Metrological Traceability of Assays and Comparability of Patient Test Results

David Armbruster, PhD, DABCC

KEYWORDS

• Metrology • Traceability • Standardization • Harmonization • Comparability

KEY POINTS

- As of 2003, metrological traceability of assay calibrators has been a regulatory requirement and necessary to ensure accuracy and comparability of patient test results.
- Calibrator traceability and comparability of test results from different assays are necessary for the use of electronic health records and optimal patient care.
- Calibrator traceability is one significant aspect of the standardization of clinical laboratory practice, which includes standardization of other facets, including reporting units, test nomenclature, and evidence-based laboratory medicine guidelines.

INTRODUCTION

The clinical laboratory field is experiencing globalization. Laboratory practice is moving toward harmonization and the ability to produce comparable patient test results. Greenberg observed, "An increasingly important objective in laboratory medicine is ensuring the equivalency of test results among different measurement procedures, different laboratories and health care systems, over time."[1,2] Metrological traceability is required to provide equivalence of results from diverse analytical systems.[3] Laboratories no longer work in isolation, and harmonization of laboratory testing is far-reaching, including all aspects of the total testing process (TTP).[4] The goal is "Right result, Right patient, Right time, Right form, Right test choice, Right interpretation, and Right advice." Test results must be equivalent to use universal clinical guidelines for disease diagnosis and patient management. Impediments to harmonization include inadequate measurand (analyte) definition, lack of analytical specificity, noncommutability of reference materials, lot-to-lot variability of reference materials and

Conflict of Interest: D. Armbruster is an employee of Abbott Diagnostics.
Clinical Chemistry, Abbott Diagnostics, Department 09AC, Building CP1-5, 100 Abbott Park Road, Abbott Park, IL 60064, USA
E-mail address: david.armbruster@abbott.com

Clin Lab Med 37 (2017) 119–135
http://dx.doi.org/10.1016/j.cll.2016.09.010 **labmed.theclinics.com**

assay reagents, and a lack of systematic approaches to standardization. These issues affect patient care because physicians fail to understand the limitations of laboratory measurements, including the lack of interchangeability of results from different analytical methods.[5]

Generating comparable results remains a holy grail due to use of multiple assays for the same analyte, potentially causing different clinical interpretations.[5,6] Clinical decision values (cutpoints) are decided by international expert groups without consideration of analytical disparity. Even advances in technology are not always an improvement. As noted by White, "frustration at the lack of significant progress…was captured in the title 'Accuracy in Clinical Chemistry — Does anybody care?', in which Tietz identified that the accuracy of many routine laboratory methods had declined as use of faster, automated methods and instrumentation increased. Since Tietz's *cri de coeur*, there has been significant progress with both the theory and the practice of implementing a coherent reference system for measurements in clinical laboratories."[6] Harmonization was not possible historically due to a lack of established reference materials and methods. Miller and Myers[7] noted, "True and precise routine measurements of quantities of clinical interest are essential if results are to be optimally interpreted for patient care. Additionally, results produced by different measurement procedures for the same measureand must be comparable if common diagnostic decision values and clinical research values are to be broadly applied."

A patient's test history would be consistent if only one laboratory performed all testing (same methodology, analyzer, and so forth), so a significant change in concentration would signal a meaningful clinical change. But patients are increasingly mobile and multiple laboratories may test their samples so results may not be consistently interpreted.[8] Harmonization can produce essentially equivalent results (not quantitatively equal but clinically equivalent) and changes in concentration can be correctly interpreted.[9] Harmonization needs to include nomenclature, units of measurement, and other factors for use of evidence-based clinical practice guidelines.[8,10] Physicians expect results to be interchangeable even though analytes can be measured by multiple methods. Many clinicians do not realize tests performed by one method cannot be reliably compared with those from another method. This lack of comparability creates barriers to sharing laboratory results across health care systems and can have adverse patient consequences.[11] For some analytes, reference materials do not exist or there is a limited supply, and new lots may not be identical to the original material.[10] It is even difficult to know which molecule is actually being measured given structural variability, for example, the various forms of human chorionic gonadotropin (HCG).

Lack of harmonization has real adverse clinical consequences, and prostate-specific antigen (PSA) is a prime example.[6,12–14] An early PSA assay (Hybritech, San Diego, CA) used the manufacturer's calibrator, and the standard 4.0 mg/L PSA cutoff for prostate cancer was established. Other assays use calibrators traceable to World Health Organization (WHO) international reference material (WHO 96/670 and 96/668). A 2004 study of 2304 patients compared PSA results from assays using the Hybritech or the WHO calibrator. Of 288 patients, 55 (19%) exceeded the PSA 4.0-mg/L cutoff based on the Hybritech calibrator result but were not candidates for prostate biopsy by the WHO-calibrated results. In another PSA study, 106 men were tested using both the Hybritech and WHO traceable calibrators and WHO calibrator results were 20% lower. Depending on the assay, some men are candidates for prostate biopsy (a definitely invasive procedure) and others are not. Many clinicians are unaware, however, that different PSA results are produced for the same patient sample if tested by assays using different calibrators, resulting in different clinical interpretation and adverse patient consequences. Lack of comparability is a concern for immunoassays,

such as thyroid and fertility hormones and cancer markers. Traceability/standardization of immunoassays is a special problem because internationally accepted clinical protocols and common reference intervals (RIs) depend on it, and "free" hormones and heterogeneous polypeptide hormones are difficult.[15]

Cholesterol is a prime example of successful harmonization. A cholesterol reference measurement system (RMS) was created over approximately 30 years (1970–2000) and produced a major reduction in mortality rates for coronary heart disease in the United States, achieving a huge savings in health care dollars.[1] Consequences of the lack of harmonization were detailed in a National Institute of Standards and Technology (NIST) report on calcium (Ca) that estimated the cost of a 0.1-mg/dL Ca bias can mean an additional $8 to $31 cost for unnecessary patient follow-up testing.[16] A Ca bias of 0.5 mg/dL could result in an additional $34/patient to $89/patient, and on an annual basis, a 0.1-mg/dL bias could translate into $75 million/y to $250 million/y (adjusted for 2016 dollars) for approximately 3.55 million patients screened for Ca.

