Clinical Chemistry 60:7
923–927 (2014)

Q&A

Harmonization of Test Results: What Are the Challenges; How Can We Make It Better?

Moderator: W. Greg Miller[1,2*]
Experts: John H. Eckfeldt,[3,4] Joseph Passarelli,[5,6] William Rosner,[7,8] and Ian S. Young[9,10]

Laboratory test results are used to inform decisions on the diagnosis and treatment of disease. Consistent and comparable results from different measurement procedures are important for developing clinical practice guidelines and for those guidelines to be applied to decisions about patient care. ISO document 17511:2003 (*In vitro diagnostic medical devices -- Measurement of quantities in biological samples -- Metrological traceability of values assigned to calibrators and control materials*) states that calibration of routine clinical laboratory measurement procedures be traceable to higher-order reference materials and reference measurement procedures. Such traceability can achieve consistent and comparable results that are sustainable over time and among different measurement procedures.

The terms "standardized" and "harmonized" are frequently used interchangeably to refer to the condition in which results are consistent and comparable among different measurement procedures. "Standardized" refers to the condition in which calibration is traceable to a reference measurement procedure that is typically calibrated with an appropriate reference material. There are situations in which the reference measurement procedure defines the analyte without any primary reference material (e.g., enzyme activity). Standardization has the advantage of trustworthy reproducibility over time and location, because the reference measurement procedure provides a stable anchor for calibration traceability of routine clinical laboratory procedures. "Harmonized" is a more general term. It may include the standardized condition, but it also includes the condition in which results are consistent and comparable in the absence of a reference measurement procedure. The process of harmonization typically depends on the availability of a suitable reference material that can be used as a common calibrator among routine clinical laboratory measurement procedures.

Despite the availability of reference measurement procedures and reference materials, results for many analytes are neither consistent nor comparable when measured with different clinical laboratory procedures. In this Q&A, experts in harmonization of test results offer their opinions on the current state of the art and how we might make it better.

Why do clinical practice guidelines include laboratory test decision values when the laboratory measurement procedures are not harmonized?

John Eckfeldt: Many of the clinical guidelines that use laboratory data were developed almost exclusively by groups of clinical experts with little knowledge of the variability of laboratory test results produced by different laboratory measurement procedures. In many cases, results from large research studies used to develop such guidelines came from a single central laboratory, and the groups promoting adoption of these guidelines were simply unaware that other widely used measurement procedures can give very different test results. A good example is the initial guidelines from clinical societies for prevention of thrombotic complications, where a prothrombin time expressed in seconds was originally recommended as an indicator of adequate warfarin anticoagulation without realizing that different thromboplastin reagents gave markedly different prothrombin times in seconds. Only with the international normalization ratio approach was the problem essentially solved.

[1] Professor of Pathology, Virginia Commonwealth University Health System, Richmond, VA; [2] Chair of the IFCC Working Group on Commutability; [3] Ellis Benson Professor of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; [4] Member (and past chair) of the College of American Pathologists Standards Committee; [5] Senior Director, Scientific Relations, Roche Diagnostics, Indianapolis, IN; [6] Roche Professional Diagnostics' representative to the ISO; [7] Professor of Medicine, Columbia University, College of Physicians and Surgeons, New York, NY; [8] Member of the Harmonization Oversight Group of the International Consortium for Harmonization of Clinical Laboratory Results; [9] Professor of Medicine, Queen's University Belfast, Belfast, UK; [10] Chair of the IFCC Scientific Division.
* Address correspondence to this author at: P.O. Box 980286, Virginia Commonwealth University Hospital, Richmond, VA 23298-0286. Fax 804-828-0353; e-mail gmiller@vcu.edu.
Received September 24, 2013; accepted September 26, 2013.

Joseph Passarelli: Although guideline producers have the goal to have clear and unambiguous recommendations, there often is a lack of understanding of many contributing factors that lead to differences of results. Careful consideration is needed with respect to test principles and chemistries, instrumentation, commutability of calibrators, biostatistical methodologies, etc., as well as a greater understanding of practical limitations of measurement technologies. Inclusion of laboratory professionals and participation from industry during the development and implementation phases of the guidelines could improve the situation.

