Metrological traceability and equivalence of measurement results in Laboratory Medicine

Authored by Elvar Theodorsson (elvar.theodorsson@liu.se, +46736209471) and Maxim Vonsky (m.vonsky@vniim.ru) and reviewed by the JCTLM Traceability Education & Promotion Working group and by other beneficiaries of the JCTLM

BIPM in Paris in 2022

This is a "living document," version 2022-03-23.

Suggestions for improvements are gratefully received.

JCTLM Secretariat

Bureau International des Poids et Mesures
Pavillon de Breteuil
92312 Sèvres
Cedex France
Tel: +33 1 45 07 70 70
Fax: +33 1 45 34 20 21
Email: jctlm@bipm.org

Metrological traceability and equivalence of measurement results in Laboratory Medicine

Metrological traceability as the basis of *equivalence of measurement results* in Laboratory medicine is crucial for making diagnoses, for therapeutic decisions, for monitoring of treatment results, for prediction of clinical outcomes, for proper adherence to clinical guidelines, for pooling of measurement results from different laboratories and for avoiding the extra cost of re-measuring samples from patients moving between healthcare facilities.

Traceability means “comparability by being connected.” The word “traceability” has its roots in Latin from the word tractus = drawn and trahere = to draw. The abbreviated term "traceability" is usually intended to mean “metrological traceability” but is sometimes used in other contexts, such as “sample traceability,” “document traceability,” “instrument traceability,” or “material traceability,” where the history ("trace") of an item is meant. Traceability in Laboratory Medicine should not be confused with, e.g., the ability to trace goods to specific factories or foodstuffs to certain farms. Therefore, the term "metrological traceability" is preferred if there is any risk of confusion.

Metrological Traceability, by definition, is “a property of a measurement result that can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” (1).

Equivalence of measurement results is “agreement of measured values among different in vitro diagnostic measurement devices intended to measure the same measurand, where the differences in measured values on the same human samples do not affect clinical interpretation” ISO 17511:2020, 3.13 (2). Thus “equivalence” is a clinical

concept which does not necessarily include equivalence of the molecular structures being measured.

The concept *transferability* of measurement results in patient samples has been used to describe a property of patient results that can be used for medical decisions irrespective of the place and time of measurement (2).

Even though measurement results are traceable from a metrological point of view, they are not necessarily equivalent. This may happen for several reasons. This may be because the assays are traceable to different reference standards. After all, there is excessive measurement uncertainty in the traceability hierarchy or because of non-commutable materials or differences in analytical specificity from a clinical point of view. Immunochemical measurement methods represent typical examples where antibodies raised against the analyte commonly bind to different epitopes. Different epitopes may result in various medical interpretations of the results.

Most of the work remains of characterizing specific parts of the molecular structures of biomarkers that have the most pronounced relation to the diagnostic information obtained by measuring the concentrations or activities of the biomarkers. The standardization of glycated hemoglobin illustrates the enormous amount of work required for such a task and the metrological and medical advantages of such work (3-9).

The distinct formal definition of metrological traceability is comprehensive for the measurands in *physics. Still, it takes* numerous complex aspects of the concept for granted when applied to Laboratory Medicine (10-12). We will summarize these aspects here using traceability's concept pillars/fundaments.

Metrological traceability in Laboratory Medicine implies the establishment of a documented and verifiable relation (a trace, calibration hierarchy) to a stated *metrological reference* which in the case of Laboratory Medicine must be amongst the following:

1. The definition of a *SI unit*
2. A value of a *certified reference material*
3. The result of a *reference measuring procedure*
4. The value assigned to an *international conventional reference material*
5. The values assigned by an *international harmonization protocol*

A standard metrological reference is a prerequisite for metrologically comparable- and thereby traceable measurement results.

Specification of the reference at the top of the traceability hierarchy must include the time this reference was used to establish the calibration hierarchy and any other relevant metrological information about the reference (11, 12). Measurement results traceable to SI units are traceable indefinitely. Still, measurement results traceable to international conventional reference materials (e.g., WHO International Standards) are only traceable as long as the actual batch of the material is available and within the time limits for storage.

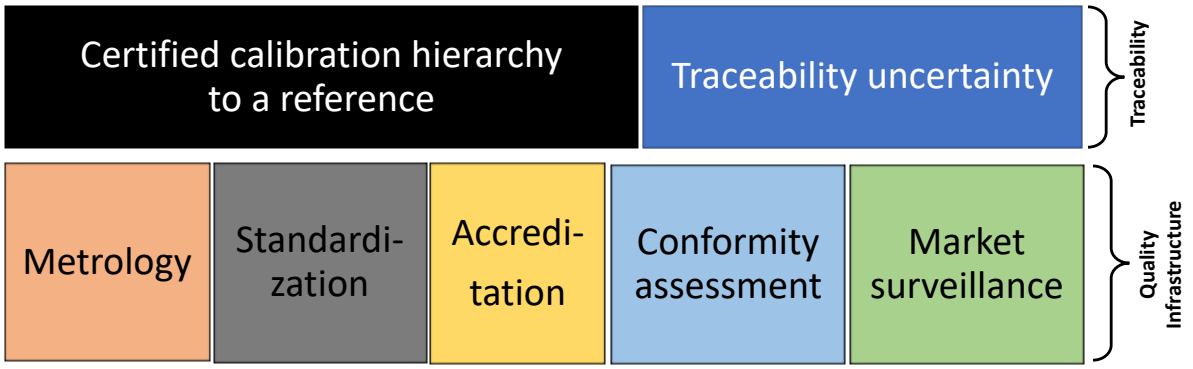


Figure 4: Traceability and corresponding quality infrastructure as defined by the International Network on Quality Infrastructure (INetQI, https://www.bipm.org/en/liaison-partners/inetqi). "Market surveillance" - in laboratory medicine corresponds to post-market surveillance and oversight from regulators, manufacturers, and customers. The two primary traceability components are illustrated at the top and the quality infrastructure at its base. In Laboratory medicine, the quality infrastructure is known as "Pillars of Traceability," as described below.

To obtain and maintain the ability to compare measurement results geographically and over time, the results need to be linked to a common stable reference- or measurement standard which serves as the common reference for all results - geographically and over time. *Linking measurement results to a standard reference* is the essence of metrological traceability. Measurement results traceable to a standard reference, with sufficiently low measurement uncertainty, can be compared between different countries, laboratories, and measuring systems. They can be reliably compared over extended periods since they are based on a standard reference. Furthermore, the traceability hierarchy (2), including reference materials and reference measurement procedures, must be detailed and the measurement uncertainty due to all steps in the traceability hierarchy documented. Like physical chains, any broken link in a metrological traceability chain/hierarchy means that a measurement result is not traceable.

Metrological traceability is essentially a property of a measurement result expressed on the *nominal* scale. This means – in principle – that measurement results are described as being either traceable or not. Claims of traceability cannot be made without information about the associated uncertainty of the measurement results, which must always accompany metrological traceability statements. Measurement uncertainty is expressed on the *ratio* measurement scale only for quantities that can be compared by ratio. A general treatment of nominal property examination uncertainty is still to be developed (13, 14). All certified reference materials must be accompanied by a statement of an assigned value and associated *uncertainty* in the entire traceability hierarchy. The uncertainty of the results produced by local laboratories (15) includes all sources, including the traceability hierarchy, reproducibility, and repeatability of U(s) laboratory. Uncertainty at the highest levels of the traceability hierarchy must be small compared to the uncertainty at the lowest levels of the traceability hierarchy (2).