EHRs contain a wealth of laboratory data on patients but the benefit is negated if the values for the same analyte are not comparable. It has been suggested that laboratory data account for approximately 70% of clinical decisions. Hallworth[17] has challenged that blanket statement but allows, "The value of laboratory medicine in patient care is unquestioned. That value is greatly diminished without comparability of test results."

HARMONIZATION AND STANDARDIZATION

Assay harmonization is a high priority but so is harmonization of terminology, reporting units, and even the pre-preanalytical and the post-postanalytical phase.[1] Clinicians expect to receive the "right test at the right time for the right patient" and also assume the "same results and interpretation for a sample irrespective of the laboratory that produced the result."[18]

In this discussion, *harmonization* is used interchangeably with *standardization*, although there is a distinction between the two.[9] Standardization means results are traceable to higher metrological order reference materials and/or methods and ideally can be reported in International System of Units (SI units). Harmonization means results are traceable to some declared reference but higher-order reference materials and/or methods are not available and SI units are not applicable. Harmonization ensures comparability of results, enables application of clinical best practice guidelines and RIs, increases patient safety, and decreases medical care costs. To achieve it requires the cooperation of laboratories, academia, professional societies, metrological institutes, government agencies, external quality assessment/proficiency testing (EQA/PT) providers, and industry.

Two standardization success stories are creatinine and glycated hemoglobin (hemoglobin A_{1c} [HbA_{1c}]).[3] Field assays for both analytes have complete traceability chains, firmly anchored by RMSs. In one HbA_{1c} study, reference samples with target values assigned by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) reference method (European Reference Laboratory for Glycohemoglobin) were tested using an enzymatic method field assay.[19] The maximum systematic bias for the commercial assay was only 1.9%, and the mean bias ranged from −1.1 to 1.0 mmol/mol. Ironically, results for creatinine and HbA_{1c} assays are still reported in different units, creatinine in mg/dL (conventional units) and μmol/L (SI units), and HbA_{1c} in %HbA_{1c} (NGSP units) and mmol/mol (SI units). Beyond analytical traceability is the challenge of harmonization of the TTP, including use of identical reporting units. Ideally, a certification process for in vitro diagnostics (IVD) manufacturers could

document and ensure harmonization.[3] It would likely be organized by national or international bodies, and stakeholders would include clinical and laboratory organizations, IVD manufacturers, government and regulatory agencies, journal editors, research organizations, metrology institutes, and standard setting organizations.

METROLOGICAL TRACEABILITY

As White[6] explains, "Metrology, the science of measurement, provides laboratory medicine with a structured approach to the development and terminology of reference measurement systems which, when implemented, improve the accuracy and comparability of patients' results." Metrological principles are relatively new in the clinical laboratory. The third edition of the *Tietz Textbook of Clinical Chemistry* made no mention of the metrology terms, *uncertainty* and *commutability*.[20] The fourth edition mentioned *uncertainty* and gave a definition of *commutability*.[21] The fifth edition includes a discussion of *uncertainty* along with *commutability*.[22] But as noted by De Bievre,[23] "Discussions with analytical chemists have revealed that basic concepts in metrology, including 'traceability' are generally not an integral part of university or college curricula and are not treated in most textbooks of analytical chemistry."

Full implementation of the IVD Directive (IVDD) (December, 2003) under European law requires calibration of quantitative IVD assays be traceable to available higher-order reference methods or materials.[24] The IVDD applies to Europe for the purposes of the Conformité Européenne [European Conformity] (CE) mark but has global implications. Manufacturers must establish metrological traceability for calibrators and controls and the uncertainty of kit calibrators. Assays have always been anchored by some kind of standards, but strict metrological traceability was not always in place or even necessarily appreciated. Powers[25] attributes the IVDD in part to European laboratory professionals striving for result accuracy and patient test result transferability, pointing to International Organization for Standardization (ISO) 17511 in which is found this statement: "It is essential that results reported to physicians and patients are adequately accurate (true and precise) to allow correct medical interpretation and comparability over time and space."[25] Metrological traceability satisfies basic clinical needs and improves patient care but the details for implementation of the process continue to be a challenge.[26]

The IVD field routinely performs measurements on an estimated 400, 600, or even 1000 different analytes, but full calibration systems with acceptable traceability currently exist for perhaps only 30 to 100 analytes.[24,25] Benefits for industry from traceability include interchangeability of data between products, competitiveness (levels the playing field), defined quality goals, lower long-term costs, clearer pathway to market access, transferable technology, and independent tools to ensure long-term performance stability. Trade-offs include diverting qualified people to participate in standards work versus other programs, risk of investing in standards not acceptable to all stakeholders, lengthy cycle time to achieve deliverables, costs of transition to new standards, less variety and fewer alternatives for customers, and barriers to innovation and market entry.[24]

Traceability originated in the metrological community and was first defined in 1993 in the International Vocabulary of Metrology — Basic and General Concepts and Associated Terms (VIM).[26] The VIM definition is the "property of a measurement result whereby the result can be related to a stated reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty."[1,26] **Fig. 1** illustrates the hierarchical order of materials and measurement procedures for an unbroken traceability chain. It follows an alternating process of assigning target values to materials