William Rosner: Most physicians not trained in laboratory medicine believe that any value arriving from the laboratory is correct and immutable. This holds true at the time when many guidelines are written and is often not addressed until unrest arises after guidelines are published. Cut points and normal ranges may have been arrived at from limited sampling (e.g., not addressing ethnicity, nationality, age, disease, etc.) and, equally as important, based on analytes that were not harmonized. The inevitable consequence is that under these conditions the guidelines are not useful for their intended audience.

Ian Young: Many clinicians are unaware of the extent to which different results may be obtained from different assays for the same analyte. This limitation occurs because most clinicians routinely see results from a single laboratory and a single method, and interlaboratory or intermethod differences are not apparent to them. Clinical practice guidelines are often generated by expert clinicians without significant laboratory input, and this can lead to inappropriate absolute decision limits on which treatment decisions are made. At its worst, whether or not a patient receives specific treatment (growth hormone replacement, for instance) may depend on which laboratory analyzes a key sample, rather than on the actual clinical need.

Since we have reference systems with reference measurement procedures for 50 or more analytes, why are results for these analytes sometimes different when measured with different clinical laboratory procedures?

John Eckfeldt: For well-defined analytes (i.e., measurands) that exist in only a single molecular form (e.g., creatinine, cholesterol) in clinical samples, use of reference measurement procedures and reference materials that have been proven to be commutable with all clinical measurement procedures for which they are intended can largely eliminate nonharmonized results. For complex measurands that exist in various molecular forms (e.g., cardiac troponin, human chorionic gonadotropin), the situation is far more difficult. Any possible solution requires a clear definition of the measurand and an understanding of the molecular forms of the related molecules, both in the clinical samples and in the reference materials.

Joseph Passarelli: There are many potential causes of why there are differences in patient results, even when reference systems are in place. Perhaps the most important of all is commutability: not only for the primary reference material but also for the secondary reference materials—calibrators and controls from the manufacturer. Significant effort is made to prepare these materials to produce patient results that are as close to "trueness" as possible, but other factors are in play as well: differing assay technologies and instrumentation, varying preanalytic processes, reagent handling, etc. In addition, immunoassays based on antibodies or protein-recognition events can lead to significant variation due to specificity differences. In the end, it is not possible to address all of these factors for every analyte.

William Rosner: Reference measurement procedures are used only rarely, if ever, by clinical laboratories. Rather, they use less expensive, less time-consuming, and more efficient routine procedures. Their results should be traceable to those obtained with the reference measurement procedure, but for a variety of reasons (including specificity of the reactions, differences in the calibrators, and lot-to-lot variability in value assignment) the results for clinical samples may differ among routine assays. Furthermore, matrix effects, e.g., the composition of the biological samples and the

presence of physiologic interfering substances, generally influence routine procedures differently, whereas a reference measurement procedure is typically insensitive to many interfering substances.

Ian Young: In theory, the existence of a full reference system to which clinical laboratory procedures are traceable should mean that clinical laboratory measurements produce consistent and reliable results in all laboratories. In practice, that is not always the case. The reasons for the discrepancies are multiple but in many cases relate to the differences between the procedures used routinely in the clinical laboratory and the reference measurement procedure. If a routine clinical method has a reduced specificity compared with the reference method, for instance due to antibody cross-reactivity or the presence of an interfering substance, then different results may be produced. It is important that the clinical laboratory remains alert to such possibilities and does not assume that traceability to a reference measurement system is sufficient to ensure result comparability.

Reference materials are available for quite a few analytes for which there are no reference measurement procedures. Why are results different when measured with different clinical laboratory procedures that are calibrated with the same reference materials?

John Eckfeldt: I believe that failure to test commutability of reference materials is one of the main reasons for the lack of harmonization of clinical laboratory test results that have well-established reference systems. One of the problems is that reference-material producers, often national metrology institutes, lack ready access to a wide variety of clinical samples from patients for whom the test is intended to be used, while in vitro diagnostics (IVD) manufacturers using these reference materials lack the resources or incentive to organize round-robin testing with other IVD manufacturers' measurement procedures to properly assess commutability. Only with close cooperation among reference-material producers, IVD manufacturers, and clinical laboratories can commutability be properly tested and clinical measurement procedures be harmonized.

Joseph Passarelli: Once again, commutability of the reference material is a major issue. To improve the situation to some extent, it would be helpful if manufacturers could be involved as early as possible before establishing a new reference material so that commutability could be carefully assessed. In addition, manufacturers use different internal standardization/harmonization procedures, and information sharing of best practices would help. For example, most often reference materials are supplied only at one concentration, and the nonlinearity of certain assays and internal dilution processes can further exacerbate the situation. Finally, there can be uncertainty as to the relationship between the reference material and those factors that might affect the analytical result of native patient samples.

