Clinical Chemistry 57:8
1108–1117 (2011)

Special Report

Roadmap for Harmonization of Clinical Laboratory Measurement Procedures

W. Greg Miller,[1*] Gary L. Myers,[2] Mary Lou Gantzer,[3] Stephen E. Kahn,[4] E. Ralf Schönbrunner,[5] Linda M. Thienpont,[6] David M. Bunk,[7] Robert H. Christenson,[8] John H. Eckfeldt,[9] Stanley F. Lo,[10] C. Micha Nübling,[11] and Catharine M. Sturgeon[12]

Results between different clinical laboratory measurement procedures (CLMP) should be equivalent, within clinically meaningful limits, to enable optimal use of clinical guidelines for disease diagnosis and patient management. When laboratory test results are neither standardized nor harmonized, a different numeric result may be obtained for the same clinical sample. Unfortunately, some guidelines are based on test results from a specific laboratory measurement procedure without consideration of the possibility or likelihood of differences between various procedures. When this happens, aggregation of data from different clinical research investigations and development of appropriate clinical practice guidelines will be flawed. A lack of recognition that results are neither standardized nor harmonized may lead to erroneous clinical, financial, regulatory, or technical decisions.

Standardization of CLMPs has been accomplished for several measurands for which primary (pure substance) reference materials exist and/or reference measurement procedures (RMPs) have been developed. However, the harmonization of clinical laboratory procedures for measurands that do not have RMPs has been problematic owing to inadequate definition of the measurand, inadequate analytical specificity for the measurand, inadequate attention to the commutability of reference materials, and lack of a systematic approach for harmonization. To address these problems, an infrastructure must be developed to enable a systematic approach for identification and prioritization of measurands to be harmonized on the basis of clinical importance and technical feasibility, and for management of the technical implementation of a harmonization process for a specific measurand.

In October 2010 the AACC convened a conference to address how to improve harmonization of laboratory test results for which there are no higher-order reference measurement procedures (RMPs),[13] and for which it was unlikely that such procedures could be developed. The conference participants are listed in the Supplemental Data file in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue8. Examples of such tests include thyroid-stimulating hormone, human chorionic gonadotropin (HCG), prostate-specific antigen, troponin I, natriuretic peptides, carcinoembryonic antigen, luteinizing hormone, hydroxylated vitamin D vitamers, Epstein-Barr virus, and BK virus, all of which have complex molecular forms and closely related molecules that in many cases vary under different pathophysiologic conditions. The primary goal of the conference was to develop consensus on organizational and technical processes to achieve harmonization of results from clinical laboratory testing procedures for these types of measurands. The term measurand means the quantity intended to be measured. In this report the term measurand also includes multiple molecular forms when the clinically important component may not be fully understood.

Results for a given measurand should be numerically equivalent, within clinically meaningful limits, among different laboratories using different measurement procedures. The term "standardized" is used

[1] Virginia Commonwealth University, Richmond, VA; [2] AACC, Washington; [3] Siemens Healthcare Diagnostics, Newark, DE; [4] Loyola University Health System, Maywood, IL; [5] Life Technologies, Benicia, CA; [6] Ghent University, Ghent, Belgium; [7] National Institute of Standards and Technology, Gaithersburg, MD; [8] University of Maryland, Baltimore, MD; [9] University of Minnesota, Minneapolis, MN; [10] Medical College of Wisconsin, Milwaukee, WI; [11] Paul-Ehrlich Institut, Langen, Germany; [12] Royal Infirmary of Edinburgh, Edinburgh, UK.
* Address correspondence to this author at: P.O. Box 980286; Richmond, VA 23298-0286. Fax 804-828-0353; e-mail gmiller@vcu.edu.
Received February 24, 2011; accepted May 13, 2011.
Previously published online at DOI: 10.1373/clinchem.2011.164012

[13] Nonstandard abbreviations: RMP, reference measurement procedure; HCG, human chorionic gonadotropin; SI, International System of Units (*Système International d'Unités*); CLMP, clinical laboratory measurement procedure; ISO, International Organization for Standardization; JCTLM, Joint Committee for Traceability in Laboratory Medicine; IVD, in vitro diagnostics; STARD, Standards for Reporting of Diagnostic Accuracy; TEa, total allowable error; CVi, within-subject biological variation, PT, proficiency testing; EQA, external quality assessment.

Table 1. ISO 17511:2003 categories for reference systems.

Category	Reference measurement procedure	Primary (pure-substance) reference material	Secondary (value-assigned) reference material[a]	Examples
1	Yes	Yes	Possible	Electrolytes, glucose, cortisol
2	Yes	No	Possible	Enzymes
3	Yes	No	No	Hemostatic factors
4	No	No	Yes	Proteins, tumor markers, HIV
5	No	No	No	Epstein-Barr virus, varicella zoster virus

[a] More than 1 secondary reference material, with potentially different properties, may be available for the same measurand.

when results for a measurand are equivalent and, in addition, the results are traceable to the International System of Units (*Système International d'Unités*) (SI) through a higher-order primary (pure substance) reference material and/or RMP. The term "harmonized" is generally used when results are equivalent either by being traceable to a reference material or based on a consensus approach, such as agreement to an all-methods mean, but neither a higher-order primary reference material nor an RMP exists.