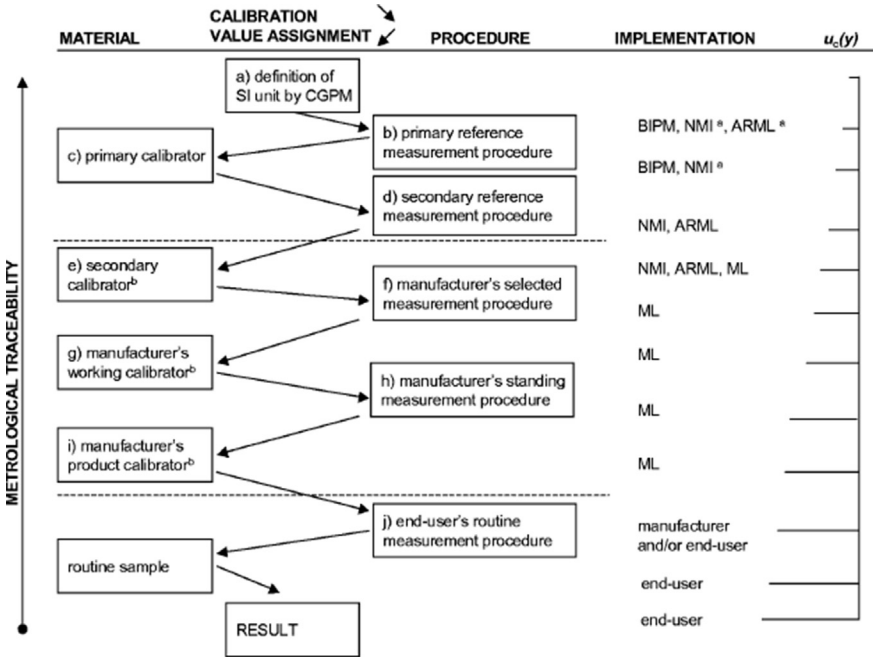

Fig. 1. General metrological traceability diagram. [a] ARML, accredited reference measurement laboratory (such a laboratory may be an independent or a manufacturer's laboratory); [b] SI unit of measurement; CGPM, General Conference on Weights and Measures; ML, manufacturer's laboratory; $u_c(y)$, combined standard uncertainty of measurement. (*From* ISO 17511. In vitro diagnostic medical devices—Measurement of quantities in biological samples—Metrological traceability of samples assigned to calibrators and control materials. International Organization for Standardization; 2003; with permission.)

used to calibrate the next lower-order measurement procedure.[26] Traceability is established using the ISO standards 17511, 15193, 15194, and 15195.[13,16,27]

The traceability to internationally recognized and accepted reference materials and measurements is the key element assuring accuracy and comparability of results.[28]

Diagnostic manufacturers must ensure analytical systems are traceable to certified reference materials and measurement procedures and that calibrator uncertainty is documented.[13] Analytes are either type A (well-defined chemical entities, traceable to SI units) or type B (heterogenous analytes in human samples and that are not directly traceable to SI units). Type A analytes represent a small number of well-defined compounds (approximately 65) belonging to classical clinical chemistry, for example, electrolytes, minerals, cholesterol, creatinine, steroid hormones, and vitamins. Type B analytes are all the rest, including most of the proteins (usually measured immunochemical methods) and more esoteric compounds whose results are expressed in terms of arbitrary units, for example, WHO international units or mass units.[13] The esoteric type B analytes measured by immunoassays are more challenging to standardize due to use of different calibrators by manufacturers because of internationally recognized reference material/measurement procedures not available, comparison of assays to different predicate devices, use of antibodies recognizing different antigens/epitopes on the same analyte, and use of different capture/detection antibodies in 2-step immunoassays for the same analyte.[14]

The IVDD requirements are incorporated in ISO 15189 (medical laboratories—particular requirements for quality and competence), the basis for many laboratory accreditation programs.[28] Metrologists are principally interested in accurate measurements and when ISO standards were drafted, committee debates about metrological principles and terminology, interesting on a philosophic level, often overshadowed concerns for the intended clinical use of assays, explaining the academic tone of the standards.[28] The standards meant a major shift for manufacturers away from using in-house materials and methods to the use of reference materials and methods vetted by metrology. Ideally, SI reporting units are used.[6] Broad implementation of SI units has facilitated scientific exchange and the Bureau international des poids et mesures [BIPM] provides a coherent system of measurements traceable to the SI, ensuring equivalence of measurements, including those used in laboratory medicine.[6]

Metrology theory introduces complications because a measurement result is an estimate of the true value of the measurand and, because the true value cannot be exactly known, the concept of measurement uncertainty (MU) was developed.[5,6] MU assumes significant bias in the reference material is eliminated and calculates an interval of values for the measurand (analyte) within which the true value lies with a stated level of confidence. Metrology defines MU as a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.[6] Clinical and Laboratory Standards Institute (CLSI) EP29 (expression of MU in laboratory medicine) is a recent guideline for estimating uncertainty.[29] "The GUM [Guide to the Expression of Uncertainty in Measurement] approach to uncertainty is rapidly gaining acceptance in metrological institutes and industry, and must be applied in ISO (International Organization for Standardization) and CEN (European Committee for Standardization) standards. It should be used in accredited laboratory work but chemists often find the implementation difficult and therefore hesitate. Additionally, sometimes, there is a fear that honest GUM uncertainty intervals, which may be wider than classical precision intervals, are bad for business."[30]

Metrology must be adapted to clinical laboratory science for harmonization but in a practical manner due to the differences between the disciplines. Metrology is a pure science in contrast to the mixed science of clinical chemistry that combines diverse disciplines and technologies. National metrology institutes (NMIs) are ivory towers in comparison to clinical laboratories, which are more like the trenches. Metrologists test pure, well-defined analytes in simple matrices whereas clinical laboratorians test complex, ill-defined analytes in challenging matrices (serum, plasma, urine, and so forth). Metrology uses expanded uncertainty (with bias eliminated) to set accuracy goals whereas clinical laboratories tend to use a total error allowable (TEa) methodology (TEa = bias + imprecision). Metrology seeks "absolute scientific truth" by reference method analysis but clinical laboratories must deal with relative truth by field method analysis. Good metrology does not necessarily equal good clinical laboratory science, but clinical laboratories need to adapt metrological concepts and adapt them for practical application.

THE PILLARS OF HARMONIZATION

The Treaty of the Meter (1875) enabled comparability of measurements, and metrology is the science of measurement.[1] Metrology is a separate science in its own right but its concepts are relevant to many other disciplines, including clinical laboratory science.[31] The IVDD calls for traceability of calibrators to higher-order reference materials and/or reference methods. But *higher-order* was not defined by the legislation beyond assigning responsibility for traceability to national notified bodies. The premise is

manufacturers will be responsible for traceability, but manufacturers need to know which reference materials and methods can anchor assays.[14] The IVDD made it necessary to identify a final arbiter of traceability. Stenman noted that many organizations deal with standardization but it is not clear who is responsible for what and it is desirable that one international organization manage standardization.[14]

In anticipation of the IVDD, the Joint Committee for Traceability in Laboratory Medicine (JCTLM) was formed in 2002 (http://www.bipm.org/en/committees/jc/jctlm).[1,2,13] The JCTLM is an international consortium (government, clinical laboratory profession, and industry), sponsored by the BIPM, the IFCC, and the International Laboratory Accreditation Cooperation. Its mission is to support worldwide comparability, reliability, and equivalence of measurement results to improve health care.[32,33] The JCTLM established 3 pillars of traceability: (1) reference measurement procedures (RMPs), (2) reference materials, and (3) a network of reference measurement laboratories. The JCTLM maintains a searchable database for all 3 pillars on the BIPM Web site.[13,14,34] JCTLM Working Group 1 (reference materials and reference methods), and JCTLM Working Group 2 (reference measurement services) and their review teams judge database submissions for blood cell counting, coagulation factors, drugs, metabolites and substrates, microbial serology, nonelectrolye metals, nonpeptide hormones, nucleic acids, proteins, vitamins and micronutrients, electrolyte and blood gases, and enzymes.[1] The components of an RMS are definition of the analyte, RMPs that specifically measure the analyte, primary and secondary reference materials, and reference measurement laboratories. Analytes fall into 2 categories: type A (well defined, concentration in SI units, results not method dependent, and full traceability chain) and type B (not well defined, heterogeneous, present in both bound and free state, not traceable to SI, and rigorous traceability chain not available).

