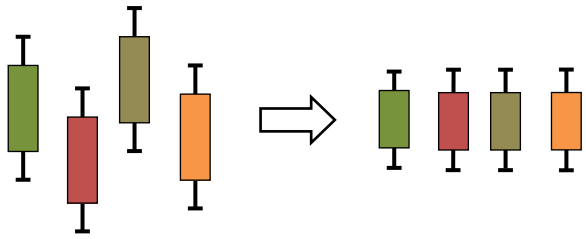


International Consortium for Harmonization of Clinical Laboratory Results

Toolbox of technical procedures to be considered when developing a process to achieve harmonization for a measurand

The procedures described here were developed by the following task force:
Cas Weykamp, Chair, Queen Beatrix Hospital, Winterswijk, the Netherlands
John Eckfeldt, University of Minnesota, USA
Hubert Vesper, Centers for Disease Control and Prevention, USA
Linda Thienpont, University of Ghent, Belgium
Chris Burns, National Institute for Biological Standards & Control, UK
Angela Caliendo, Brown University, USA
Thomas Ciesiolka, Roche Diagnostics, Germany
Joe Passarelli, Roche Diagnostics, Germany

Contents

Introduction

Background
Results between different clinical laboratory measurement procedures should be equivalent, within clinically meaningful limits, to enable optimal use of clinical guidelines for disease diagnosis and patient management. In October 2010 the AACC convened a conference to address how to improve harmonization of laboratory test results for which there are no higher-order reference measurement procedures, and for which it was unlikely that such procedures could be developed. The major outcome of the conference was that a systematic approach to harmonization must be developed that would identify measurands for which harmonization is needed, prioritize the measurands based on clinical importance and technical feasibility, and organize the implementation of harmonization activities by all interested stakeholders on a global basis.

Following the conference, a Steering Committee was formed to develop this new entity. The Steering Committee formed three task forces (TF) to develop the recommendations: TF 1 to develop an operational structure and administrative procedures for the International Consortium and the Harmonization Oversight Group (HOG); TF 2 to develop a toolbox of technical procedures to address procedures to harmonize a measurand when a reference measurement procedure is not available; and TF 3 to develop checklists for prioritization and feasibility to harmonize a measurand. This document presents the recommendations of TF 2.

Task force 2 has created a toolbox of well developed generic processes as a starting point for use by a Harmonization Implementation Group (HIG) to achieve harmonization for a particular measurand. The processes described here are intended to be used and modified as needed for harmonization of a specific measurand that has no reference measurement procedure. The processes include experimental and mathematical components.

In part 1 the parameters to be determined in the process of harmonization are described. In part 2 options to address each of these parameters are described. In part 3 two detailed protocols are described. It is expected that further evolution of the protocols can be expected based on the experience of HIG's with different measurands.

Part 1. Parameters to be determined

Introduction

Results between different clinical laboratory measurement procedures should be equivalent, within clinically meaningful limits, to enable optimal use of clinical guidelines for disease diagnosis and patient management. When results are not equivalent, they should be made equivalent through harmonization. Achievement of harmonization depends on a number of technical parameters that should be determined in an assessment study. If the assessment demonstrates feasibility to achieve harmonization, then a harmonization effort should be initiated and its success monitored in a structured way.

Definitions

Variability of results
The degree of variability in results of a laboratory test, expressed as the intermethod (interassay) CV

Clinically meaningful limits
The maximum degree of variability in results of a laboratory test that allows optimal patient care

Equivalent results
Agreement among results of a laboratory measurement procedure that does not exceed the clinically meaningful limits

Harmonization
A process that reduces the variability of results of a laboratory measurement procedure to a level below the clinically meaningful limits

Assessment study
A study to investigate whether it is achievable, through harmonization, to reduce the variability of a laboratory measurement procedure to a level below the clinical meaningful limits.

Parameters related to properties of tests
Properties of both the laboratory test and the measurand with an impact on the variability of results of laboratory measurement procedures.

Parameters related to requirements of harmonization tools
Properties of both the laboratory measurement procedure and the measurand related to requirements of tools needed for harmonization

Harmonization effort
All structured actions to reduce the variability of results of a laboratory measurement procedure to a level within clinical meaningful limits.

Harmonization success
Proof that variability of results of a laboratory measurement procedure is within clinical meaningful limits.

Calibrator
The term calibrator refers to “International Conventional Calibrator” according to ISO 17511, Category 4, Section 5.5.

Parameters to be determined

The parameters that should be determined to evaluate whether harmonization can be achieved right away, or to determine the issues that prevent immediate implementation of harmonization, are categorized into two groups: a) parameters related to properties of the measurement procedures and b) parameters related to harmonization tools.