Clinical practice guidelines based on standardized or harmonized laboratory tests contribute to quality care and may reduce overall healthcare costs. When laboratory test results are neither standardized nor harmonized, a different numeric result may be obtained for the same clinical sample. Unfortunately, some clinical practice guidelines base a diagnosis or treatment on test results from a specific clinical laboratory measurement procedure (CLMP) without considering the possibility or likelihood of differences between various CLMPs. When this happens, aggregation of data from different research clinical trials and development of appropriate clinical practice guidelines will be flawed by the lack of standardized or harmonized results. Clinicians and researchers may be unaware of the extent to which test values vary among different CLMPs, thus both hampering the effective use of clinical practice guidelines and setting up a potential source of errors in patient care. Lack of recognition that results are neither standardized nor harmonized may lead to erroneous decisions—clinical, financial, regulatory, and technical.

The experience of clinical laboratories has been that harmonization efforts have typically been initiated only after a measurand has been available for some time and limitations have been recognized with use in clinical practice or application of clinical practice guidelines. The importance of harmonization must be more prominently promoted and supported to enable the development and application of uniform clinical practice guidelines that depend on consistent laboratory results. A US Institute of Medicine report *(1)* specified the 6 aims for healthcare as safe, timely, efficient, effective, equitable, and patient centered. These aims provide excellent guidance for the goals of a harmonization process.

Traceability of Laboratory Results

Today, standardization and harmonization are based on traceability principles described in the International Organization for Standardization (ISO) standard 17511, which includes 5 categories of reference systems (Table 1) *(2, 3)*. Calibration traceability for measurands in categories 1, 2, and 3 is based on the availability of an RMP. There are well-established procedures to address standardization of measurands in these categories and to develop reference materials and RMPs for measurands. For example, the IFCC and the Joint Committee for Traceability in Laboratory Medicine (JCTLM) have promoted international implementation of traceability in these categories *(4, 5)*. When possible, it is preferable to establish calibration traceability to SI by using an RMP. As a result of the extensive global effort in standardization, categories 1, 2, and 3 will not be addressed in this report.

ISO traceability category 4 includes measurands for which 1 or more reference materials with a protocol for value assignment are available for calibration, but there is no RMP. Category 5 includes measurands for which neither RMPs nor reference materials for calibration are available. Measurands in categories 4 and 5 have been technically more difficult to address *(6)*. There have been few effective procedures implemented for harmonization in these categories *(7)*.

One successful example of traceability to a reference material in category 4, as well as category 1 for some measurands, is ERM (European Reference Materials)/IFCC-DA470k Proteins in Human Serum from the IRMM (Institute for Reference Materials and

Measurements), which has been effectively used to improve harmonization for the 12 measurands in the material *(8)*. However, many of the current reference materials available for category 4 measurands have not been validated for commutability with clinical samples, and CLMPs with traceability to such materials have been shown to produce nonharmonized patient sample results for several measurands *(9–12)*.

A potential limitation of reference materials in category 4 is that replacement preparations (new lots) may have different relative amounts of the molecular species related to the measurand of interest and different matrix characteristics from the preceding reference material. These limitations introduce the possibility for inconsistency between lots in the assigned concentrations and in the commutability with clinical samples *(13)*.

An important technical limitation is that many of the existing reference materials developed for measurands in category 4 have not been validated for commutability with clinical samples, and some that have been investigated have been shown not to be commutable. Consequently, a revision of the current practice that better meets the needs of clinical laboratories, in vitro diagnostics (IVD) manufacturers and healthcare providers is required. In addition, consensus procedures for harmonization of category 5 measurands are needed.

Challenges for Harmonization of Measurands in ISO Categories 4 and 5

DEFINITION OF THE MEASURAND

One technical issue in ISO categories 4 and 5 is measurement of heterogeneous analytes when the measurand of clinical interest is not well defined. For example, troponin I *(14)* and HCG *(15)* have different molecular forms in different clinical conditions. Consequently, the form(s) of clinical interest must be clearly understood so that measurement procedures can be developed that are specific for those forms. Detailed studies such as those undertaken for HCG *(15)* are desirable for any complex measurands to be harmonized. In some cases, more than one measurement procedure may be needed to provide suitable analytical specificity for a given clinical condition. It should be recognized that harmonization may not be possible for some measurands until the clinically important molecular form(s) are clearly identified.

ANALYTICAL SPECIFICITY FOR THE MEASURAND

The analytical specificity, e.g., the epitope recognized by an antibody, may be different for different CLMPs. Harmonization may require modification of existing CLMPs, or development of new CLMPs to improve specificity for the molecular form(s) of greatest clinical importance. Potential approaches for addressing measurement of complex protein mixtures have been suggested *(16, 17)* and must be considered when developing a technical plan for harmonization of such measurands. It should be recognized that harmonization may not be possible for some measurands until technical advances enable more specific measurement procedures.

COMMUTABILITY OF REFERENCE MATERIALS

In 1967 Radin proposed calibration traceability to reference standards to achieve harmonization among results from different CLMPs *(18)*. At that time, the potential for matrix-related biases when using noncommutable reference materials was not appreciated. Since then, the importance of using only commutable reference materials for calibration traceability and for assessment of harmonization of results among different CLMPs has been recognized *(2, 19–21)*.