Primary and matrix-based secondary references are both needed.[35] Secondary reference materials (pooled human serum, plasma, and urine) are critical for IVD manufacturers to anchor calibrators. RMPs, such as isotope dilution mass spectrometry, are developed by metrology institutes but for analytes, such as enzymes, standardization is only possible by method-specific protocols. Metrology institutes do not have sufficient medical/clinical expertise to set ideal specifications for secondary reference materials and must develop them in close collaboration with laboratory experts (eg, American Association for Clinical Chemistry [AACC] and IFCC). Even then, success is not guaranteed. Studies performed using human pooled serum spiked with NIST SRM (Standard Reference Material) 2921 (human cardiac troponin complex) demonstrated this material does not behave like individual patient samples with elevated troponin I, probably due to lack of commutability.[35]

Even defining a measurand (analyte) is difficult.[6] Not all are well characterized with a known molecular structure and weight (eg, glucose or sodium). Complex molecules, such as proteins, may be structurally heterogeneous due to post-translational modification, glycosylation, complex formation, and so forth. A prime example is HCG with 7 significant isoforms. Relative concentrations of these isoforms can differ markedly depending on the clinical condition.

A special requirement for harmonization is commutability. Rej and colleagues[36] introduced the term in 1973 to designate the property that calibrators and controls should exhibit (ie, analytical response indistinguishable from that of authentic patient samples). Noncommutability was particularly noticeable for enzymes because of sample differences, for example, nonhuman animal sources, isoenzymes, stability, matrix differences, and effects of sample preparation procedures (eg, lyophilization and preservatives).[37] Other variables are differences in substrates, cofactors, pH, and reaction temperatures used by assays.

Commutability is equivalence of the mathematical relationships between the results of different measurement procedures for a reference material and for representative samples from healthy and diseased individuals.[9,38,39] Without commutability, results from routine methods cannot legitimately be compared to identify a calibration bias, and the reference material cannot be used as a calibrator without commutability, so traceability to the reference system is invalid. Fresh patient samples and calibrators need to provide an identical analytical response (Fig. 2). Many secondary reference materials are not commutable and have failed to achieve harmonized results. Noncommutable EQA/PT samples require peer group grading because they are not amenable to accuracy-based grading, that is, comparison of results reported by laboratories to a target value determined by reference method analysis. Commutability is not a universal property of reference materials and must be proved with every field method. Well recognized by metrology, commutability was not widely appreciated by clinical laboratories and commutability of calibrators was not routinely established. Noncommutability results in biases with field assays due to matrix effects, use of nonhuman forms of analyte, lack of antibody specificity, or other causes. Producing sufficiently large pools of commutable material for EQA/PT samples is a practical difficulty because of the large volume required.[38] JCTLM now requires a commutability assessment before listing a reference material in its database. CLSI EP30 (characterization and qualification of commutable reference materials for laboratory medicine) is a recent guideline providing commutability guidance.[40] Commutability of each calibrator in a calibration hierarchy is essential for traceability.[6] Noncommutability breaks traceability. During manufacture, secondary calibrators may suffer matrix modifications due to lyophilization, freeze-thawing, filtration, and so forth, and commutability may be lost.[6]

The JCTLM faces several challenges. Like many similar professional organizations, it depends on volunteers and their expertise. Many laboratory professionals and organizations (metrology institutes, governmental regulatory agencies, manufacturers, and so forth) are actively engaged in JCTLM activities and support involvement. But JCTLM participation is an "extracurricular activity" for the volunteers and not part of the day job or a top priority for employers. It is difficult for stakeholders to allocate human and other

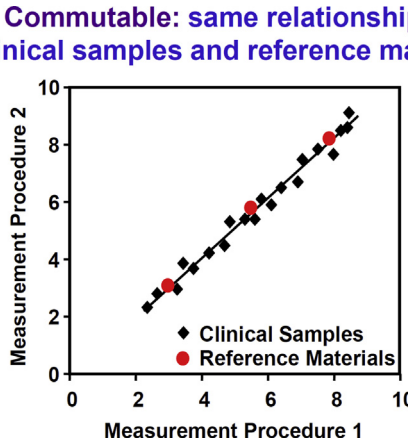

Fig. 2. Commutability is demonstrated if fresh patient samples and reference materials, for example, calibrators, demonstrate an equivalent analytical response when tested by 2 methods.

resources to support the JCTLM.[14] NMIs providing reference materials and methods may have limited resources to maintain the necessary metrological infrastructure. The JCTLM operates by consensus and obtaining consensus among the members may require considerable discussion. After all, if agreement on internationally accepted reference materials and methods were simple, there would have been no need to form the JCTLM in the first place. There is not a fixed JCTLM budget and the participating organizations are expected to fund activities on a pay-as-you-go basis."[14]

In addition to the 3 pillars defined by the JCTLM, the laboratory community has identified 3 more: universal RIs and medical decision levels, EQA/PT programs to ensure traceability of field assays is maintained (eg, accuracy-based grading programs, such as the College of American Pathologists [CAP] requirement of $\pm$ 6% of the target value for HbA_{1c}), and harmonization of clinical laboratory practice and the TTP, for example, standardized nomenclature/terminology, reporting units, and evidence-based laboratory medicine.

The fourth pillar—universal RIs—cannot be erected without the adoption of RMSs and assay harmonization. RIs for some analytes are affected by various partitioning factors, for example, age, gender, ethnicity, and body mass index; thus, universal ranges may not be feasible. But such decisions cannot be made until harmonization has been achieved and study results are comparable.