Parameters related to properties of measurement procedures

These are the parameters that contribute to variability in the result of a laboratory measurement procedure.

1. Reproducibility within one measurement procedure
Variability of results within one measurement procedure can derive from within laboratory imprecision, batch to batch variability of reagents or calibrators, lack of robustness resulting in between laboratory, between instrument, between operator and between environmental conditions differences.

2. Linearity of methods
In ideal measurement procedures, there is a linear relation between the measured quantity and the concentration of the measurand. A non-linear relationship will have an impact on equivalence of results of different measurement procedures. The term “predictability of analyte target concentration” is used in the assessment study to evaluate linearity.

3. Heterogeneity
For many analytes, the measurand is heterogeneous (e.g. different epitopes in immunoassays) and in many measurement procedures the reagent is heterogeneous (e.g. reacts with different epitopes in immunoassays). This heterogeneity in both measurand and reagent should be taken into account. This heterogeneity cannot be measured directly but can be derived from the scatter of data observed when plotting results from 2 different measurement procedures. The influence of interfering compounds (lack of specificity) can also be different for different measurement procedures and contribute to variation.

4. Calibration
Variability in results of a laboratory measurement procedure can derive from differences in calibration among different procedures. Such differences are typically reflected by simple relations like $Y = aX$ or $Y = aX + b$.

5. Overview
The assessment should clarify the contribution of each of these parameters to the variability. If the major source of variability is calibration, harmonization can be achieved right away (using a simple factor). If the major source is a non-linear relationship, harmonization - although less simple - can be achieved with appropriate statistical methods to manage non-linear relationship factors. If either reproducibility or heterogeneity is the major source of variability for one or more measurement procedures, harmonization is likely not possible until the affected measurement procedures are changed.

Parameters related to Harmonization Tools

When the properties of the measurement procedures allow harmonization, one can conclude that harmonization is in principle possible. But that is not enough. Appropriate tools must be available to do the harmonization. The tool used is a sample or a set of samples (termed an International Conventional Calibrator, referred to as a calibrator in the following text) with assigned values for the quantity of measurand. Note that a set of patient samples may provide the function of an International Conventional Calibrator. There are crucial technical and practical parameters that should be considered when assessing the technical feasibility of achieving harmonization.

6. Commutability
Commutability of the calibrator is a technical requirement: the calibrator should behave as a patient sample in all measurement procedures for which it is intend to be used.

7. Stability
It must be possible to store the calibrator for a long time.

8. Sustainability
Once a batch of calibrator is finished, the next batch should have the same properties to maintain the continuity of harmonization.

9. Value Assignment
There must be a sustainable and generally accepted way to assign a target value to the calibrator.

10. Availability
The availability of sufficient amounts of the calibrator must be ensured.

11. Costs
Costs of calibrators must be acceptable.

12. Will to Harmonize
All parties involved must support the harmonization effort (commercial aspects such as marketing and costs of implementing and maintaining harmonization can be limiting factors).

Other Aspects

13. Is Harmonization Achievable?
Ideally the assessment, considering measurement procedure properties and availability of tools, should identify the current and potentially achievable equivalence among measurement procedure results in comparison to the desirable equivalence. On the basis of the achievability assessment, one can determine if a harmonization effort is worth doing or not.

There is some overlap with the "feasibility" assessment conducted by the HOG with the assistance of a Special Working Group (SWG). This decision would occur before a HIG was formed. However, it may be that there will not be enough information available for the HOG to make a determination of feasibility, and some experiment would then be needed. So an experimental design to assess feasibility is provided.

14. Effort
Once there is agreement that harmonization is technically possible, an effort can start to implement the harmonization.

15. Success
Once harmonization is implemented, success of harmonization should be monitored. The expectation is that the HIG will develop a strategy to have a monitoring scheme developed. The actual monitoring task would typically fall to a PT/EQA organization to implement. However, technical advice and suggestions how to develop the materials to use for monitoring would come from the HIG. The toolbox includes the technical items to consider and recommend to a PT/EQA provider on how to implement a surveillance program for a given measurand.

Figure 1 summarizes the parameters and aspects to be addressed to achieve harmonization and thus the tools described here.