The quality infrastructure necessary for traceability

Traceability and measurements of its uncertainty components must rest on a solid fundament of *quality infrastructure,* which is defined as follows by the International Network on Quality Infrastructure (INetQI, https://www.bipm.org/en/liaison-partners/inetqi):

> "The system comprises the organizations (public and private) and the policies, relevant legal and regulatory framework, and practices needed to support and enhance the quality, safety, and environmental soundness of goods, services, and processes. […] It relies on metrology, standardization, accreditation, conformity assessment, and market surveillance."
> (https://www.bipm.org/en/liaison/quality-infrastructure)

This traceability and corresponding quality infrastructure are depicted in Figure 4.

In Laboratory Medicine the emphasis is traditionally on the following components of infrastructure for traceability:

1. Fitness for the intended use of the measuring systems/measurement procedures and reference materials.
2. The laboratory's documented quality management includes regular inspection/evaluation by an independent external authority, e.g., when ISO standards accredit laboratories.
3. Regular participation by the laboratory in a trueness-based external quality assessment which, if possible, applies commutable materials and reference materials with reference values measured by reference measurement systems

4. An appropriate timeline of the monitoring and documentation of the traceability of the measurement results.

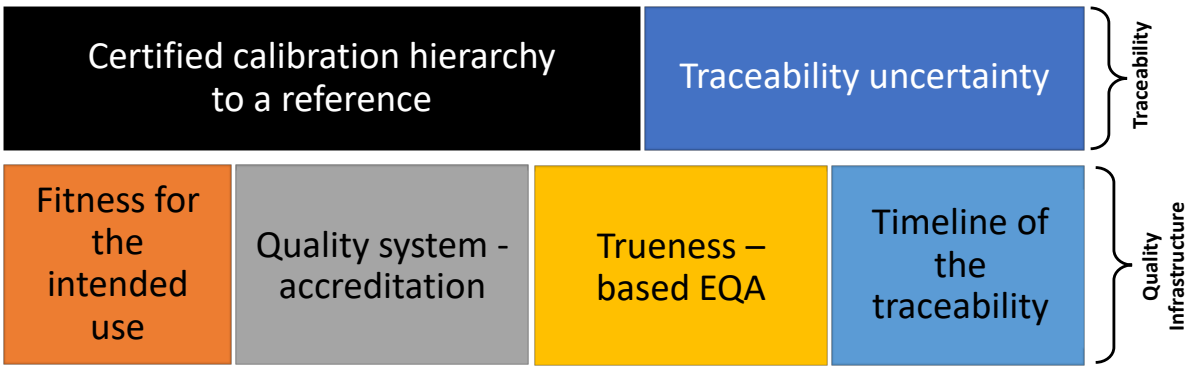


Figure 5: Traceability and essentials of a corresponding quality infrastructure as perceived in Laboratory Medicine. See also "Pillars of traceability" explained below. The quality system must incorporate accreditation demands, including internal quality control. In case the traceability is to SI, the timeline of traceability is indefinite. However, international conventional certified reference materials (e.g., WHO) have a limited lifetime due to limited availability and storage time.

Metrological traceability of a measurement result defined by its traceability hierarchy and measurement uncertainty is no guarantee that the measurement uncertainty is fit for a given, intended use or that mistakes are minimized. Quality infrastructure is needed to substantiate claims of traceability in Laboratory Medicine. It provides evidence and insurance through an extended period of the fitness for the intended use and the quality of the measurement results. The fitness for the intended use of the measuring systems and certified reference materials or reference measuring systems are prerequisites for traceability. An externally- and regularly revised quality system in the laboratory represents a necessary foundation for avoiding mistakes. Trueness-based external quality assurance offers further confidence in the timeline of the quality and traceability of particular measurement results by a specific laboratory.

In addition, end-users must be aware of this property of an impact or results being traceable. For example, if two laboratories perform the same test and the results are traceable/comparable, which is not known by the clinician, the test may be repeated unnecessarily. Similarly, a clinician may wrongly assume results from two labs are equivalent and make a diagnostic error.

The pillars of traceability

Traceability in Laboratory Medicine is the combined end-result of several factors on which it rests, commonly depicted in Laboratory Medicine as pillars in an ancient temple (16-19). Which pillars are emphasized may vary, but reference materials,

reference procedures, reference measuring systems, a network of reference measurement laboratories, and laboratory quality systems are always included (Figure 6).

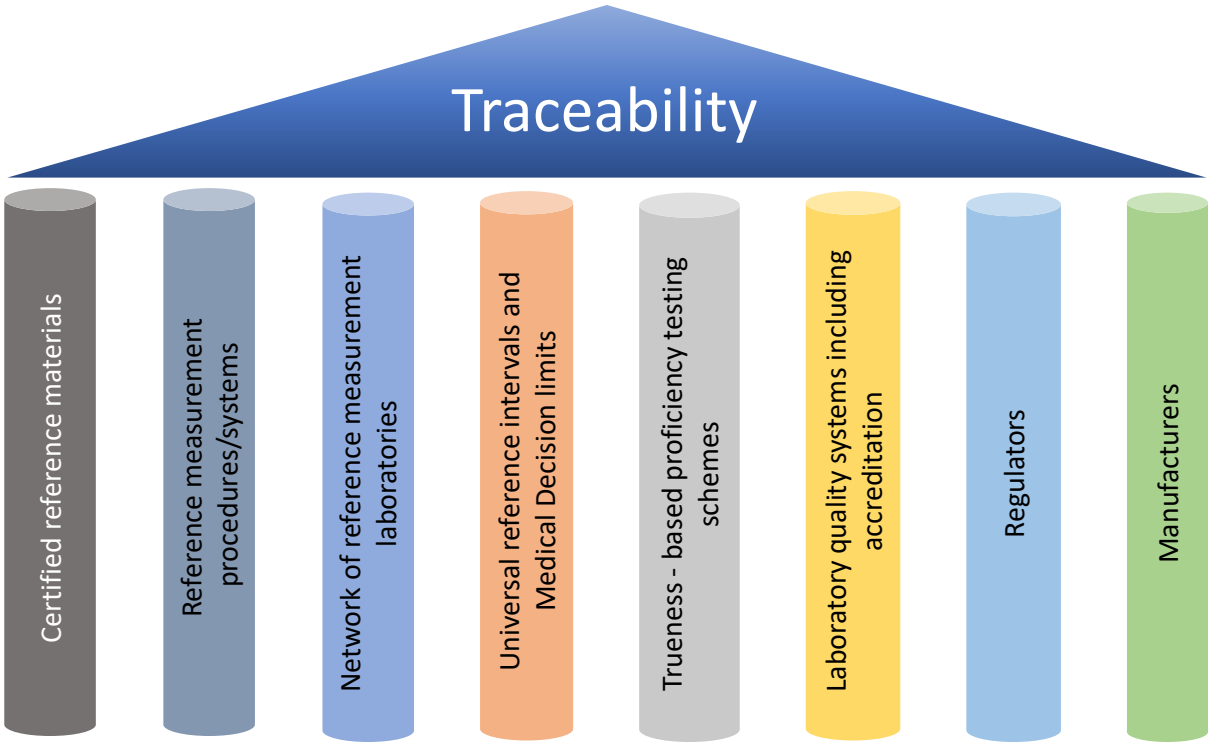


Figure 6: The Joint Committee of Traceability in Laboratory Medicine (JCTLM) established the three initial *pillars of traceability*: 1. Reference measurement procedures (RMP)/reference measurement systems, 2. Reference materials (RM) (including commutable reference materials) and 3. Network of Reference Measurement Laboratories (RELA studies). The International Federation of Clinical Chemistry (IFCC) described a fourth pillar: 4. Universal reference intervals and Medical Decision Limits. Prof. Mauro Panteghini (20) suggested a fifth pillar, 5. Accuracy-based grading of proficiency testing schemes to ensure and maintain international reference systems. *Proficiency testing schemes* for monitoring the maintenance of traceability are essential, especially if the plans are trueness-based. The roles of *manufacturers* and *regulators* are rarely mentioned despite their evident and increasing roles in Laboratory Medicine.