The requirement for result trueness and comparability requires a fifth pillar: validation of manufacturers' traceability by EQA/PT. EQA/PT surveys were originally educational exercises to compare laboratories to their peers. Although still useful for this purpose, PT surveys now also serve a regulatory purpose.[38] Target values for PT samples should be determined using reference methods and materials because peer group comparison leaves open the question of the absolute accuracy because there can be multiple true values, each peer group mean value representing a relative true value. Comparison to the true value as determined by an RMP allows both an absolute and relative performance yardstick.[14]

Regulatory programs may have wider acceptance limits because it is undesirable for too many participant laboratories to fail challenges. Passing EQA/PT surveys means a laboratory meets some minimum regulatory requirement but does not guarantee clinically acceptable and desired performance. EQA/PT is a one–time point assessment subject to random error, and performance can vary from one survey cycle to another. EQA/PT programs using commutable samples with reference method target values allow accuracy-based grading.[41] Ideally, PT/EQA surveys should be sent to laboratories as blind samples indistinguishable from actual patient samples so they are handled as patient specimens, for example, survey samples should be analyzed only once.[38] Blind testing is a challenge. Horowitz[41] notes, "Far too many laboratories consider proficiency testing just a necessary evil, little more than periodic pass–fail exercises we perform solely to meet regulatory requirements" and "Even for central-laboratory techniques, traditional PT suffers from 'matrix effects,' in that samples used for testing often react differently from native patient samples. Therefore, comparisons must be made only to peer groups, rather than to the 'true value.' What if the peer group as a whole is wrong?"[41] EQA/PT has typically been used to measure proficiency at performing a test and not the trueness of the test method or its performance relative to other methods.[10] Miller and colleagues[42] conclude, "Traditional PT materials are not suitable for field-based postmarketing assessments of a method's trueness."

Collection and pooling of unaltered samples to prepare EQA/PT aliquots, stored frozen ($\leq$70), is considered the best method for preparing commutable samples.[38] Target values should be assigned using reference methods. In the absence of an

RMS, all-participant means or median values may be used as the target value. Because EQA/PT can be driven by either regulatory, clinical decision considerations or by biological variability goals, passing EQA/PT challenges may indicate a laboratory meets minimum standards but it does not necessarily guarantee clinically acceptable performance. In a well-designed study using commutable samples with reference method target values for 10 analytes, glucose, iron, potassium, and uric acid methods exhibited the best performance, with all peer groups meeting the minimum and more than 87.5% of peer groups meeting the desirable, biological variability bias goals.[42] But the other 6 analytes did not meet bias goals. Accuracy-based EQA/PT is ideal but more demanding than peer group grading. EQA/PT results reflect laboratory performance at a given point in time and continuous participation in EQA/PT is necessary to ensure continual acceptable performance.

Interlaboratory comparability of EQA/PT results allows evaluation of calibration traceability.[43] In one study, commutable serum-based material assigned target values by reference methods for 6 enzymes (alanine aminotransferase [ALT], aspartate aminotransferase [AST], creatine kinase [CK], γ-glutamyltransferase [GGT], lactate dehydrogenase [LD], and amylase) was tested by 70 laboratories using 6 field methods.[44] Results were graded on accuracy using biological variability targets. For ALT, results were deemed acceptable for greater than 94% of 6 commercial assays. Performance for the other 5 enzymes was variable and all methods demonstrated significant bias for CK. It was concluded method bias should be reduced by improved traceability to internationally accepted reference systems. Tacrolimus is measured by liquid chromatography–mass spectrometry (LC-MS) and immunoassays and all methods are calibrated without traceability to a recognized reference method or material.[45] Tacrolimus results thus may not be comparable between methods, with potential risks to cancer patients monitored by therapeutic drug monitoring. A global comparability study conducted to assess analytical variability found an immunoassay (ARCHITECT, Abbott, Chicago, IL, USA) demonstrated the best precision (coefficients of variation [CVs] of 3.9%–9.5%) whereas CVs for another immunoassay (Dade Dimension, Dade, Glasgow, DE, USA) and LC-MS methods ranged from 5% to 48.1% and 11.4% to 18.7%, respectively. Higher LC-MS imprecision was primarily due to between-laboratory variability. An advantage of the commercial immunoassays is they use the same extraction procedure, instrumentation, detection systems, and calibration, whereas the LC-MS method parameters varied. Even the use of a common calibrator did not harmonize the LC-MS results. No LC-MS tacrolimus method has yet been listed in the JCTLM database, and thus none is recognized as defining analytical truth for the analyte.

The sixth harmonization pillar is the TTP. Plebani[46] observed, “Although the focus is mainly on the standardization of measurement procedures, the scope of harmonization goes beyond method and analytical results: it includes all other aspects of laboratory testing, including terminology and units, report formats, reference intervals and decision limits, as well as test profiles and criteria for the interpretation of results.” Harmonization of reporting units seems easy, but is it not. A UK survey revealed 80% of laboratories reported hemoglobin using grams per deciliter although grams per liter is the recommended unit.[46] Harmonization of basic terminology and units is necessary but the international laboratory community has yet to reach agreement. See **Table 1** for examples of disharmony.

CHALLENGES TO HARMONIZATION

Thienpont[47] has lamented that major IVD manufacturers have not agreed to a new measurement paradigm in clinical chemistry and moved to accuracy-based assays,

Table 1
Examples of disharmony

Analyte	Conventional Units	SI Units
ALT	U/L	mkat/L
Bilirubin	mg/dL	mmol/L
Cl	mEq/L	mmol/L
Glucose	mg/dL	mmol/L
Creatinine	mg/dL	mmol/L
HbA_{1c}	% Hb A_{1c}	mmol/mol

demonstrating transparency by comparing assays with accepted RMPs. Embracing metrological harmonization is a paradigm shift for the IVD industry. Manufacturers traditionally sought to differentiate themselves from competitors (eg, claiming a greater dynamic range, lower limit of detection, better precision, smaller sample size, and so forth) and producing comparable patient results was not a priority. Lack of harmonization among field assays is evident from EQA/PT data, often requiring peer group reporting. Manufacturers are responding to the need for comparability by providing calibrator traceability/uncertainty information, restandardizing assays, establishing commutability, and so forth. Manufacturers play an integral role in educating laboratories about assay harmonization and modern clinical laboratory practice in general. But the old question remains, "Where do manufacturers' obligations end and the obligations of laboratory directors begin?" Manufacturers must provide fit-for-purpose tests, but laboratories must use the assays properly and effectively. When an assay failure occurs (and failure can apply to myriad issues and causes), does the fault lie with the manufacturer or with the laboratory and its use of the test?

