference method and primary standards. One option mentioned was “harmonization”, in which different methods are calibrated to read the same, ideally to a “best available” candidate reference method. Stemman pointed out that this approach is the least desirable toward improving interlaboratory/method comparability compared with an ideal of absolute standardization with analytically correct standards and an agreed on reference method of higher order. At the same time, it is also apparent, given the practical and technical issues that were raised, that absolute standardization of immunoassays is a long way off, if even possible, for most immunoassay analytes. With this reality, as well as the reality that absolute standardization does not necessarily remove significant between-method differences, perhaps it would be useful to look more closely at harmonization as a practical, real-world approach to improving comparability of results between laboratories. As Lequin (5) pointed out in response, absent the ability to achieve absolute trueness in standardization, the final aim is to improve comparability of results obtained for patient samples between laboratories.

To facilitate discussion of the issues involved in standardization, the analyte to be standardized has been categorized into one of two general types: an analyte (or hapten) that is traceable to SI units; and an analyte for which measurement results are not traceable to SI units, but rather are reported in arbitrary units (5). The latter type is typified by protein analytes, which may be of a complex or ill-defined nature. With chemically well-defined analyte and assignable SI units, there may be a general tendency to consider the “hapten” bucket as more straightforward to standardize. Although this may be true in the absolute analytical sense, it does not mean that harmonization of results will be the outcome. Variability among methods for 17β-estradiol, for example, despite being “standardizable” in theory, has prompted suggestions for a harmonization approach. Thiennpont and De Leenheer (6) have pointed out that, despite this well-defined situation, standardization may be far easier in concept than in practice. Split-sample comparisons and calibrator value assignments, along with the willingness of manufacturers to participate, were highlighted as a potential breakthrough in harmonization of estradiol testing. The association made between calibration traceability and the traceability of results (“accuracy based standardization”) (7) would seem to point the standardization discussion away from analytical ideals toward the direction of the patients being served.

The difficulties with proteins, even when absolute SI units can be assigned to reference materials, are well known (8–10). Even with a well-characterized reference material (such as recombinant troponin ITC complex), immunoassays differing in epitope specificity will require compensating calibration value assignments to accomplish comparability of results.

Perhaps the most successful example of addressing between-method differences in patient results are the ongoing National Glycohemoglobin Standardization Program and IFCC efforts to harmonize different methodologies for measurement of glycohemoglobin. Methodologies specific for different fractions of glycohemoglobin are harmonized to report a single subfraction (hemoglobin A1c), which has been chosen as the target analyte for standardization. Assay methods measuring glycohemoglobin fractions in addition to the A1c fraction were harmonized to report only the A1c fraction, thus converting them via standardization to a seemingly more specific methodology. Given the well-established relationship be-

---

1 Although technically folate assays using FBP are not “immunoassays”, for purposes of discussion of standardized issues, the term immunoassay will be considered as applicable to folate methods.
tween the A1c subfraction and the total glycosylated fraction, such a standardization conversion can be justified scientifically. The National Glycohemoglobin Standardization Program leaves it up to assay manufacturers to determine the manner of harmonization, which may be accomplished by reassignment of calibrator values or by use of a conversion equation (11). Thus, whereas Lequin (5) is correct to point to the international A1c effort as a good example of how the standardization process should work, it might also be acknowledged that this is an equally good example of the principle of harmonization as a standardization approach and how it can improve comparability of results among laboratories and methodologies with patient samples.

The terms harmonization and standardization are frequently used interchangeably, and some clarification of terminology might be useful. Harmonization has been referred to as both a specific standardization approach and as a desired outcome of standardization. Harmonization as a standardization approach might best be described as a method of standardizing in which calibrator value assignments are made on the basis of values obtained by a higher order reference method (or candidate reference method) rather than on the basis of absolute known quantities. In contrast, “absolute” standards might be thought of as standards assigned on the basis of absolute known quantities, or quantities best estimated by the highest order methodology available. In this respect, efforts to standardize troponin and A1c testing are best understood as a blend of absolute standardization and harmonization of different assay methods to the chosen reference method and materials. It may be appropriate to give greater recognition to the prominent place of harmonization in addressing the larger goal of improving comparability of results among laboratories and methods for real patient samples. From an analytical standpoint, harmonization can certainly be applied to the hierarchy of standardization (12). Indeed, the marriage of absolute standardization (analytical ideal) and harmonization can come from the choice of which “higher order” reference method and materials with use for harmonization. In cases in which progress on absolute standardization is lacking, a best possible “candidate” or “designated” reference method can provide the basis for harmonizing. As has been pointed out (13), harmonizing is best carried out with fresh patient samples to preclude potential influences of matrix effects that can occur with processed or artificial matrices used for calibrator material.