Regulatory issues

Traceability is far from only of academic interest as it has become increasingly crucial for regulators around the globe, as exemplified by the following text in an EU Directive:

The EU Directive 98/79/EC on IVD MDs Annex I, Essential requirements A.3 reads: "The traceability of values assigned to calibrators and control materials must be assured through available reference measurement procedures and available reference materials of a higher order."

Furthermore, the Competent Regulatory Authorities of the European Union and European Commission consider it essential that WHO International Standards are within the "higher-order" of reference materials and thus acceptable for use as references for IVDs. [Report on WHO consultation 2004-06-07/08].

This introduction of the concept of "Reference materials of a higher order" means that traceability, which is established to SI, can also be established to values assigned to international conventional reference materials and values assigned by international harmonization protocols (2, 21).

Documented and independently and regularly reviewed quality system

Traceability underpinning equivalence and transferability of measurement results are dependent on several *attributes of individual laboratories* - including the following:

1. The *laboratory employees* must be appropriately educated and competent in using the measuring systems and procedures of the laboratory.
2. The *measuring systems* must be fit for the intended use as demonstrated by appropriate validations and verifications as appropriate.
3. The measuring systems must be appropriately maintained.
4. The knowledge and skills of the laboratory employees must be developed and maintained over time.
5. Maintaining a comprehensive internal quality control system and *external quality assessment*/proficiency testing.
6. The practice of a regular *external and independent evaluation of the laboratory's quality system*.

Traceability depends on international agreements and efforts at national and international levels to provide appropriate reference materials and reference measuring systems. Individual laboratories then need to establish and maintain traceability to these materials, which is also a requirement of ISO-17025:2017 and ISO-15189:2012. Regulatory efforts exert pressure on laboratories and manufacturers alike in this direction (22, 23).

Appropriate evidence for the technical competence of the laboratory in the framework of accreditation by ILAC member organizations (https://ilac.org) and the claimed metrological traceability is likely to include at least the following clauses: (numbers refer to clauses in ISO-17025:2017 and ISO-15189:2012):

ISO-17025:2017

6.2 Documentation and records for the competence of personnel

6.3 Documentation and records for facilities and environmental conditions
6.4 Records for equipment that can influence laboratory activities
6.5 Documentation and records for metrological traceability of measurement results
6.6 Audits of the calibration laboratory
7.2.2.4 Records of calibration method validation
7.6 Procedures for evaluation of measurement uncertainty
7.7 Documentation and records for ensuring the validity of results
8.8 Audits of the calibration laboratory

ISO-15189:2012

5.5.1.2 Verification of examination procedures
5.5.1.3 Validation of examination procedures
5.5.1.4 Measurement uncertainty of measured quantity values
5.5.2 Biological reference intervals or clinical decision values
5.6.2 Quality control

Requirements that must be fulfilled for a valid claim of traceability.

The reference which constitutes the basis of the traceability chain must be amongst the following: the definition of a SI unit, a certified value of reference material, the result of a reference measuring system, the value assigned to an international conventional calibrator, and the values assigned by the international harmonization protocol.

Metrological traceability requires an established and characterized calibration hierarchy. Suppose the measurement model used in each step of the calibration hierarchy involves more than one input quantity. Each input quantity (e.g., mass, time, volume) should be metrologically traceable if that quantity represents a substantial contribution to the measurement result.

The reference specification must include the time at which this reference was used in establishing the calibration hierarchy, along with any other relevant metrological information about the reference, such as when the first calibration in the calibration hierarchy was performed (1).

The following components are also implicitly a requirement (24)

1. A clear definition of the quantity being measured (definition of the measurand).
2. A complete description of the measuring procedure and the reference materials used to perform the measurement.
3. A measurement result includes a documented measurement uncertainty.
4. The validation includes the fitness for the intended use evaluation of the measuring system.

5. Details of the internal quality assurance program used for establishing the status of the measuring system or measurement standard related to the claim of traceability (25).
6. Details of the participation in a trueness-based proficiency testing program the time the measuring system or reference material was compared to it. The commutability of the calibrators used in the traceability chain must be established, preferably using methods described by Miller et al. (26-28).

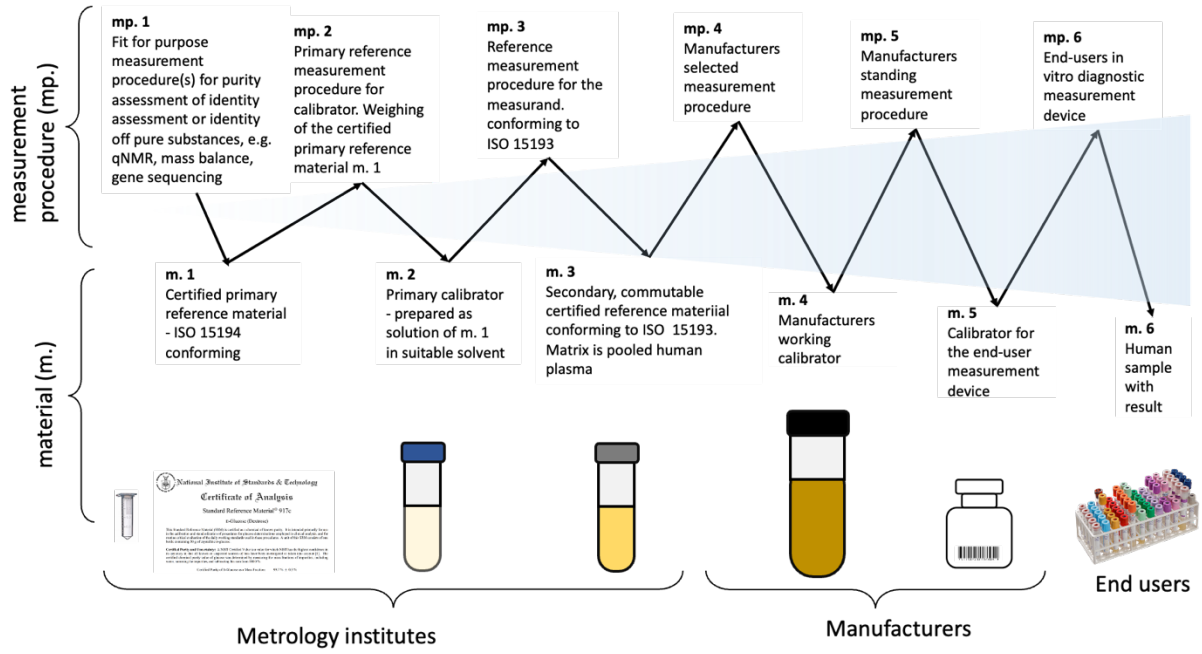


Figure 7: Illustration of a general traceability hierarchy for measuring a measurand in a plasma sample from a patient. The upper half of the illustration lists the chain of measuring systems, and the lower half lists the reference materials used to carry the values of the unbroken chain of reference measurements in the calibration hierarchy. To be a proper part of a calibration hierarchy, manufacturers should have accreditation as calibration laboratories using reference measurement procedures according to ISO 15195.