A major manufacturer challenge is to choose a TEa goal from many available options: US-specific CLIA requirements, CAP, Royal College of Pathologists of Australasia, Guidelines of the Germany Federal Medical Society (RiliBÄK [Richtlinien der Bundesarztekammer]), or other EQA/PT provider specifications. Another popular source of TEa goals is biological variability, but there are 3 targets from which to choose:

$$\text{Minimum TEa} < 1.65\,(0.75\,CV_i) + 0.375\,(CV_i^2 + CV_g^2)^{1/2}$$

$$\text{Desirable TEa} < 1.65\,(0.5\,CV_i) + 0.25\,(CV_i^2 + CV_g^2)^{1/2}$$

$$\text{Optimum TEa} < 1.65\,(0.25\,CV_i) + 0.125\,(CV_i^2 + CV_g^2)^{1/2},$$

where *CVi* is individual biological variability and *CVg* is group biological variability.

The IFCC Working Group on Allowable Error for Traceable Results (WG-AETR) was formed to define clinically acceptable limits for harmonization and better clinical application and to cooperate with manufacturers, regulatory bodies, and end-users.[48] WG-AETR concluded, "Although manufacturers are compelled by the European IVDD, 98/79/EC, to have traceability of the values assigned to their calibrators if suitable higher order reference materials and/or procedures are available, there is still no equivalence of results for many measurands determined in clinical laboratories." For some common analytes, such as sodium, current assays are too imprecise to meet TEa targets based on biological variation. Due to cost and limited resources, IVD manufacturers

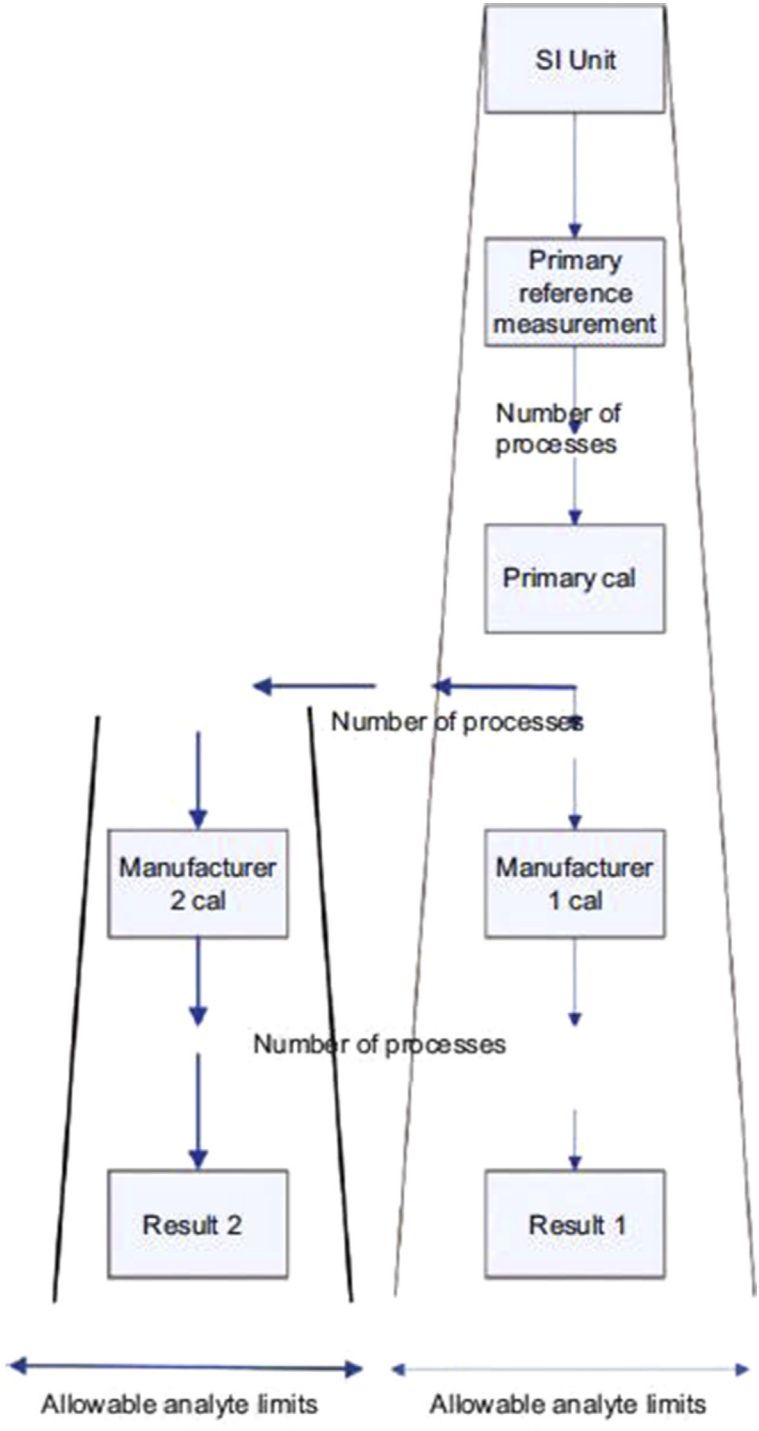

Fig. 3. Manufacturers may prepare calibrators (cal) starting with traceability to the same reference material and/or reference method but the calibrator manufacturing process

do not always follow the full traceability chain steps to value assign every new calibrator lot but rely on value transfer from an internally stored master calibrator material. In most cases, this procedure is probably valid, but a common complaint is calibrator lot-to-lot variability. WG-AETR noted when there are 2 traceability paths for a measurand, calibrators from different manufacturers may both be derived from valid traceability chains but produce nonequivalent results, as illustrated by **Fig. 3**. Equivalent results from 2 systems may only be possible by using a correction factor determined by a correlation study.