Folate testing represents yet another situation in which the greater aim of improving comparability of results between laboratories and assay methods may not be compatible with absolute “trueness” of standardization across all methods. Recognizing this, we changed course in our standardization approach for the ARCHITECT® Folate assay from the one taken with two previous folate products, IMX® Folate and AxSYM® Folate.

Despite the availability of pure, chemically well-defined preparations of “folate” and the use of a common binding protein (FBP) for capture, absolute standardization appears to be at odds with harmony across methods. The underlying reasons may be numerous. First consider the analyte. Folate is a generic term encompassing several forms of folenic compounds. The form of folate used for food fortification and vitamins is the pteroylglutamic acid (PGA) form because of its well-known stability. When ingested, PGA is metabolized predominantly to a physiologically active 5-methyltetrahydrofolate (5mTHF) form (14). 5mTHF contains two asymmetric carbon atoms, and the physiologic form of 5mTHF is the optically active [6S,aS] 5mTHF diastereomer (also referred to as L-5mTHF). 5mTHF is highly susceptible to oxidation; therefore, its use as a folate assay standard reference material poses practical difficulties. Because FBP can, under the right reagent conditions, bind PGA with an affinity comparable to that for 5mTHF, assay calibrators for folate are typically the stable PGA form. With proper handling precautions, assay standardization can be accomplished with either PGA or 5mTHF. Commercially available 5mTHF, because of the steps in its synthesis, has until recently existed as a 50:50 racemic mixture of natural and unnatural diastereomers. Therefore, the path from practical folate assay calibrators to the form of folate actually being measured in patient samples has not been a direct one.

A second source of difficulty for harmonization of folate testing is the FBP. This protein exhibits marked aggregation tendencies, which affect its affinity for folate (15, 16). It is less clear how the relative affinities for PGA and 5mTHF are affected, particularly by the method of FBP incorporation into assay reagents. In addition, FBP can exhibit different selectivity for the diastereomers of 5mTHF. Using the natural 5mTHF diastereomer prepared via an enzymatic process, Shane et al. (17) showed that different folate methods using purified FBP as the capture protein exhibited markedly different responses to the natural diastereomer and racemic 5mTHF. It is not known how variables of aggregation and/or incorporation into assay reagents can impact FBP stereoselectivity. An implication is that standardization of folate methods with racemic 5mTHF will not ensure between-method comparability with patient samples. Standardization of folate methods with PGA reference material also would not improve things because the same problem of variable affinity for the predominant physiologic form would remain. Obviously, headway could be made only with optically pure (6R) 5mTHF reference material.

A third challenge in harmonizing folate methods is associated with markedly different FBP affinity profiles for PGA and 5mTHF as a function of pH (18). A consequence of this is that slight differences in reagent pH can directly impact patient results if the assay calibrators use the stable PGA form of folate. During the development of Abbott’s folate assays, all of which use PGA for calibrators, we found that uptake of atmospheric CO2 by one of the sample denaturant reagents, such as can occur when the reagent bottle is left open, can significantly alter patient results if not carefully controlled. As a result, a 5mTHF quality-control sample is provided to monitor reagent pH integrity. If a consensus reference material of 5mTHF or PGA were made available as absolute folate standards, this would not normalize or control for differ-
ences among methodologies that can arise from even minor pH variations.

Previous folate standardization efforts at Abbott used gravimetrically correct, spectrophotometrically value-assigned primary standards of the purest 5mTHF form of folate available. This approach attempted to mimic the form of folate found in patient specimens and to be as analytically correct as possible. In addition, reagent pHs were adjusted such that the PGA and 5mTHF dose–response curves were matched, permitting the use of stable PGA for assay calibrators. Although this scenario was as close to analytical correctness as was possible, differing dose–response characteristics between the standards and actual patient samples created linearity problems. In essence, there was greater affinity for the standards than for real patient specimens. This difference in affinity may have been attributable to a stereoselectivity bias from the FBP toward the unnatural diastereomer of the racemic 5mTHF.

During the development of the ARCHITECT Folate assay, the absolute standardization approach was abandoned, and harmonization of the assay to a designated external reference method was attempted. The first step in harmonizing is, of course, the choice of which method to harmonize with. Because there has been no recognized folate reference method or suitable candidate reference method described until recently (19,20), the Bio-Rad Quantaphase II radioassay was designated as the “reference method” for several reasons: (a) details of the gravimetrically correct PGA standardization of this method have been published (21); (b) it was used as the reference method in the CDC round-robin folate assay proficiency survey (2); (c) it is an equilibrium, boil method that is relatively insensitive to factors that can suppress results from individuals with renal impairment [a sample type that can give falsely depressed results in nonequilibrium automated folate methods (D. H. Wilson and R. Herrmann, unpublished observations)]; and (d) the method exhibits good overall analytical performance characteristics.