Challenges in traceability and equivalence of measurement results in Laboratory Medicine

Optimally, quantitative measurement results in Laboratory Medicine (analogous to quantitative measurement results of physical quantities) should be made traceable to the SI unit when the analyte can be uniquely identified, e.g., by a chemical structure, e.g., sequence of nucleic acid, etc. This means that measurements in Laboratory Medicine are optimally expressed as “amount of substance,” the relevant basic quantity in the international measuring system (SI), or have the nature of a count. Unfortunately, this is possible only for a minority of the biomarkers used in clinical practice and Laboratory Medicine since the “analytes” corresponding to the measurands need to be available in a

pure form and the same stable form in healthy and compromised human beings. Macromolecules crucial for the proper function of the human organism are commonly present *in vivo* as multiple molecular forms posing the risk of resulting in different quantity values for different measurands when measured by other measuring systems.

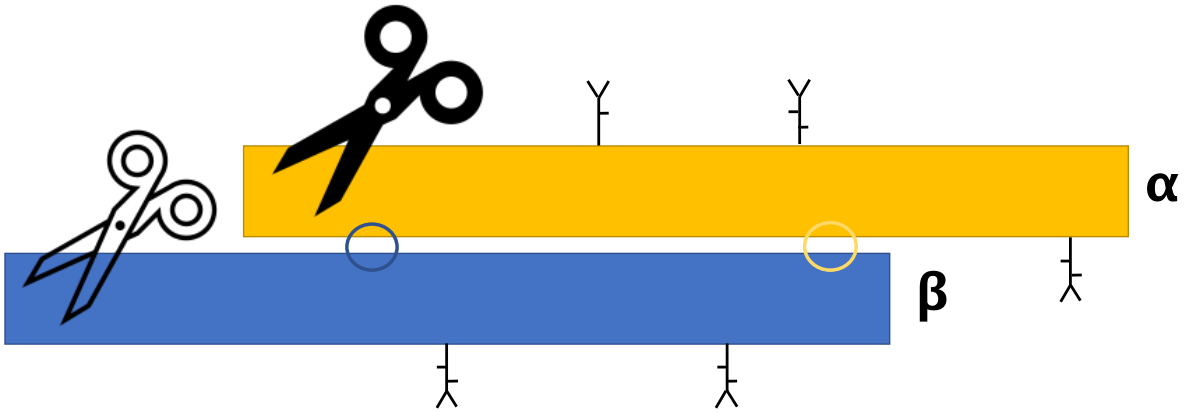


Figure 8: A schematic graphic showing a macromolecule with essential characteristics similar to thyrotropin.

Thyrotropin (TSH) is a typical example of a common molecular complexity in laboratory medicine. TSH is a glycoprotein with a molecular weight of approximately 30 kilo Daltons which consists of two subunits: alpha and beta (Figure 8). The beta-subunit carries the TSH-specific immunological and biological information. In contrast, the alpha-chain takes species-specific information and has an identical amino acid sequence to the α-chains of lutropin, follitropin, and gonadotropin. Synthesis of a mature TSH molecule requires specific enzymatic cleavage of signal peptides from both TSH alpha- and beta-subunits, trimming mannose and further adding fucose galactose and sialic acids. Mature TSH molecules are complex carbohydrate structures capped with sulfate and sialic acid molecules. The post-translational modifications of TSH primarily in the carbohydrate sidechains are hormone- and disease-dependent (29-32).

This means that it is not feasible to produce a homogenous and stable form of TSH reflecting the TSH in the blood plasma from all patients. The reason is that there are multiple forms created by enzymatic activities and glycosylation processes in the human organism. Biological mechanisms influence these processes in health and disease.

It is possible to establish traceability of measurement results to "*international conventional reference materials*," for example, WHO reference materials, and agreed on reference measurement methods. Still, calibration hierarchies to SI are impossible in these circumstances, as explained above. Traceability for all non-SI traceable measurement results in Laboratory Medicine and where international conventional

reference materials are not available must be handled separately, commonly through *harmonization protocols* (2, 21) since the imagined "analyte" may only partially comprise molecules represented in quantity measured representing the measurand. This is an area of ongoing development with practical examples still pending at the time of writing.

Several in vitro diagnostic measuring systems claim to measure the exact "analytes" but base their measurements on different physiochemical principles, resulting in different measurement results for the same human sample or reference material. The most likely reasons are *differences in measurement selectivity characteristics,* including tertiary molecular structures, microheterogeneity, or chemical configurations of the intended "analyte." Activities must therefore be undertaken at all levels of the calibration hierarchy to prevent problems caused by differences or changes in the measured quantity values among the different measuring systems at the various levels in the calibration hierarchy. The essence is to recognize and minimize the differences between the quantity being measured and the quantity intended to be measured (measurand). An example is quantifying viral load by quantitative polymerase chain reaction (qPCR). The intended analyte is the viral genome, but the target sequence presents the de facto analyte. Cases with variable microheterogeneity of the analyte (e.g., isoforms and posttranslational modifications) within the calibrators or human samples are significant.

Surrogate markers in Laboratory Medicine

In clinical medicine, a "surrogate marker" is " …a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy" (33). A usual goal in cancer treatment is to decrease the volume of the tumor cells in the body. Numerous imaging techniques can be used to measure tumor volume as tumor markers. Cancer markers, e.g., prostate-specific antigen (PSA), represent indirect measurands for tumor volume and are therefore surrogate markers.

Analytical Chemistry commonly measures quantities in samples to investigate whether the concentrations are above limits decided by law, regulation, or other concrete targets. In Laboratory Medicine, the fundamental question is whether the patient is healthy, at-risk, or sick or whether they have improved by treatment or not. A surrogate marker known to correlate with a correct answer to the clinical question is measured to answer these fundamental questions. A patient blood sample is taken, and the plasma or serum portion of whole blood from the patient is separated from the cells. The concentration of the surrogate marker is measured using a combination of chemical and

physical methods using reference materials subjected to the same chemical and physical methods for calibration.