Commutability is a property of a reference material whereby the same numeric relationship, within clinically meaningful limits, can be demonstrated between 2 or more measurement procedures for both the reference material and a panel of representative individual patient samples *(3, 19, 20)*. The concept of commutability is illustrated in Fig. 1, which shows one of several approaches to demonstrate the equivalence of a numeric relationship between results for reference materials and patient samples *(22)*. Fig. 1A shows results for commutable reference materials that have the same numeric relationship between 2 measurement procedures, as do the patient samples. Calibration of routine CLMPs to be traceable to a commutable reference material produces harmonization of results for patient samples. Fig. 1B shows that noncommutable reference materials have a different numeric relationship than that for the patient samples.

A noncommutable reference material cannot be used for calibration traceability because it does not have the same numeric relationship between measurement procedures as do the patient samples. Establishing calibration traceability to noncommutable reference materials produces differences in calibration and nonharmonization among results from different CLMPs *(19, 23, 24)*. Until recently, it has not been a common practice to validate commutability for secondary reference materials, but the importance of this requirement is now recognized *(2, 19–21, 22)*. Inconsistent results among methods may also be caused by inadequate analytical specificity that causes some CLMPs to be influenced by interfering substances or by molecular forms other than the intended measurand. Inadequate analytical specificity may also prevent a ref-

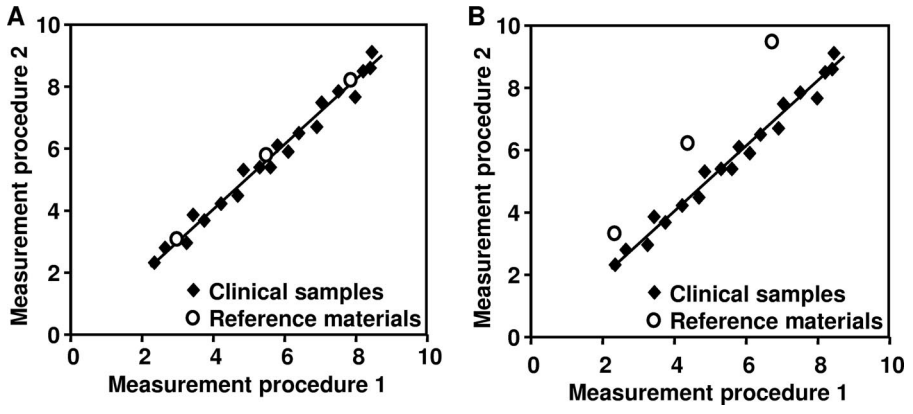

Fig. 1. Concept of commutability of results for reference materials with results for a panel of individual clinical samples.

(A), results for commutable reference materials that have the same numeric relationship between 2 measurement procedures as observed for a panel of patient samples. (B), results for noncommutable reference materials that have a different numeric relationship between 2 measurement procedures than observed for the patient samples.

erence material from being commutable for some CLMPs.

Noncommutability has been attributed to (1) manipulation and alteration of the sample matrix during preparation, even if the sample originated or was derived from human sources; (2) added nonnative forms of a measurand that can produce a different measurement than that expected for native forms; and (3) exclusion of clinically important forms owing to purification steps *(2, 25)*. The terms "matrix-related bias" and "matrix effect" are used to refer to the component of bias that is caused by noncommutability of a reference material rather than that due to a difference in calibration between measurement procedures.

Recommendations for a Systematic Approach to Harmonization

A systematic approach to harmonization must be developed that would identify measurands for which harmonization is needed, prioritize the measurands based on clinical importance and technical feasibility, and organize the implementation of harmonization activities. The process must be performed with the recognition that, when possible, standardization to the SI through a higher-order reference material and/or RMP is the preferred approach for measurands in ISO categories 1, 2, or 3. When an RMP is not available, a harmonization approach for measurands in ISO categories 4 and 5 as addressed here is needed.

PRIORITIZATION OF MEASURANDS TO BE HARMONIZED

Based on the Standards for Reporting of Diagnostic Accuracy (STARD) model *(26)*, a measurand prioritization checklist is proposed to ensure a thorough assessment and that all relevant information is available. Online Table S2 presents "Domains" and subheadings termed "Elements," formulated as questions, to assist with evaluation of each of the domains. For each domain, a higher or lower weight may be accommodated by simply altering the total point value. The use of ranges of scores for high, medium, and low priorities will allow a high-priority measurand to be actively harmonized independently of its sequential rank.

ESTABLISHING THE CLINICAL REQUIREMENTS FOR HARMONIZATION

The clinical requirements for a laboratory test have often been locally derived (e.g., based on in-house experience) and/or based on regional or national experience rather than on international consensus. Clinically based specifications for analytical performance are likely to be most critical when a decision to treat is based on small changes in results; for example, use of troponin for diagnosis in the emergency department *(14)*, prostate-specific antigen for population screening, *(10, 27)* or hepatitis C viral-load monitoring *(28)*. For patient monitoring based on laboratory results, determination of what constitutes a clinically relevant change in concentration is required *(29)*. Awareness of how a test is used in clinical practice is a prerequisite for clinical specifications for acceptable between-CLMP variability. Appropriately designed computer-simulation