The international clinical laboratory community has embraced harmonization. A good example is the AACC International Consortium for Harmonization of Clinical Laboratory Results (ICHCLR).[3] The ICHCLR prioritizes analytes globally for harmonization and development of reference materials and RMPs. ICHCLR stakeholders include clinical laboratory and medical professional societies, IVD manufacturers, metrology institutes, public health organizations, regulatory agencies, and standard-setting organizations. A similar initiative is Pathology Harmony in the United Kingdom.[49] Pathology Harmony states, "As we the move towards full electronic reporting of pathology results, we appreciate more fully that variations in things such as test names, reference intervals and units of measurement associated with our results is something that hinders progress.'' In Australia, the Pathology Information, Terminology and Units Standardisation Project is dedicated to harmonization, in particular focusing on the interoperability of pathology test requesting and reporting.[50] Harmonization of report formats, RIs, and decision limits and best practice evidence for test requesting are also concerns.[51] Staff requesting and receiving test results and information system developers may be unaware of reporting differences, and clinicians often assume results from different methods are comparable and there is no risk of misinterpretation or adverse patient outcomes.

IN VITRO DIAGNOSTICS MANUFACTURERS AND HARMONIZATION

Clinical laboratories assume manufacturers have implemented calibrator metrological traceability and uncertainty. But full traceability may not be available or may lack sufficient detail, and calibrator uncertainty may be lacking or not be reasonable due to the variety of methods for calculating it.[33] IVD manufacturers have set aside traditional commercial competition for the goal of assay comparability and equivalent patient results. Even with rigorous calibrator traceability, manufacturers may be reluctant to provide full information for fear of disclosing proprietary data or opening themselves to technical criticism. Full disclosure of this information remains a challenge for Industry.[33]

Manufacturers all have substantial product development priority lists and requirements for which personnel and financial resources are committed over long-term periods. Reprioritization is possible and welcomed by industry to benefit clinicians, patients and health care systems, but global harmonization creates competing priorities for companies. As manufacturers support harmonization, timelines reflecting development cycles (years) require companies to simultaneously reprioritize

may diverge at some point, resulting in significantly different results for the same measurand in the same patient sample if tested by the 2 field methods, despite metrologically acceptable traceability for each assay's calibrators (cal). (*From* Bais R, Armbruster D, Jansen RT, et al. Defining acceptable limits for the metrological traceability of specific measurands. Clin Chem Lab Med 2013;51(5):975; with permission.)

resources while maintaining projects driving innovation and product portfolio development.

The IVD industry has readily accepted and supported the JCTLM.[13] Benefits from the JCTLM include realistic timelines to restandardize assays using traceability chains, an efficient source of traceability information (ie, the JCTLM database), and an effective communication forum on traceability and standardization. Manufacturers still face traceability challenges, including difficulty identifying reference materials and methods even with the JCTLM database, unavailability of some reference materials (eg, enzyme standards), lack of secondary reference materials for some measurands (eg, analytes in matrices, such as whole blood, serum, plasma, and urine), reference materials not expressed in SI units, noncommutable reference materials, reference methods not easily transferable or available (eg, definitive methods, such as isotope dilution–gas chromatography/mass spectrometry), time required to restandardize (often optimistically estimated at 18–24 months due to manufacturing process changes, labeling changes, inventory obsolescence, communications to customers/PT providers/regulatory agencies, changes in RIs, and so forth), producing assays compatible with evidence-based laboratory medicine, and laboratory medicine practice guidelines. Another consideration is whether harmonization always provides a benefit. Accuracy is important but there are situations in which existing assays may be relatively harmonized yet the reference method is very different from commercialized assays. The cost of harmonization, which includes physician education, patient safety, and investment in product redevelopment, should be weighed to prove the benefit of harmonization.

Health care consumers (physicians and patients) expect (take for granted) laboratory test that results are of high quality and suitable for diagnosis and management and that accurate results are produced by all laboratories at all times. But it is a daunting task for different laboratories to analyze the same patient specimens and generate equivalent results, within acceptable analytical tolerance limits, regardless of the measurement system used.[31]

Uncertainty is routine in metrology but a new concept in the clinical laboratory. Estimating uncertainty is valuable for manufacturers, allowing them to identify and minimize variability of calibrators. It remains unclear to what extent uncertainty is useful for clinical purposes and whether it should be reported with test results. For example, clinicians could be confused if ALT enzyme results are expressed in SI units with uncertainty as serum ALT; catalytic concentration = (1.15 $\pm$ 0.23 ukatal [kat]/L).[17]

SUMMARY

There is no doubt global harmonization of the clinical laboratory field is the next frontier and efforts to achieve this goal will continue in the twenty-first century. Success depends on creativity, such as creating the JCTLM as a new international organization devoted to assay traceability and bringing together a wide variety of stakeholders in this effort.[52] Analytical standardization to allow assays to produce clinically equivalent test results is only one facet of this movement. Success will be defined by harmonization of the TTP encompassing the preanalytical, analytical, and postanalytical phases.[53]

REFERENCES

1. Greenberg N. Update on current concepts and meanings in laboratory medicine: standardization, traceability and harmonization. Clin Chim Acta 2014;432:49–54.
2. Armbruster D, Donnelly J. Harmonization of clinical laboratory test results: the role of the IVD industry. EJIFCC 2016;27:37–47.