Metrological traceability is a challenge, mainly occur when the principle of the measuring system is based on the *detection of a surrogate for the analyte of interest,* e.g., a peptide epitope in a large protein rather than the entire protein molecule or a fragment of the protein molecule, a short segment of the DNA or RNA macromolecule. For example, complete single-strand viral genomic RNA of Human immunodeficiency virus 1 (HIV-1) consists of 9,181 nucleotides (34). Regular monitoring of HIV viral load performed by RT-qPCR via genomic RNA target sequences measurements is considered the most accurate and meaningful measure of influential ART. It is recommended that clinicians assess response to ART using viral load assays (35). Several different commercial RT-qPCR assays (like OBAS® TaqMan® (Roche Molecular Systems), Abbott RealTime m2000rt (Abbott Molecular), NucliSENS EasyQ® (bioMérieux), VERSANT® kPCR (Siemens Healthcare Diagnostics), Generic HIV Viral Load (Biocentric), VERSANT HIV RNA 3.0 Assay (bDNA), artus® HI Virus-1 RG RT-PCR and artus® HI Virus-1 QS-RGQ Kit (QIAGEN)) are used for the monitoring with the sensitivity vary from 81.02 to 95.24 and specificity from 55.16 to 96.74 % (36). The typical target for RT-qPCR covers no more than 150 nucleotides of 9,181 nucleotides of the viral genome. Different assays target different viral genome subsequences, which stipulate other analytical characteristics. It is essential that during RT-qPCR, broken viral genome RNA fragments may be amplified with the same efficiency as the intact viral genome, which results in viral load overestimation. We have to pay the price for using “surrogate analytes” for our inability to measure full-size macromolecules.

Similarly, the calibrators used by measuring systems may contain a homolog molecule that is a surrogate for the analyte found in human samples. Two or more measuring systems with immunochemical measurement methods all claim to measure the amount of substance concentration of a single protein hormone (e.g., prostate-specific antigen [PSA]) even though they use binding reactions that are only selective for a 4-8 amino acid part of the analyte which is adjacent in the three-dimensional structure of the molecule in solution. A measuring system that involves several features of molecules of interest has higher selectivity than a measurement system relying on one. For example, LC/MS/MS uses a combination of chromatographic characteristics of molecules combined with specific molecular masses of fragments of molecules created when a stream of electrons fragments molecules in the sample.

Meaning of "surrogate marker" in Laboratory Medicine	Examples
Biomarker as surrogate marker for an inflammatory condition.	Measuring the concentration of C-reactive protein (CRP) in the patient´s plasma/serum is a surrogate marker for inflammation since macrophages in the organism of the patients incorporate cell remnants that are the results of inflammation and induce the liver to produce CRP by producing cytokines.
Biomarker as surrogate marker for a disease process	The use of tumor marker as surrogate marker for the total volume of cancer in the organism, e.g. the measurement of prostate specific antigen as marker for prostate cancer
Biomarker as surrogate marker for lack of nutrients.	Methyl Malonic Acid as marker for the lack of vitamin B_{12}
Measuring an "epitope" of a macromolecule, e.g. 4-8 amino acids in a protein macromolecule as a surrogate marker for the macromolecule	Using immunochemical methods to measure a epitope of a peptide hormone compared to measuring properties of the entire peptide hormone e.g. using LC/MS/MS

Table 2: The different meanings of the concept "surrogate marker" in Laboratory Medicine

Equivalence of measurement results

The equivalence of measurement results from different measuring systems may be observed in very selective (but different) measurement principles (e.g., a mass-spectrometric measurement procedure vs. an immunoassay for a protein hormone in patient plasma). However, other values are commonly measured because each measuring system measures different quantities, e.g., the binding of the selective antibodies to different epitopes of the intended molecules and molecular heterogeneity due to post-translational processing and variable protein binding influencing immunoassay, for example.

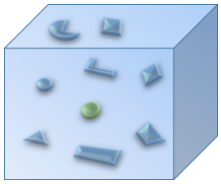
Amongst further challenges facing the specialties of Laboratory Medicine in producing traceable measurement results are the following:

1. There are “matrix effects ” (substances and factors in the sample except for the analyte of interest that may influence the results).
2. Inability to produce the analyte in a unique and pure form that can be weighed.
3. Molecular heterogeneity, for example, transferrin, LH, FSH, TSH.
4. Selectivity for different epitopes of the molecule of interest.
5. Lack of knowledge of which epitopes of molecules are medically most relevant, for example, most substantial biological activity or best diagnostic properties.
6. Changes in posttranslational modification of molecules in health and disease, such as LH and FSH, occur during the menstrual cycle.

Similarities and differences between reference materials in Laboratory Medicine and physical measurement standards

The general principles and terminology of metrology are applied both for physics and Analytical Chemistry and Laboratory Medicine fields(1). This means that a reference can be a “measurement standard” or a “reference material,” traceability is a concept used in all areas of metrology, and SI units are used whenever possible in all countries that have comprehensively implemented the SI system. However, reference materials in Laboratory Medicine, not even certified reference materials, realize the SI unit to the extent that physical measurement standards do because of the presence of “influence quantities” in the calibrators and the patient samples. Physical quantities such as length and mass can usually be measured without significant influence from unwanted influences. However, the quantities measured in, for example, human plasma samples can never be measured without the risk of the impact of surrounding molecules which commonly are present in orders of magnitude higher concentrations than the analyte of interest.

The vital matter of commutability of reference materials will be discussed in a separate chapter in this series.

Cause for lack of selectivity	**Graphical illustration**
The presence of “matrix factors” – all other molecules and other factors in the calibrator or sample that influence the measurement result – except the analyte of interest.	

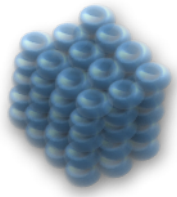
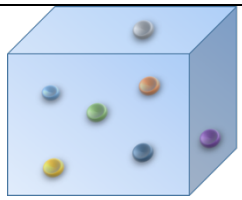
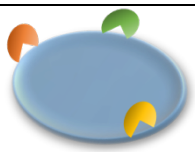
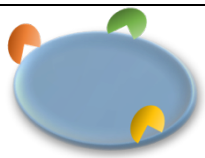
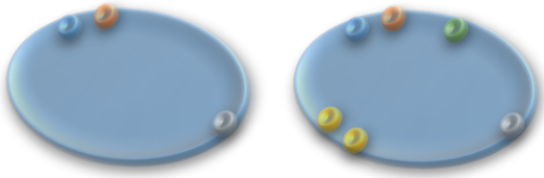
Inability to produce the analyte of interest in a pure form that can be weighed.	
Molecular heterogeneity of the "analyte", for example transferrin, LH, FSH, TSH.	
Selectivity, for example of antibodies or nucleic acids for different parts (epitopes or sequences) of the target molecule..	
Lack of knowledge of which epitopes of molecules are medically most relevant, for example most substantial biological activity or best diagnostic properties.	
Changes in posttranslational modification of molecules in various physiological or pathological conditions, for example LH and FSH during the ovarian cycle.	

Figure 9: Illustration of common causes of lack of selectivity which results are challenges for traceability in Laboratory Medicine.

The International Laboratory Accreditation Cooperation (ILAC) criteria for confirming metrological traceability

An unbroken hierarchy of comparisons going back to stated national or international standards references acceptable to the end-users of the measuring systems in Laboratory Medicine.

The measurement uncertainty for each step in the traceability hierarchy must be calculated or estimated according to agreed methods. It must be stated that an overall uncertainty for the whole measurement hierarchy can be calculated.

Each step in the measurement hierarchy must be performed according to documented and generally accepted procedures, and the measurement results must be recorded.