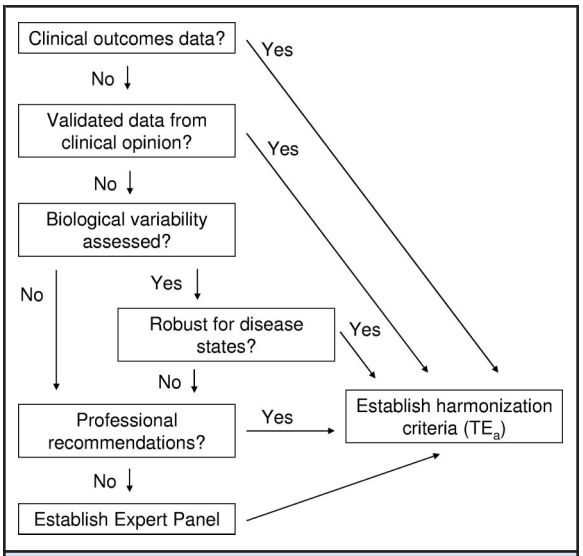

Fig. 2. Proposed schema for developing method harmonization criteria based on Kenny et al. *(31)*.

TE_a derived from allowable bias and imprecision, for the agreement of results among different measurement procedures.

studies can contribute valuable insight into the extent to which CLMP-related differences affect clinical interpretation of results *(27, 30)*.

A hierarchical approach for determining allowable bias and imprecision for laboratory tests is summarized in Fig. 2 *(31)*. Total allowable error (TEa) based on clinical outcomes data is ideal for such determinations but frequently is not available. Clinical opinion can be valuable but may be limited by inadequate presentation of the clinical scenario and analytical conditions. Data on biological variability, ideally derived from studies fulfilling the STARD criteria *(26)*, may enable a useful definition of TEa. When properly determined, data on within-subject biological variation (CVi) provides a measure of the presumably random average fluctuation observed for a "typical" individual in a group of reference individuals *(32)*. To be useful, the CVi needs to be known for healthy and diseased cohorts. Existing data on biological variability may not be sufficiently robust *(33, 34)*, nor adequately represent diseased cohorts. Consequently, levels of evidence based on CVi are variable and there is no overall coordination of this approach. Biological variation is more difficult to determine for measurands for which there may be a substantial CLMP nonspecificity contributing to the uncertainty in the derived CVi. Data applicable to discrimination between healthy and diseased conditions may be useful. Finally, when no other data are available, approaches based on recommendations from professional organizations, or an expert panel, are used to establish TEa for a given measurand.

ASSESSING THE CURRENT ANALYTICAL PERFORMANCE OF A CLMP

Sources of information to assess harmonization status include:

- Peer-reviewed scientific publications using panels of well-characterized patient samples,
- Interlaboratory comparisons that use commutable samples provided by proficiency testing (PT)/external quality assessment (EQA) schemes,
- Independent reports or reviews submitted or undertaken as part of regulatory approval for market, and
- Reports from postmarket surveillance.

These sources of information are all subject to various limitations, and to some extent each provides only a single time-point estimate of performance. Scientific publications must be scrutinized to ensure that the data reported are valid, i.e., that good experimental design was followed, particularly in relation to the number and quality of patient samples, and the number of different CLMPs included. The patient and analytical considerations included in the STARD guideline for studies of diagnostic accuracy *(26)* provide a good guide for assessing published reports. Data from studies in which panels of clinical patient samples were used will be most reliable. Data from PT/EQA schemes and other interlaboratory comparisons can be used only when the samples are commutable with clinical patient samples *(19, 20, 23–25, 35–37)* and of clinically relevant concentrations. Data from noncommutable PT/EQA samples will provide misleading information. Data submitted as part of a regulatory approval process are likely to have been obtained under conditions that may be more strictly controlled for clinical variables than those encountered in routine clinical laboratories, whereas postmarket surveillance generally relies on reports from individual users and may not be representative.

A GENERAL APPROACH FOR HARMONIZATION OF A MEASURAND

The specifics of any harmonization effort will vary for different measurands. Fig. 3 presents a general framework for development of measurand-specific protocols. All available measurement procedure comparison data and physicochemical information about the measurand should be compiled to assist in defining the measurand and identifying potential underlying causes for any observed nonequivalency between methods. If suitable data are not available, then an experimental assessment of harmonization status must be performed.