3. Miller WG, Myers GL, Gantzer ML, et al. Roadmap for harmonization of clinical laboratory measurements procedures. Clin Chem 2011;57:1108–17.
4. Tate JR, Johnson R, Barth J, et al. Harmonization of laboratory testing- current achievements and future strategies. Clin Chim Acta 2014;432:4–7.
5. McLawhon RW. Patient safety and clinical effectiveness as imperatives for achieving harmonization inside and outside the clinical laboratory. Clin Chem 2011;57:936–8.
6. White GH. Metrological traceability in clinical biochemistry. Ann Clin Biochem 2011;48:393–409.
7. Miller WG, Myers GL. Commutability still matters. Clin Chem 2013;59:1291–3.
8. Beastall GH. Adding value to laboratory medicine: a professional responsibility. Clin Chem Lab Med 2013;51:221–7.
9. Gantzer ML, Miller WG. Harmonisation of measurement procedures: how do we get it done? Clin Biochem Rev 2012;33:95–100.
10. Malone B. A jumpstart on harmonization. Clinical Laboratory News 2010;36:12.
11. Hernandez JS, Dale JC, Bennet K, et al. Challenges and opportunities for medical directors in pathology and laboratory medicine. Am J Clin Pathol 2010;133:8–13.
12. Stephan C, Siemben K, Camman H, et al. Between-method differences in prostate-specific antigen assays affect prostate cancer risk prediction by nomograms. Clin Chem 2011;57:995–1004.
13. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. Clin Biochem 2009;42:236–40.
14. Armbruster D, Miller RR. The Joint Committee on Traceability in Laboratory Medicine (JCTLM): a global approach to promote the standardization of clinical laboratory results. Clin Biochem 2007;28:105–13.
15. Morris HA. Traceability and standardization of immunoassays: a major challenge. Clin Biochem 2009;42:241–5.
16. NIST. Planning report 04-1, the impact of calibration error in medical decision making. Gaithersburg (MD): National Institute for Standards and Technology; 2004.
17. Hallworth MJ. The '70% claim': what is the evidence base? Ann Clin Biochem 2011;48:487–8.
18. Tate JR, Johnson R, Barth JH. Harmonization of laboratory testing- a global activity [Editorial]. Clin Chim Acta 2014;432:1–3.
19. Teodoro-Morrison T, Janssen MJW, Mols J, et al. Evaluation of a next generation direct whole blood enzymatic assay for hemoglobin A1c on the ARCHITECT c8000 chemistry system. Clin Chem Lab Med 2015;53:125–32.
20. Burtis CA, Ashwood ER, editors. Tietz textbook of clinical chemistry. 3rd edition. Philadelphia: W.B. Saunders Co; 1999.
21. Burtis CA, Ashwood ER, Bruns DE, editors. Tietz textbook of clinical chemistry and molecular diagnostics. 4th edition. Philadelphia: W.B. Saunders Co; 2006.
22. Burtis CA, Ashwood ER, Bruns DE, editors. Tietz textbook of clinical chemistry and molecular diagnostics. 5th edition. Philadelphia: W.B. Saunders Co; 2011.
23. De Bievre P. Metrological traceability of measurement results in chemistry: concepts and implementation. IUPAC Technical Report. Pure Appl Chem 2011;83:1873–935.
24. Greenberg N. The IVD directive and availability of reference systems for IVD medical devices: a view from Industry. IVD Technology 2001;7:18–27.
25. Powers DM. Traceability of assay calibrators: the EU's IVD Directive raises the bar. IVD Technology 2000; July – August.
26. Vesper HW, Thienpont LM. Traceability in laboratory medicine. Clin Chem 2009;55:1067–75.

27. ISO 17511. In vitro diagnostic medical devices - measurement of quantities in biological samples - metrological traceability of values assigned to calibrators and control materials. Geneva, Switzerland: International Organization for Standardization; 2003.
28. ISO 15189. Medical laboratories — requirements for quality and competence. International Organization for Standardization; 2012.
29. CLSI EP21. Expression of measurement uncertainty in laboratory medicine. Wayne (PA): Clinical Laboratory Standards Institute; 2012.
30. Dybkaer R. From total allowable error via metrological traceability to uncertainty of measurement of the unbiased result. Accred Qual Assur 1999;4:401–5.
31. Armbruster D. Measurement traceability and US IVD manufacturers: the impact of metrology. Accred Qual Assur 2009;14:393–8.
32. Panteghini M. Traceability, reference systems and result comparability. Clin Biochem Rev 2007;28:97–104.
33. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. Clin Chim Acta 2014;432:55–61.
34. Available at: www.bipm.org/jctlm. Accessed March 19, 2016.
35. Bunk DM. Reference materials and reference measurement procedures: an overview from a national metrology institute. Clin Biochem Rev 2007;28:131–7.
36. Fasce CF, Rej R, Copeland WH, et al. A discussion of enzyme reference materials: applications and specifications. Clin Chem 1973;19:5–9.
37. Rej R. Accurate enzyme activity measurements. Arch Pathol Lab Med 1993;117:352–64.
38. Miller WG, Jones GRD, Horowitz GL, et al. Proficiency testing external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.
39. Miller WG, Myers GL, Rej R. Why commutability matters. Clin Chem 2006;52:553–4.
40. CLSI EP30. Characterization and qualification of commutable reference materials for laboratory medicine. Wayne (PA): Clinical Laboratory Standards Institute; 2010.
41. Horowitz GL. Proficiency testing matters. Clin Chem 2013;59:335–7.
42. Miller WG, Myers GL, Ashwood ER, et al. State of the art in trueness and interlaboratory harmonization for 10 analytes in general clinical chemistry. Arch Pathol Lab Med 2008;132:838–46.
43. Miller WG. The role of proficiency testing in achieving standardization and harmonization between laboratories. Clin Biochem 2009;42:232–5.
44. Jansen R, Schumann G, Baadenhuijsen H, et al. Trueness verification and traceability assessment of results from commercial systems for measurement of six enzyme activities in serum: an international study in the EC4 framework of the Calibration 2000 project. Clin Chim Acta 2006;368:160–7.
45. Levine DM, Maine GT, Armbruster DA, et al. The need for standardization of tacrolimus assays. Clin Chem 2011;57:1739–47.
46. Plebani M. Harmonization in laboratory medicine: the complete picture. Clin Chem Lab Med 2013;51:741–51.
47. Thienpont LM. Accuracy in clinical chemistry- Who will kiss Sleeping Beauty awake? Clin Chem 2008;46:1220–2.
48. Bais R, Armbruster D, Jansen RTP, et al. Defining acceptable limits for the metrological traceability of specific measurands. Clin Chem Lab Med 2013;51:973–6.
49. Berg J, Lane V. Pathology Harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory. Ann Clin Biochem 2011;48:195–7.

50. Legg M. Standardisation of test requesting and reporting for the electronic health record. Clin Chim Acta 2014;432:148–56.
51. Tate JR, Johnson R, Legg M. Haromonisation of laboratory testing. Clin Biochem Rev 2012;33:81–4.
52. Jones GRD, Jackson C. The Joint Committee for Traceability in Laboratory Medicine (JCTLM)- Its history and operation. Clin Chim Acta 2016;453:86–94.
53. Aarsand AK, Sandberg S. How to achieve harmonization of laboratory testing-the complete picture. Clin Chim Acta 2014;432:8–14.