There are different approaches to harmonizing assay methods. These include use of an after-measurement conversion factor or equation (22), calibrator assignment or value reassignment based on an agreed on reference material, and calibrator value assignment based on split-sample testing of patient samples. For the ARCHITECT Folate, the latter procedure was used, whereby split patient samples were tested with the candidate reference assay and the ARCHITECT assay. Data from a set of serum samples from the reference assay plotted against the ARCHITECT raw signal are shown in Fig. 1. This relationship was used to establish the “patient” dose–response characteristics in the ARCHITECT assay. A zero calibrator (human serum albumin with Tris buffer) was used to anchor the dose response to a folate-free “zero” signal. These data were then fitted by a four-parameter logistical curve (4PLC) to define the targeted dose response of the calibrators.

Also shown in Fig. 1 are the dose responses of gravimetrically prepared PGA folate standards. Gravimetrically correct standards exhibited a steeper dose response than did samples from real patients. The differences between the gravimetric calibration curve and the “patient curve” were quantified mathematically and used to provide gravimetric correction factors for the calibrators. This process gave harmonized standards that were prepared to mimic the dose response of patient samples as defined by the designated reference method. The gravimetric correction factors were determined mathematically by converting the raw instrument signal from the gravimetric calibrators to folate concentration values by use of the 4PLC equation defined by the fitted patient sample curve. The resulting ratios of the gravimetric quantities to the values determined from the 4PLC equation gave the correction factors needed for preparing harmonized primary reference standards. On determination of the gravimetric correction factors, subsequent batches of standards can be prepared, as needed, by duplicating the quantities of the initial batch; i.e., the process becomes gravimetric.
cally based and no longer needs to involve testing of patient samples.

In general, long-term independent monitoring of harmonized calibrators for stability and/or between-batch consistency can occur with any suitable independent method that is available. The values of the initial batch can be characterized by the independent method and can provide long-term target values for quality-control monitoring. For ARCHITECT Folate, routine assay calibrators are manufactured by making fine adjustments in the calibration concentrations until the raw instrument signal of each in-process calibration concentration matches the instrument signal of the corresponding harmonized primary reference standard to within a predetermined specification. As a general approach to harmonization, the above method was chosen because factors that can cause “disharmony” between gravimetrically correct calibrators and the patient dose response may not be linear across the assay range. This approach allows immediate mathematical determination of the correction factors based on the 4PL fit of actual patient sample data.

Harmonization can provide analytical performance benefits in addition to improving interlaboratory/method comparability. Intralaboratory precision benefits have been reported for some methods as a result of the harmonization effort in glycohemoglobin testing (12). With ARCHITECT Folate, the nonlinearity noted with the IMx Folate and AxSYM Folate was resolved with harmonized standardization. As a specimen is diluted, the dilution response should, in the absence of matrix effects introduced by the diluent, follow a “specimen” dose response. Determining results for diluted samples based on a calibrator dose response that is mismatched to the true sample response will give inaccurate results (Fig. 1, inset).

Harmonization of ARCHITECT Folate patient sample results with the Quantaphase II Folate method was demonstrated in a recent method comparison reported by Owen and Roberts (3) for both serum and whole-blood folate. In that report, correlation slopes in method comparisons between ARCHITECT Folate and Quantaphase II for serum and whole-blood folate measurement were 0.99 and 1.06, respectively (R =0.95), compared with 0.60–2.99 for other automated methods. This is not to suggest that one method is better than another method, but the potential for harmonizing all folate methods exists, which would address what has been described as an urgent need (2).

From a broader standpoint, harmonization is a principle that should be more openly discussed and recognized as an analytically valid procedure for improving comparability of different methods, particularly in situations where absolute standardization appears unlikely, unrealistic, or even counterproductive. Absolute standardization efforts might be most productively leveraged toward defining stable higher order methodologies and reference materials to which immunoaassay methods can harmonize. A mindset of seeking to force-fit immunoaassay methods to analytical ideals and absolute standardization criteria will likely remain at odds with the ultimate objective of improving comparability between laborato-

References


David H. Wilson* 
Gregg Williams 
Robert Herrmann 
Dallas Wiesner 
Paul Brookhart

Abbott Diagnostics Division 
Abbott Laboratories, Abbott Park, IL 60064

*Author for correspondence. Fax 847-937-9805; e-mail david.h.wilson@abbott.com.

DOI: 10.1373/clinchem.2004.042358