The laboratories performing one or more steps in the measurement hierarchy must supply evidence for their technical competence (e.g., by demonstrating that they are accredited). When possible, the top of the measurement hierarchy should end at certified reference materials for the realization of the SI units. Calibrations must be repeated at appropriate intervals; the length of these intervals will depend on several variables (e.g., measurement uncertainty required, frequency of use, way of use, stability of the equipment)" (37).

Traceability to SI

Prerequisites for traceability to SI are:

1. The molecules of interest are present in the organism in a single and unique molecular form.
2. The availability of the molecules of interest in pure form for making reference materials where the "analyte" can be characterized by primary measurement methods (for example, measuring weight or isotope dilution mass spectrometry, IDMS) to produce the reference material.
3. The definition of the measurand includes relevant descriptions of the sample and matrix. For example, sodium may give different results with changes in protein concentration with "direct" and "indirect" ion-sensitive electrodes due to the definition of mmol per liter of the sample rather than mmol/liter of solvent (water) in the sample.

Electrolytes and other small molecules generally fulfill these criteria enabling traceability to a pure preparation of the substance that can be characterized for purity weighed, thereby establishing traceability to SI.

Traceability to international conventional reference materials

Large molecules in the organism are commonly subject to post-translational processing, including enzymatic cleavage and conjugation reactions, e.g., glycosylation. This means that the analyte is present in more than one molecular isoform in the organism. An extract of human tissue containing high concentrations of the molecules, including all its potential multiple molecular isoforms, is commonly used to manufacture reference materials. Thus, a unique molecular form representing the analyte cannot be produced, and SI traceability is impossible.

Reference materials are needed for many such molecules (38, 39). To serve the need for traceability of measurands in this category, *international conventional reference*

materials are prepared with state-of-the-art purification and identification techniques. Their biological function can be tested using *bioassay* to establish their biological activity. An international unit (IU) is then assigned by convention, e.g., one international unit is set to 1 mg preparation of the tissue sample. Once defined, the IU is passed to all further IS preparations, preferably using bioassay via interlaboratory comparison. The amount of reference material representing the unit commonly decreases over time due to improved purification techniques.

This means that the international unit of a measurand is commonly defined as biological function and not as a chemical structure, and therefore a moving target until a possible situation occurs when a unique chemical structure is identified, representing the biological activity and the diagnostic properties of the measurand. If such developments are possible, the measurand can be made traceable to SI, which is the optimal situation.

When a substance of biological, biotechnological, or synthetic origin has its activity defined by, e.g., the World Health Organization (WHO) in terms of an International Unit (IU), then the unit is IU/L. These materials contain proteins, protein hormones, antigens, vaccines, antisera, blood products, or nucleic acids.

In these instances, the true amount-of-substance concentration expressed in several moles of analyte molecules per unit of volume and its homologs in patient samples is unknown. However, the same reference material is used globally to measure clinical studies' systems and procedures, determining reference intervals and decision limits. In that case, these materials can still be used as reference materials for calibration. Suppose more than one such material is used for measuring the same measurand. In that case, international harmonization protocols must be used to harmonize results for harmonized clinical guidelines to be applied appropriately (2, 21).

Traceability to international harmonization protocol

Equivalence between measuring systems is commonly not obtained even though international conventional reference materials are used. International harmonization protocols and commutable international harmonization reference materials are used to harmonize measurement results between two or more measuring systems, in particular when measurands are only traceable to the manufacturer´s internal arbitrarily defined reference material(s). A set of natural human samples are then used in the harmonization efforts to include as many factors influencing routine measurement results as possible when catering for equivalence between the final measurement results.

Two or more in-vitro diagnostic measuring systems are *harmonized* regarding measurements of a specific analyte when equivalence is achieved among the quantity

values measured in natural patient samples. Equivalence may be found amongst measuring systems when international conventional calibrators are used, but this is no guarantee that equivalence and harmonization are present amongst natural patient samples. Therefore, traceability of complex and heterogenous macromolecules quantitative characteristics to international conventional calibrators is still needed using “All Procedure Trimmed Mean” (APTM, expressed in IU) as a surrogate reference measurement procedure. The APTM is derived from a dedicated comparison study data with natural human samples. As many relevant routine measuring systems as possible participate in (32), and all contribute to determining the values of the international harmonization reference sample by their contributions to the averages. The method comparison study is also known as “split sample multiple method comparison study” (21).

Using the harmonization protocol, the APTM targeted panel of patient samples is crucial in the harmonization process. Its primary advantage is that it is composed of samples containing the typical variants of the measurands in blood plasma. They are commutable by definition and representative of the patient samples commonly encountered as possible.

Notably, after being assigned with the APTM is derived from the measurement values from measuring systems calibrated by international conventional calibrators. The panel will be a new appropriate traceability reference since the international conventional reference material unit is transferred to the panel. Note that for the process to be successful, the calibration of the new harmonization reference materials needs to be sustainable. There must be a correlation between the involved measuring systems and the international harmonization reference materials. The international unit is transferred from the first panel to the follow-up panels via analogous method comparison studies using an appropriate number of natural patient samples. However, the first panel remains the reference panel (32, 40).

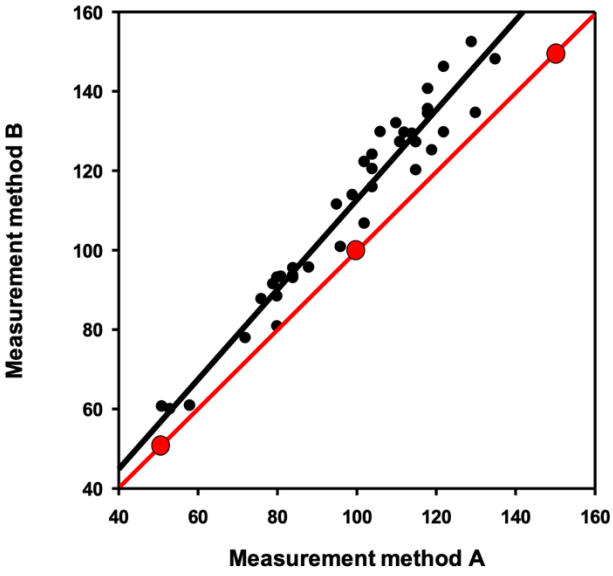


Figure 10: Non-commutable reference material depicted in red while natural patient samples are shown in black.

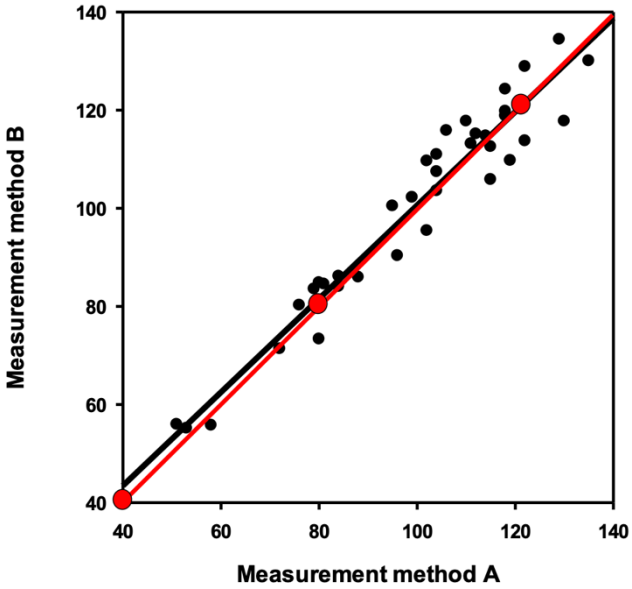


Figure 11: Commutable reference material depicted in red while natural patient samples are shown in black.