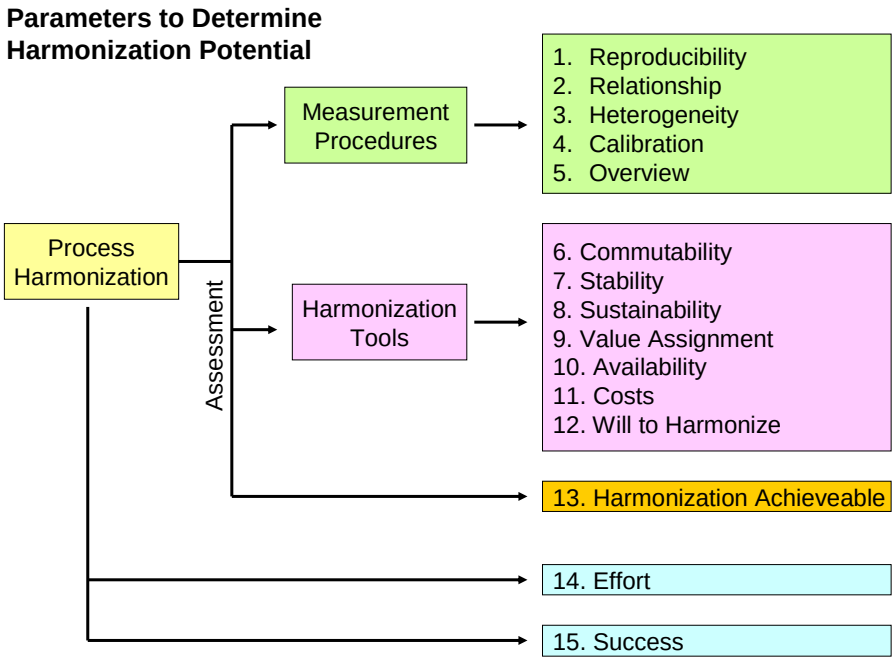


Figure 1. Parameters to be determined when considering harmonization.

Part 2. Options to address the Parameters

In this part the preceding 15 parameters and aspects of the harmonization process are considered in more detail. Each section starts with considerations, then potential paper options to address the parameter are listed. Paper options are ways to get information without doing experiments (e.g. scientific literature, manufacturer instructions for use or calibrator certificate of analysis). Paper options are inexpensive because they are based on existing information. However, the available information needs to be critically reviewed to ensure the data were derived from sound experimental designs and are reliable. In addition, a critical assessment must be made if data from different sources can be compared.

Experimental options are ways to get information via experiments. There are two experimental options described in part 3: a) the Integrated Harmonization Protocol (IHP), a toolbox with 15 tools to address each of the 15 parameters; and b) the Step-up Design for Harmonization (SDH). Experimental options will give reliable information and in many cases are necessary to address harmonization of a given measurand. However, an assessment needs to be made if the costs and time required are acceptable. Such an assessment will be based on the clinical importance of harmonizing a measurand and will be reflected in the priority assigned by the HOG.

1. Reproducibility

Considerations

Many variables have an impact on the variability within one measurement procedure:

- measurements within one analytical run or between analytical runs
- measurements performed in one laboratory or in several laboratories
- measurements performed using one instrument or using several instruments
- measurements performed with one batch of reagents (or calibrators; or other consumables) or with different batches
- measurements performed by one operator or by more operators
- measurements performed with fresh reagents or stored reagents
- measurements performed in a standardized environment or in different environments
- etc.

A HIG should be aware of all these variables but also decide which ones should be included in the assessment study. Added value of the information and efforts in time/costs should be balanced. Information can be derived from paper options and experimental options. Paper options are cheap and easy but should be reviewed critically: i.e. was the experimental design reliable and can data from different sources be compared? Experimental options are reliable but are the costs and time acceptable?

Paper Options

- Literature; papers in which reproducibility of tests is reported
- Information from Manufacturers; reproducibility specified in instructions for use
- EQA programs; when duplicates are included in the design, and commutability does not confound the data

2. Linearity

Considerations

Harmonization is relatively easy when there is a linear relationship between the reported units of the respective measurement procedures and the concentration of the measurand. If not linear, it is important to know that there is a non-linear relation for one or more of the measurement procedures and preferably also the reason (e.g. only at highest concentration, S-shaped curve, convex, concave, other).

Paper Options

- Literature; papers in which relationship is reported
- Information from Manufacturers; relationship specified in instructions for use

- EQA programs; when linearity panel is included in the design, and commutability does not confound the data

3. Heterogeneity

Considerations

Heterogeneity is a very important parameter. Harmonization is only possible when the respective measurement procedures measure the same measurand (or at least measurands that occur in proportion, e.g. glycated hemoglobins and HbA1c). For many measurement procedures (e.g. immunoassays) this will not be the case. The measurand may not be defined (e.g. the epitope of clinical interest). If different measurands are measured, their relationship may not be proportional in different patients (e.g. one patient has many epitopes X, while another patient has many epitopes Y). The measurement procedures do not measure the same measurand or do not measure in the same proportion (e.g. procedure1 A has high reactivity towards epitope X, method B has a high reactivity towards epitope Y). These phenomena are referred to as heterogeneity. The more heterogeneity, the more difficult to harmonize results of measurement procedures. An additional aspect is the specificity of measurement procedures: they can be more or less susceptible to interference by other compounds than the measurand (e.g. icteric, hemolytic, drugs, proteins with nearly the same structure).