Ian Young: Availability of a reference material is an important component of a reference measurement system but on its own is insufficient to ensure comparability of results between different routine clinical laboratories and methods. There are many examples of methods that purport to measure the same measurand but in reality do not, as a result of different antibody specificities or minor differences in chemical reaction conditions. Use of a common validated reference material to provide an anchor for methods is not sufficient to achieve comparability of clinical results between methods in such cases.

What are the main challenges for harmonization of results from laboratory-developed tests?

John Eckfeldt: Initially, most new laboratory-developed tests lack any accepted reference system and reference materials known to be commutable in clinical measurement procedures, which often use widely different methodological approaches with substantial differences in analytical specificity. For many "emerging analytes," it can take many years or even decades to fully understand the molecular heterogeneity and presence of all the molecular species related to the biomarker that can influence the measurement process in both clinical samples and candidate reference materials.

Joseph Passarelli: Tests developed by manufacturers are under strict design control and oversight by regulatory authorities worldwide. As a result, harmonization studies (comparison to other methods, measurement of candidate reference materials, traceability, etc.) are planned and executed during the development process before commercial launch. Generally, this is not the case with laboratory-developed tests, and the challenge to harmonize these is much greater and maybe impossible in certain situations. In addition, other sources of variation, such as reagent lot-to-lot differences, reagent stability, instrumentation inconsistencies, etc., are often overlooked or not even assessed for laboratory-developed tests.

William Rosner: The overarching reason (in addition to the details mentioned previously) is the "informa-

tional isolation" of each laboratory. The individual laboratory has neither the broad means of communication nor the incentives to see the test they have developed opened to the general laboratory community and thence have the opportunity to become harmonized with similar procedures developed by other laboratories.

Ian Young: Achieving harmonization of results from laboratory-developed tests provides an even greater challenge. Minor differences in methods used by individual laboratories will be very common, and between-laboratory variation is likely to be greater. Mass spectrometry is the area of technology where in-house assay development is most common. Results will be influenced by many factors related to instrumentation, extraction procedures, signal processing, etc. There may not be any available reference material, although as discussed above, this would not be sufficient on its own to ensure comparability of results. Internal QC will help to ensure consistency of results in an individual laboratory. Regular monitoring through external quality assessment (proficiency testing) is also important, but since many quality-assessment schemes do not routinely utilize commutable samples, the data need to be interpreted carefully.

It seems the laboratory community is always trying to fix lack of harmonization after several measurement procedures for an analyte are in use. What can we do to ensure that different measurement procedures for a new biomarker will give comparable results?

John Eckfeldt: For "new" analytes, there is typically an incomplete understanding of their molecular forms in the clinical samples and often some disagreement as to their bioactivity and what forms are best measured for the diagnostic or therapeutic questions at hand. This leads to uncertainty as to which IVD manufacturer's test is really most useful clinically. In addition, harmonization with its goal that all IVD manufacturers' results become equivalent tends to drive the reagent price down by making the laboratory test a commodity, thus giving a further economic disincentive to harmonization. However, strong pressure by clinician- and laboratory scientist–based societies can overcome these scientific and economic factors and make harmonization a priority based on patient-safety goals.

Joseph Passarelli: Perhaps I am stating the obvious, but this situation could be significantly improved if the development of the reference material/reference method were made in parallel with the development of the measurement procedure for a new biomarker. To minimize any conflict of interest and to allow for broader access and utilization, officially recognized reference materials could be made available to all manufacturers very early on through an impartial third-party organization. Such organizations could also facilitate cooperation and collaboration among all relevant constituencies (diagnostics, pharmaceuticals, laboratories, regulatory agencies, payers), thus strengthening acceptance by the scientific community in the end.

William Rosner: This is really difficult. I don't see how harmonization for new biomarkers can be accomplished, short of a government regulatory agency setting up a requirement that the developer of an assay provide standards for others to use to harmonize their measurement procedures with the original. Also, such standards may need to be amended/changed as new information becomes available.