Calibration hierarchies and standardization using harmonization

The term *standardization* is here used to mean achieving equivalent results among different measuring systems by having calibration traceable to higher-order references. The term *harmonization* refers to achieving standardization by having calibration traceable to an international harmonization protocol as the highest level of metrological

traceability when there are no entirely appropriate certified reference materials or reference procedures for a given measurand. This means that harmonization is one of the means to accomplish standardization and not an alternative to the standardization (41, 42).

The standard ISO-17511:2020 describes six comprehensive calibration hierarchies (CH1 to CH6) of reference measuring systems that fulfill the requirement for metrological traceability of calibration to "higher-order references" (23). The first three of them have been practiced for decades and require the availability of appropriate reference materials and reference measuring systems. The remaining three hierarchies are new and use harmonization protocols and intend to fulfill the demands for traceability for measurands whether reference materials, conventional calibrators, or reference measurement procedures are available.

Calibration hierarchy	Primary reference material traceable to SI	Reference measure-ment procedure	Defined by a reference measurement procedure	Reference measurement procedure calibrated with a particular primary calibrator traceable to SI	An international conventional calibrator or a certified reference material with a consensus-based protocol for value assignment is available	Other type of common reference is available
CH1						
CH2						
CH3						
CH4						
CH5						
CH6						

Table 2: The six types of calibration hierarchies defined by ISO-17511:2020 (2) (depicted as table rows) and the six categories of criteria (shown as columns) defining each category. Which criteria define a specific calibration hierarchy is displayed using colored squares.

Hierarchy number	Characteristics

CH1	For measurands where both primary reference material(s) and reference measurement procedure(s) with full metrological traceability to SI are available.
CH2	For measurands defined by a primary reference measurement procedure with metrological traceability to SI, but where no primary reference material for the quantity traceable to SI is available.
CH3	For measurands defined by a reference measurement procedure calibrated with a particular primary calibrator traceable to SI.
CH4	For measurands where an international conventional calibrator with a consensus-based protocol for value assignment is available but there is no reference measurement procedure and primary reference materials or primary calibrators and no traceability to SIs.
CH5	For measurands for which neither a reference measuring system nor a certified reference material or international conventional calibrator is available and traceability is supported by by an international harmonization protocol..
CH6	For measurands where metrological traceability to a common reference is not possible. Traceability can therefore only be to the internal calibrator chosen by the manufacturer of a measuring system. A consensus harmonization protocol is needed for this situation to make the results functionally equivalent among different measuring systems- and methods when analyzing measurands in human samples (43-45). This situation evidently includes measurands that cannot be standardized using traceability schemes available in categories #1-#5. Standardization of the results from measuring system results based on such a harmonization protocol provides metrological traceability of the calibrators used in a particular measuring system to that protocol. Standardization using a global harmonization protocol requires involvement and administration by an authoritative international body (e.g. the IFCC Scientific Committee or the ICHCLR part of the IFCC) to achieve equivalence among results for different measuring systems in order to meet requirements for use of the results in medical decisions.

Table 3: The six types of calibration hierarchies defined by ISO 17511:2020 (2) (shown as table rows) and the characteristics of each calibration hierarchy.

Using a harmonization protocol requires both a standard for the purpose and involvement and administration by an authoritative international body. The standard is already published – *ISO-21151:2020* - In vitro diagnostic medical devices – Requirements for international harmonization protocols establishing metrological traceability of values assigned to calibrators and human samples (21). This standard is applicable in cases #5 and #6 and when certified reference materials or international conventional calibrators exist but are not fit-for-purpose. For example, they are not commutable with natural patient samples.

Small, homogenous, and structurally stable molecules including electrolytes, glucose, and steroid hormones can be standardized traceable to SI. Enzymes are standardized using reference measuring systems and international conventional calibrators even though no pure substance reference materials are traceable to SI. Other large molecules, including hemostatic factors, proteins, protein hormones, viral genomes, or envelope proteins, need more complex standardization strategies.

Are the main challenges in traceability in Laboratory Medicine already solved?

A substantial portion of the users of laboratory results expects that challenges regarding traceability are already appropriately solved, even though the actual situation, unfortunately, is not up to such expectations. This unfortunate fact was evident in the years after the Second World War when the laboratories themselves made most of their reagents, as shown by Belk and Sunderman in 1947 (46-48). Fortunately, the problem persists to a minor degree in our time when the in vitro diagnostic industry has shouldered most of the production of reagents from the laboratories (49-54).

Traceability will only be solved when it can be shown to be solved. Thus, there is a vital role for traceable, commutable EQA programs, feeding back to clinicians, laboratories, manufacturers, regulators, and accreditors, allowing the celebration of the benefits of success and improvements in the areas where needed.

References

1. JCGM. International vocabulary of metrology — Basic and general concepts and associated terms (VIM 3): Bureau International des Poids et Mesures; 2012 [3 edition:[Available from: https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf.
2. ISO. ISO 17511:2020 In vitro diagnostic medical devices — Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples. Technical Committee : ISO/TC 212 Clinical laboratory testing and in vitro diagnostic test systems Geneva, Switzerland: International Organization for Standardization; 2020.
3. Hoelzel W, Weykamp C, Jeppsson JO, Miedema K, Barr JR, Goodall I, et al. IFCC reference system for measurement of hemoglobin A1c in human blood and the national standardization schemes in the United States, Japan, and Sweden: a method-comparison study. Clinical chemistry. 2004;50(1):166-74.
4. Jeppsson JO, Kobold U, Barr J, Finke A, Hoelzel W, Hoshino T, et al. Approved IFCC reference method for the measurement of HbA(1c) in human blood. Clinical Chemistry and Laboratory Medicine. 2002;40(1):78-89.
5. Jonsson M, Carlson J, Jeppsson JO, Simonsson P. Computer-supported detection of M-components and avaluation of immunoglobulins after capillary electrophoresis. Clin Chem. 2001;47(1):110-7.
6. Helander A, Eriksson G, Stibler H, Jeppsson JO. Interference of transferrin isoform types with carbohydrate-deficient transferrin quantification in the identification of alcohol abuse. Clinical chemistry. 2001;47(7):1225-33.
7. Finke A, Kobold U, Hoelzel W, Weykamp C, Miedema K, Jeppsson JO. Preparation of a candidate primary reference material for the international standardisation of HbA1c determinations. Clinical chemistry and laboratory medicine : CCLM / FESCC. 1998;36(5):299-308.
8. Eckerbom S, Bergquist Y, Jeppsson J-O. Improved method for analysis of glycated haemoglobin by ion exchange chromatography. Ann Clin Biochem. 1994;31:355-60.
9. Jeppsson J-O, Kristensson H, Fimiani C. Carbohydrate-deficient transferrin quantified by HPLC to determine heavy consumption of alcohol. Clinical chemistry. 1993;39:2115-20.
10. Belanger BC. Traceability: An Evolving Concept. ASTM Stand News 1980;8:22-8.
11. Ehrlich CD, Rasberry SD. Metrological timelines in traceability. Metrologia. 1997;34(6):503-14.
12. Ehrlich CD, Rasberry SD. Metrological timelines in traceability. Journal of Research of the National Institute of Standards and Technology. 1998;103(1):93-105.
13. Mari L, Narduzzi C, Nordin G, Trapman S. Foundations of uncertainty in evaluation of nominal properties. Measurement. 2020;152.
14. Nordin G, Dybkaer R, Forsum U, Fuentes-Arderiu X, Pontet F. Vocabulary on nominal property, examination, and related concepts for clinical laboratory

sciences (IFCC-IUPAC Recommendations 2017). Pure Appl Chem. 2018;90(5):913-35.