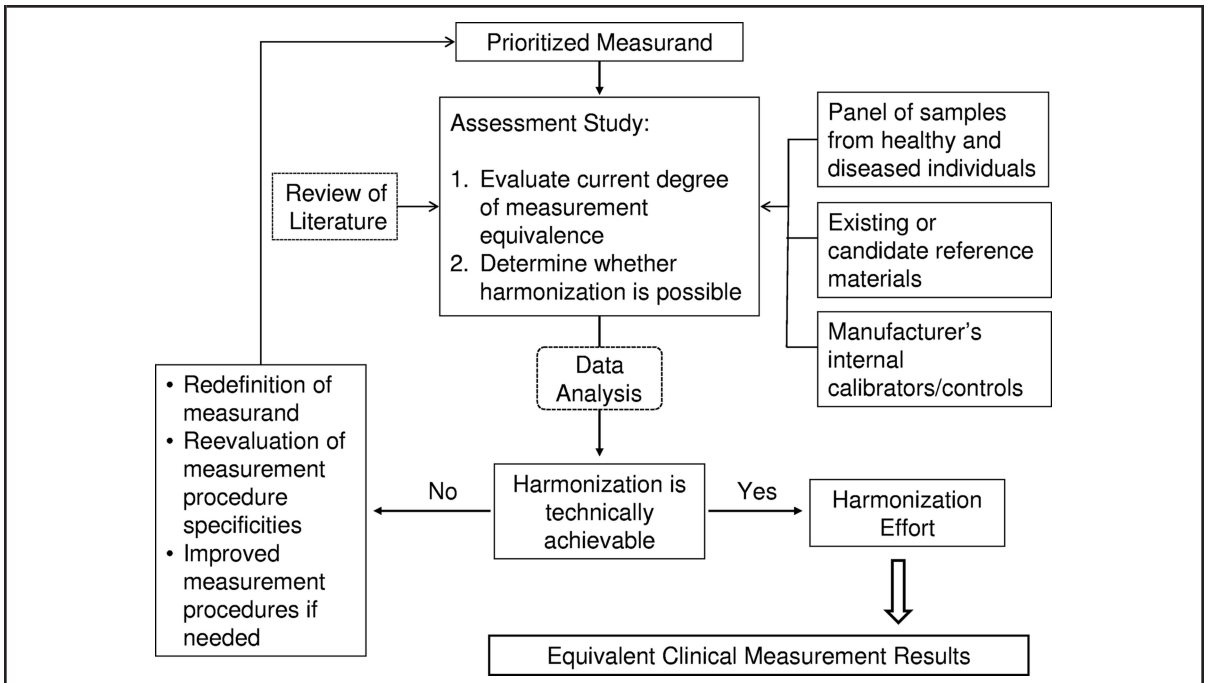

Fig. 3. General process for assessing and achieving harmonization (equivalency) of clinical laboratory measurement results.

An assessment of all CLMPs should be performed by using a panel of individual samples from people who are healthy, those who are diseased, and those who have related pathological conditions that may influence the measurand. The assessment has 2 main goals: to evaluate the current degree of measurement equivalency, and to determine experimentally whether harmonization is feasible. For convenience and efficiency, samples from healthy and diseased individuals that reasonably span the measuring interval should be included in the assessment. Because some measurands are not present in healthy individuals (e.g., drugs) or are present at very different concentrations or with different molecular forms in diseased vs healthy individuals, the appropriate sample panel may require a substantial percentage to be collected from diseased individuals. Inclusion of samples from healthy and diseased persons will also facilitate the identification of any physicochemical differences in the measurand between healthy and diseased conditions. The panel of patient samples would be used in a split-sample comparison involving, ideally, all available CLMPs. This assessment also provides an opportunity to evaluate the commutability of any existing higher-order reference materials for the measurand and to evaluate candidate reference materials to be used in a future harmonization effort. If candidate reference materials are needed for a measurand, it is desirable to prepare them in advance so they can be included in the initial split-sample assessment.

The data from an assessment study should be evaluated by using appropriate statistical approaches such as:

- Pairwise comparisons of the results from each measurement procedure, including evaluation of regression lines and/or difference plots, data scatter around regression lines, and other relevant statistical parameters,
- Comparisons to a trimmed, all–measurement-procedures mean, and
- Comparisons to an arbitrarily selected comparison measurement procedure.

The assessment may demonstrate that attempts to produce equivalency of measurement results will not likely succeed because the measurand is biochemically heterogeneous and various CLMPs are measuring different measurands (e.g., as indicated by excessive scatter and number of aberrant results among CLMPs). This condition may occur because different molecular forms of a measurand exist in different samples from healthy and diseased individuals. In such circumstances, the measurand of clinical interest may require redefinition based on its molecular heterogeneity, the

specificities of the CLMPs may have to be reevaluated in relation to the redefinition of the measurand, or some CLMPs may have to be redesigned before harmonization can be successful. If the assessment results show consistent numeric relationships among the different CLMPs, then a harmonization effort would be technically achievable.

When nonequivalency of measurement results is observed, an investigation of influence quantities (e.g. interfering substances or alternate molecular forms) that occur in clinical samples should be carried out. For measurands that are not sufficiently well characterized, additional experiments may be needed to understand why CLMP results are not equivalent. This investigation would lead to better characterization of the measurand and aid in development of commutable reference materials and possibly a designated comparison measurement procedure or an RMP.

When the assessment indicates that harmonization can be technically achieved, a variety of resources will be required. A panel of human samples will be needed to recalibrate the CLMPs. In most cases, this "calibration" sample panel will be the same as the sample panel used in the assessment study and the data from the assessment panel can be used for recalibration. Thus, it is recommended that the initial experimental design should include collecting data on each manufacturer's in-house calibrators and controls that will be needed for a follow-up recalibration to achieve harmonization. Clinical sample pools or other candidate reference materials should also be included in the assessment protocol to investigate if they have acceptable commutability to be useful in a manufacturer's value-assignment procedures for product calibrators. There may be instances in which interfering substances, or other influence quantities, are suspected to be present in the assessment panel, requiring a refined collection protocol for the calibration panel, and a second split-sample study. Statistically valid processes should be developed to compare data from the analysis of the calibration panel to implement recalibration to achieve equivalency of CLMP results within the agreed clinical requirements *(12)*. Harmonizing the units in which results are reported should also be addressed because differences in units may lead to clinical misinterpretation.

Information from a harmonization project, including the physicochemical characteristics of measurands, clinical data on the individuals used to prepare sample panels that may be important to the understanding of any nonequivalency of results observed in the assessment and harmonization studies, and the split-sample comparison data and its evaluation should be maintained in a publicly available data repository.