Ian Young: Whenever a new biomarker is being developed, the emphasis, rightly, is on measurement reproducibility and demonstration of clinical utility. When several manufacturers (or laboratories) independently develop assays, they usually pay little attention to agreement with assays that have already been approved or marketed. Regulatory authorities also pay insufficient attention to this issue. Responsibility for addressing it should be shared by manufacturers, clinical laboratories, and regulatory authorities. There is a need to define what is meant by an acceptable degree of harmonization between assays. Manufacturers should feel a responsibility to ensure that any new assay they seek to market meets this definition, and clinical laboratories should implement only an assay that is suitably qualified. Ultimately, if manufacturers and clinical laboratories cannot address the issue, regulatory approval may need to be made dependent on satisfying predefined harmonization criteria.

Do regulatory requirements help or hinder harmonization of measurement procedures? Is there room for improvement?

John Eckfeldt: At times, the regulatory process seems to inhibit harmonization of laboratory IVD measurement procedures. I have been told of instances in which regulatory agencies have required a full new IVD regulatory submission (including newly established reference ranges, imprecision studies, etc.) for a measurement procedure simply being recalibrated to a new reference material or accuracy base for a reference measurement procedure. In many cases, simple mathematical adjustments of previously submitted data by using known conversion factors from pre- vs. postrecalibration test results would suffice. Hopefully, this issue

can be addressed going forward through better collaboration between regulators and individual IVD manufacturers or their organizations.

Joseph Passarelli: Yes and no, and there is certainly room for improvement. In a positive way, regulatory requirements (proof of traceability, comparison to predicate devices, assessment of uncertainty, etc.) significantly highlight and enhance the importance and benefits of standardization/harmonization. Overall, this is seen as positive by device manufacturers. Unfortunately, the requirements and guidance are not always clear and often vary between different countries, markets, and federations/unions. Therefore, significant effort, rework, and complicated and expensive studies are often required to commercialize a new device worldwide. A more consistent and predictable worldwide regulatory process would greatly streamline and improve the current situation.

Ian Young: Regulatory requirements could help harmonization, but at present their effects are mixed. The requirement in some jurisdictions (Europe, for instance) to demonstrate traceability to a higher-order reference measurement system where one is available is helpful. On the negative side, where harmonization is a voluntary activity, changes in an assay by a manufacturer to achieve harmonization may be made more expensive by regulatory requirements. This hinders the harmonization process. It could be resolved if harmonization were to become a mandatory requirement. However, this too is not without risks. A new, more specific assay might have to be adjusted to harmonize with established, less specific assays, and this would hinder scientific advancement and possibly prove harmful to patients.

What is the clinical laboratory community doing to improve harmonization of test results?

John Eckfeldt: Professional organizations such as the AACC, the IFCC, and others have provided a platform to address harmonization from an international perspective. Without international cooperation, there is a risk that various regional groups will develop a reference system and an accuracy base that are inconsistent with those of other regional organizations. However, I believe strong pressure from clinical organizations is needed to really drive the harmonization process forward for a given analyte or group of analytes.

Joseph Passarelli: Most manufacturers collaborate with various professional organizations that have a focus or some interest in the standardization/harmonization of laboratory results. A few of the most notable are activities coordinated through the Scientific Division of the IFCC via a number of committees and working groups: the Joint Committee for Traceability in Laboratory Medicine, the AACC via the International Consortium for Harmonization of Clinical Laboratory Results, external quality assessment programs, and several others. In addition, there are a few standardization networks that consist of a number of reference laboratories with a focus on a specific measurand (for example hemoglobin A_{1c}). For success, participation from all key stakeholders (laboratories, industry, regulatory agencies, payers, etc.) is necessary.

Ian Young: There is no doubt that the clinical laboratory community is aware of the need for harmonization, although awareness among clinical users and patients is much more patchy. The issues involved are complex and difficult to explain to individuals and groups without knowledge of standardization issues. The IFCC Scientific Division has managed a program of standardization activities for many years, which has resulted in considerable improvements in some areas (for instance, enzyme activity measurements). The program has developed in a piecemeal way, however, and there has not been any systematic approach to prioritizing measurands. The AACC initiative to establish the International Consortium for Harmonization of Clinical Laboratory Results will help to address this issue.

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: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: W.G. Miller, *Clinical Chemistry*, AACC, and CLSI; J. Passarelli, Roche Diagnostics; I.S. Young, *Clinical Chemistry*, AACC, and IFCC.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: None declared.
Expert Testimony: None declared.
Patents: None declared.
Other Remuneration: I.S. Young, IFCC.

Previously published online at DOI: 10.1373/clinchem.2012.201186