15. Wegscheider W. The relationship between traceability and measurement uncertainty. Chimia. 2003;57(1-2):13-.
16. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. Clinica chimica acta; international journal of clinical chemistry. 2014;432:55-61.
17. De Bièvre P, Günzler H. Traceability in chemical measurement. Berlin ; London: Springer; 2004.
18. Possolo A, Bruce SS, Watters RL. Metrological Traceability Frequently Asked Questions and NIST Policy. 2021. Contract No.: NIST Technical Note 2156.
19. Ellison SLR, Williams A. Eurachem/CITAC Guide: Metrological Traceability in Chemical Measurement. A guide to achieving comparable results in chemical measurement. https://eurachem.org/images/stories/Guides/pdf/ECTRC_2019_EN_P1.pdf. Eurachem and CITAC - Cooperation on International Traceability in Analytical Chemistry; 2019.
20. Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. Clin Biochem. 2009;42(4-5):236-40.
21. ISO. ISO 21151:2020 In vitro diagnostic medical devices — Requirements for international harmonisation protocols establishing metrological traceability of values assigned to calibrators and human samples. Technical Committee : ISO/TC 212 Clinical laboratory testing and in vitro diagnostic test systems. Geneva, Switzerland: International Organization for Standardization; 2020.
22. Bossuyt P, Evaluation EWGT. What type of clinical evidence study designs are required in the new IVDR era? Clinica Chimica Acta. 2019;493:S747-S.
23. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU [Internet]. 2017. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0746&from=EN.
24. NIST. Traceability 2020 [Available from: https://www.nist.gov/traceability/traceability-table-contents.
25. Eurachem. The Fitness for Purpose of Analytical Methods (2014) https://www.eurachem.org/index.php/publications/guides/mv. Eurachem; 2014.
26. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. Clinical chemistry. 2018;64(3):455-64.
27. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design. Clinical chemistry. 2018;64(3):447-54.

28. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC Working Group Recommendations for Assessing Commutability Part 3: Using the Calibration Effectiveness of a Reference Material. Clinical chemistry. 2018;64(3):465-74.

29. Thienpont LM, Faix JD, Beastall G. Standardization of FT4 and harmonization of TSH measurements - a request for input from endocrinologists and other physicians. Clinical Endocrinology. 2016;84(2):305-6.

30. 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 3: Total Thyroxine and Total Triiodothyronine. Clinical chemistry. 2010;56(6):921-9.

31. 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 2: Free Thyroxine and Free Triiodothyronine. Clinical chemistry. 2010;56(6):912-20.

32. 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. Clinical chemistry. 2010;56(6):902-11.

33. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? JAMA. 1999;282(8):790-5.

34. NCBI Reference Sequence: NC_001802.1 [Internet]. 2022. Available from: https://www.ncbi.nlm.nih.gov/nuccore/NC_001802.1.

35. HIV Guidelines [Internet]. 2022. Available from: https://cdn.hivguidelines.org/wp-content/uploads/20210611140723/NYSDOH-AI-Virologic-and-Immunologic-Monitoring_6-11-2021_HG.pdf.

36. WHO Library Cataloguing-in-Publication Data: Technical and operational considerations for implementing HIV viral load testing: interim technical update [Internet]. World Health Organization. 2014. Available from: https://apps.who.int/iris/bitstream/handle/10665/128121/9789241507578_eng.pdf?sequence=1.

37. ILAC. ILAC Policy on Traceability of Measurement Results. 2002.

38. Dybkaer R, Storring PL. Application of Iupac-Ifcc Recommendations on Quantities and Units to Who Biological Reference Materials for Diagnostic Use. Eur J Clin Chem Clin. 1995;33(9):623-5.

39. Dybkaer R. Quantities and units for biological reference materials used with in vitro diagnostic measuring systems and antibodies. Scand J Clin Lab Invest. 1996;56:385-91.

40. 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 2: free thyroxine and free triiodothyronine. Clinical chemistry. 2010;56(6):912-20.

41. Myers GL, Miller WG. The roadmap for harmonization: status of the International Consortium for Harmonization o Clinical Laboratory Results. Clinical Chemistry and Laboratory Medicine. 2018;56(10):1667-72.

42. Miller WG, Myers GL, Lou Gantzer M, Kahn SE, Schonbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clinical chemistry. 2011;57(8):1108-17.

43. Miller WG. Harmonization: Its Time Has Come. Clinical chemistry. 2017;63(7):1184-6.

44. Miller WG, Eckfeldt JH, Passarelli J, Rosner W, Young IS. Harmonization of test results: what are the challenges; how can we make it better? Clinical chemistry. 2014;60(7):923-7.

45. Miller WG, Myers GL. Commutability still matters. Clinical chemistry. 2013;59(9):1291-3.

46. Beek WP, Sunderman FW. A survey of the accuracy of chemical analyses in clinical laboratories. Manit Med Rev. 1948;28(4):215.

47. Belk WP, Sunderman FW. A survey of the accuracy of chemical analyses in clinical laboratories. Am J Clin Pathol. 1947;17(11):853-61.

48. Belk W, Sunderman FW. A survey of the accuracy of chemical analyses in clinical laboratories. By William P. Belk and F. William Sunderman, 1947. Arch Pathol Lab Med. 1988;112(4):320-6.

49. Jansen R, Schumann G, Baadenhuijsen H, Franck P, Franzini C, Kruse R, 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. Clinica chimica acta; international journal of clinical chemistry. 2006;368(1-2):160-7.

50. Cobbaert C, Weykamp C, Franck P, de Jonge R, Kuypers A, Steigstra H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples--five year experience from the Netherlands. Clinica chimica acta; international journal of clinical chemistry. 2012;414:234-40.

51. Jansen R, Jassam N, Thomas A, Perich C, Fernandez-Calle P, Faria AP, et al. A category 1 EQA scheme for comparison of laboratory performance and method performance: An international pilot study in the framework of the Calibration 2000 project. Clinica Chimica Acta. 2014;432:90-8.

52. Infusino I, Frusciante E, Braga F, Panteghini M. Progress and impact of enzyme measurement standardization. Clinical Chemistry and Laboratory Medicine. 2017;55(3):334-40.

53. Infusino I, Schumann G, Ceriotti F, Panteghini M. Standardization in clinical enzymology: a challenge for the theory of metrological traceability. Clinical chemistry and laboratory medicine : CCLM / FESCC. 2010;48(3):301-7.

54. Braga F, Frusciante E, Infusino I, Aloisio E, Guerra E, Ceriotti F, et al. Evaluation of the trueness of serum alkaline phosphatase measurement in a group of Italian laboratories. Clinical Chemistry and Laboratory Medicine. 2017;55(3):E47-E50.