SUSTAINING HARMONIZATION

Surveillance of the effectiveness of harmonization is best accomplished through PT/EQA programs that use commutable samples. Criteria for evaluation of the success of harmonization will generally be based on the bias component of the TEa required for clinical use. Manufacturers and clinical laboratories should participate at least annually in a program that uses commutable samples when available. PT/EQA providers should develop programs that use commutable samples whenever possible. The concentrations of these samples should cover a clinically relevant range and focus on clinically significant decision points. Because there are no RMPs for category 4 and 5 measurands, the target values may be established as the all-procedure trimmed mean *(12, 16)*. Acceptance criteria based on the dispersion of results is another option, with reduced dispersion over time being the goal, thereby indicating improved harmonization. Data from PT/EQA schemes that use commutable samples would be strengthened if data from all programs were publicly available to enable the data to be more readily compared. The data should be accompanied by clear guidance regarding commutability of the samples used and how the data should be interpreted, and explanation of any caveats regarding the data.

Commutable PT/EQA materials are frequently prepared from single-donor samples or as pools of patient samples. Use of single donors entails the risk that a particular donor's sample may contain unusual endogenous (e.g., human antimouse antibodies) or exogenous (e.g., drugs that the donor happens to be taking) influence quantities that may impact different CLMP results in variable ways. It is important to recognize that pools may be unacceptable for molecular testing when a single subtype or genotype of a nucleic acid material is required.

Manufacturers need resources to support their internal calibration procedures to maintain traceability to the harmonization program developed for a given measurand. It is likely that one or more commutable reference materials and/or a panel of patient samples will be created as part of the technical implementation of a harmonization program. In many cases, reference materials are needed with well-characterized attributes, including molecular forms of interest either as measurands or as interfering substances, to validate suitability of CLMPs. Continued manufacturing and validation of these types of reference materials is a requirement to sustain a harmonization program.

A structured certification process for manufacturers may be useful to document harmonization. Such programs typically operate similarly to PT/EQA, but are organized by a national or international organization that usually issues a certificate of conformance to

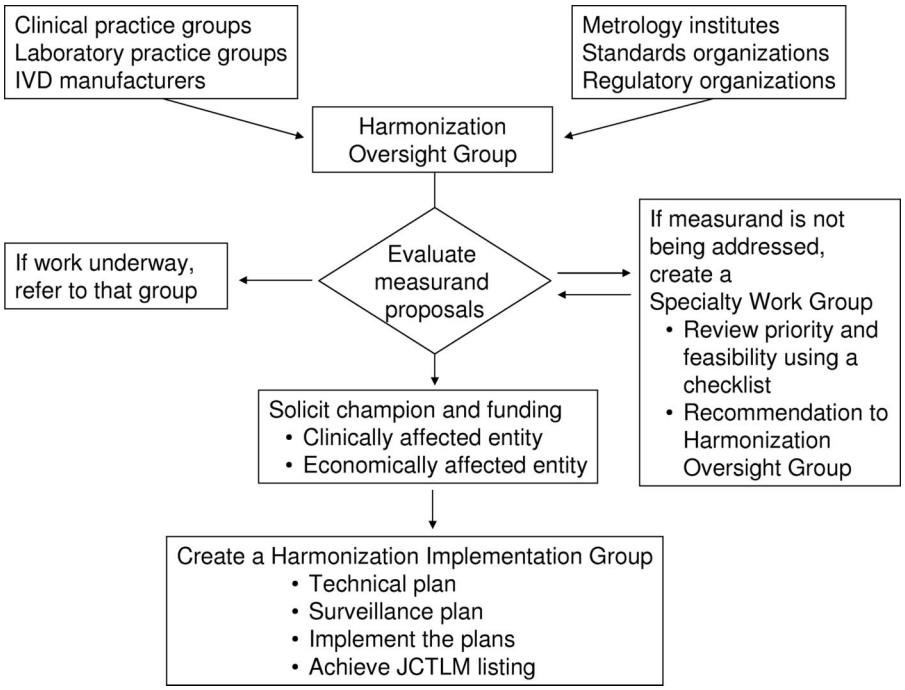

Fig. 4. Overview of a general approach to manage harmonization of a measurand.

established criteria. Self-certification by manufacturers based on use of resources described in the previous paragraph is also possible. In any type of certification program, the data demonstrating CLMP harmonization, or the commutability of a material, must be publicly available for review.

AN INFRASTRUCTURE FOR HARMONIZATION

The complicated processes required to achieve harmonization mandate an infrastructure with well-defined procedures, transparent operation, effective communication with all stakeholders, international involvement, and a consensus approach to cooperation. A proposed process to accomplish harmonization of measurands is diagrammed in Fig. 4. An entity called the Harmonization Oversight Group manages the process including receipt of proposed candidate measurands for consideration, prioritization of measurands, oversight of the implementation of harmonization schemes for different measurands, and communication with all stakeholders. Clinical organizations, laboratory organizations, IVD manufacturers, government or regulatory agencies, journal editors, and research organizations provide input on measurands and other information regarding a need for harmonization. A publicly available repository should include information regarding current and past harmonization efforts.

The Harmonization Oversight Group should develop a prioritized list of measurands to be harmonized. If a group is already addressing a measurand, the work should be referred to that group to avoid duplication of effort. If a measurand is not being addressed, a Specialty Work Group comprised of experts in the clinical use and laboratory measurement of the measurand should be created to evaluate the clinical importance of a measurand, evaluate the gap between clinical requirements and current practice, and assess the technical feasibility to harmonize that measurand.

For high-priority measurands, for which it is technically feasible to achieve harmonization, bids for financial support can be solicited from entities with a clinical or financial interest in a measurand's harmonization. Once funding has been assured, a Harmonization Implementation Group will implement the harmonization activity. The composition of this group should include clinical users, laboratory technical experts, IVD industry experts, a regulatory expert, and a patient representative. The Harmonization Implementation Group would develop the criteria for acceptable agreement among CLMPs, the technical plan, any needed reference materials, and implement procedures to achieve harmonization. This group would also ensure that a process for surveillance of the success of harmonization is developed.

An important objective is to have all new reference materials created and all processes established for harmonization to be documented in the literature and to lead to JCTLM listing when in accordance with the JCTLM quality system and reviewed by the respective JCTLM expert group. JCTLM listing is important for IVD manufacturers to satisfy regulatory requirements and be able to sell IVD products with calibration traceable to a reference material or to a harmonization process *(38)*. Regulatory agencies must be engaged in the harmonization process to address any regulatory hurdles associated with recalibration. IVD manufacturers will face a financial burden associated with recalibration of their CLMPs, and the financial costs associated with harmonization must justify the benefits.

Implementation of a harmonization process will require the involvement of international clinical and medical organizations, IVD manufacturers, clinical laboratories, metrology institutes, standard-setting organizations, and regulatory agencies. Long-term success will depend on collaboration among stakeholders who have a commitment to providing improved patient care and those with the financial resources needed to contribute to the implementation of harmonization for specific measurands. The laboratory and medical communities must be informed about the impact on patient care of poor between-CLMP agreement, the benefit of harmonization, and any change in reported measurand values. Early involvement of clinical opinion leaders to educate practitioners regarding result-interpretation changes is especially important to ensure appropriate implementation of harmonized laboratory results.

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 Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: W.G. Miller, *Clinical Chemistry,* AACC, and CLSI; G.L. Myers, AACC, IFCC Scientific Division, and Diagnostic Marketing Association; S.F. Lo, National Academy of Clinical Biochemistry.
Consultant or Advisory Role: G.L. Myers, NIH Office of Dietary Supplements; S.F. Lo, College of American Pathologists.
Stock Ownership: E.R. Schönbrunner, Life Technologies.
Honoraria: G.L. Myers, NIH Office of Dietary Supplements.
Research Funding: None declared.
Expert Testimony: None declared.
Other Remuneration: Abbott Laboratories, ARUP Laboratories, Beckman Coulter, Beckon, Dickinson and Company, Denka Seiken, Life Technologies, Ortho-Clinical Diagnostics, Quest Diagnostics, SeraCare Life Sciences, and Wako Pure Chemical Industries provided financial support to AACC for the conference.

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.

Acknowledgments: The authors are grateful to NIST, Gaithersburg, MD, for graciously hosting the conference at their facility. We thank Dr. Willie May and Donna Kimball for their assistance in coordinating use of the NIST facilities. We also thank Jean Rhame and Vince Stine at AACC for organizing the conference logistics. We also appreciate thoughtful review of the manuscript by the following conference participants: Chris Burns, National Institute for Biological Standards and Control; Angela Caliendo, Emory University; Michael Flessner, National Kidney Disease Education Program; Neil Greenberg, Ortho Clinical Diagnostics; Alexander Katayev, Laboratory Corporation of America; Jeong-Ho Kim, Korean Society for Laboratory Medicine; William Koch, US Pharmacopeia; Julie Lin, American Society of Nephrology; Curt Rohlfing, National Glycohemoglobin Standardization Program; William Rosner, the Endocrine Society; Heinz Schimmel, Institute for Reference Materials and Measurements; Lothar Siekmann, University of Bonn; and David Wood, WHO.

References

1. Kohn LT, Corrigan JM, Donaldson MS, eds. To err is human: building a safer health system. Washington (DC): National Academy Press; 2000. 287 p.
2. Vesper HW, Thienpont LM. Traceability in laboratory medicine. Clin Chem 2009;55:1067–75.
3. ISO. In vitro diagnostic medical devices — measurement of quantities in biological samples — metrological traceability of values assigned to calibrators and control materials. 1st ed. ISO 17511:2003(E). Geneva: ISO; 2003.
4. Armbruster D, Miller RR. The Joint Committee for Traceability in Laboratory Medicine (JCTLM): a global approach to promote the standardisation of clinical laboratory test results. Clin Biochem Rev 2007;28:105–13.
5. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. Clin Biochem 2009;42:236–40.
6. Ekins R. Immunoassay standardization. Scand J Clin Lab Invest 1991;51(Suppl 205):33–46.
7. Thienpont LM. Accuracy in clinical chemistry: who will kiss Sleeping Beauty awake? Clin Chem Lab Med 2008;46:1220–2.
8. Zegers I, Keller T, Schreiber W, Sheldon J, Albertini R, Blirup-Jensen S, et al. Characterization of the new serum protein reference material ERM-DA470k/IFCC: value assignment by immunoassay. Clin Chem 2010;56:1880–8.
9. Little RR, Rohlfing CL, Tennill AL, Madsen RW, Polonsky KS, Myers GL, et al. Standardization of C-peptide measurements. Clin Chem 2008;54:1023–6.
10. Stephan C, Klaas M, Müller C, Schnorr D, Loening SA, Jung K. Interchangeability of measurements of total and free prostate-specific antigen in serum with 5 frequently used assay combinations: an update. Clin Chem 2006;52:59–64.
11. Miller WG, Thienpont LM, Van Uytfanghe K, Clark PM, Lindstedt P, Nilsson G, Steffes MW. Toward standardization of insulin immunoassays. Clin Chem 2009;55(5)1011–18.
12. Thienpont LM, Van Uytfanghe K, Beastall G, Faix JD, Ieiri T, Miller WG, et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests; part 1: thyroid-stimulating hormone. Clin Chem 2010;56:902–11.
13. Rose MP. Follicle stimulating hormone international standards and reference preparations for the calibration of immunoassays and bioassays. Clin Chim Acta 1998;273:103–17.
14. Tate JR, Bunk DM, Christenson RH, Katrukha A, Noble JE, Porter RA, et al. Standardisation of cardiac troponin I measurement: past and present. Pathology 2010;42:402–8.
15. Sturgeon CM, Berger P, Bidart JM, Birken S, Burns C, Norman RJ, et al. Differences in recognition of the 1st WHO international reference reagents for hCG-related isoforms by diagnostic immunoassays for human chorionic gonadotropin. Clin Chem 2009;55:1484–91.
16. Thienpont LM, Van Houcke SK. Traceability to a common standard for protein measurements by

immunoassay for in-vitro diagnostic purposes. Clin Chim Acta 2010;411:2058–61.
17. Schimmel H, Zegers I, Emons H. Standardization of protein biomarker measurements: is it feasible? Scand J Clin Lab Invest 2010;70(Suppl 242): 27–33.
18. Radin N. What is a standard. Clin Chem 1967;13: 55–76.
19. Miller WG, Myers GL, Rej R. Why commutability matters. Clin Chem 2006;52:553–4.
20. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. Clin Biochem Rev 2007;28:139–47.
21. Caliendo AM, Shahbazian MD, Schaper C, Ingersoll J, Abdul-Ali D, Boonyaratanakornkit J, et al. A commutable cytomegalovirus calibrator is required to improve the agreement of viral load values between laboratories. Clin Chem 2009;55: 1701–10.
22. CLSI. Characterization and qualification of commutable reference materials for laboratory medicine: approved guideline. CLSI document C53-A. Wayne, (PA): CLSI; 2010.
23. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem 1998;31:449–57.
24. Thienpont LM, Stockl D, Friedecky B, Kratochvila J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of the samples. Scand J Clin Lab Invest 2003;63:195–202.
25. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta 2003;327:25–37.
26. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003;326: 41–4.
27. Roddam AW, Price CP, Allen NE, Ward AM. Assessing the clinical impact of prostate-specific antigen assay variability and nonequimolarity: a simulation study based on the population of the United Kingdom. Clin Chem 2004;50:1012–6.
28. NIH consensus statement on management of hepatitis C: 2002. NIH Consens State Sci Statements. 2002;19(3):1–46.
29. Petersen PH, Sandberg S, Iglesias N, Sölétormos G, Aarsand AK, Brandslund I, Jørgensen LG. 'Likelihood-ratio' and 'odds' applied to monitoring of patients as a supplement to 'reference change value' (RCV). Clin Chem Lab Med 2007; 46:157–64.
30. Karon BS, Boyd JC, Klee GG. Glucose meter performance criteria for tight glycemic control estimated by simulation modeling. Clin Chem 2010; 56:1091–7.
31. Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine. Consensus agreement. Scand J Clin Lab Invest 1999; 59:475–585.
32. Fraser CG. Biological variation: from principles to practice. Washington (DC): AACC Press; 2001.
33. Braga F, Dolci A, Mosca A, Panteghini M. Biological variability of glycated hemoglobin. Clin Chim Acta 2010;411:1606–10.
34. Miller WG, Bruns DE, Hortin GL, Sandberg S, Aakre KM, McQueen MJ, et al. Current issues in measurement and reporting of urinary albumin excretion. Clin Chem 2009;55:24–38.
35. Miller WG, Myers GL, Ashwood ER, Killeen AK, Wang E, Ehlers GW, et al. State of the art in trueness and inter-laboratory harmonization for 10 analytes in general clinical chemistry. Arch Pathol Lab Med 2008;132:838–46.
36. Palmer-Toy DE. Comparison of pooled fresh frozen serum to proficiency testing material in College of American Pathologists surveys: cortisol and immunoglobulin E. Arch Pathol Lab Med 2005;129:305–9.
37. Schreiber WE. Comparison of fresh frozen serum to proficiency testing material in College of American Pathologists Surveys: α-fetoprotein, carcinoembryonic antigen, human chorionic gonadotropin, and prostate-specific antigen. Arch Pathol Lab Med 2005;129:331–7.
38. Blirup-Jensen S. Protein standardization III: method optimization. Basic principles for quantitative determination of human serum proteins on automated instruments based on turbidimetry or nephelometry. Clin Chem Lab Med 2001;39: 1098–109.