In experimental options the HIG should decide whether or not to include specificity. Inclusion of specificity implies collection of many rare samples and may take a lot of time, effort and cost. Heterogeneity cannot be measured directly but must be derived from the scatter of data as seen in X/Y plots of results from different measurement procedures.

Paper Options

- Literature; published X/Y and difference plots
- Information from manufacturers; instructions for use
- EQA: scatter when a number of single donation samples are in the design

4. Calibration

Considerations

Variability in results of a laboratory test can derive from differences in calibration among different measurement procedures. In cases when calibration can be traced back to either a (reference) measurement procedure or a (reference) material, if different reference measurement procedures or materials are used for different measurement procedures, there will be a difference in calibration and thus in results for patient samples. Measurement procedures traceable to a non-commutable reference material used as a common calibrator may also have differences in results for patient samples.

Paper Options

- Literature; slope and intercept of published X/Y and difference plots
- Information from manufacturers; test principle or traceability to reference measurement procedures or to reference materials.
- EQA: slope and intercept in multiple sample design when commutable samples are used

5. Review parameters 1 to 4

Considerations

If the major source of variability is calibration, harmonization can be achieved right away (using simple factors). If the major source is a non-linear relationship, harmonization - although less simple - can be achieved with appropriate statistical methods to manage non-linear relationship factors. If either reproducibility or heterogeneity is the major source of variability for one or more measurement procedures, harmonization is likely not possible until the affected measurement procedures are changed.

The assessment should summarize the contribution of each of these parameters to the variability. The assessment can be in qualitative terms: if a huge scatter is seen in X/Y plots this does not need to be quantified – it is clear that harmonization is impossible. If scatter is not too bad, a quantitative

approach is applied: harmonized variation can be estimated with virtual harmonization (recalculation of the results of the respective measurement procedures using an assigned value to the samples (e.g. mean of all measurement procedures for each sample on the x-axis; individual measurement procedure results on the y-axis; calculate y = ax + b; use this equation to calculate harmonized results for each measurement procedure). Although not required, a full breakdown of the total variation in individual measurement procedure results into the contributions of the four respective parameters may be interesting: e.g. reproducibility, relationship, heterogeneity and calibration contribute U%, V%, W% and X%, respectively, to the total variation.

This estimated harmonized variation (HV) can be compared with the desired variation (DV). Then the HIG can make a decision:

- Qualitative assessment showing huge heterogeneity: harmonization not possible; useless to investigate parameters 5 to 13
- HV > DV: harmonization not possible to a degree that warrants investigation of parameters 5 to 13
- HV < DV: harmonization possible; relevant to continue assessment with parameters 5 to 13
- HB ~ DV: harmonization may be possible; HIG to decide whether to go on with parameters 5 to 13 or not.
- Complicated result: some measurement procedures can, others cannot be harmonized.

Paper Options

- Literature report of such an investigation

Remark
Data evaluation may reveal that although harmonization of all measurement procedures is not possible, partial harmonization can be achieved when one or two measurement procedures are left out. No general rule can be given for whether this is an acceptable option or not. It is up to the HIG for a specific measurand to consider and decide.

6. Commutability

Considerations
The reference material used as a common calibrator should behave like a patient material in all measurement procedures. Or at least the magnitude of the non-commutability should be defined, in which case it may be possible to develop a correction factor. Basis of any approach to evaluate commutability: assay candidate calibrators and a number of patient samples with two measurement procedures and plot the results. When a candidate calibrator is on the same line as the patients results (or in the same cloud of dots on an x/y or difference plot), the candidate calibrator is commutable for both measurement procedures. Note that commutability must be evaluated for all combinations of measurement procedures for which the calibrator is intended to be used.

Paper Options

- Literature: have candidate calibrators already been investigated?
- Information from Manufacturers/Reference Institutions: specifications in instructions for use.
- EQA: if candidate calibrators are included as EQA samples along with commutable patient samples

Experimental Options

- CLSI EP30-A (formerly C53-A) in addition to the IHP or SDH protocols

7. Stability

Considerations.
Stability of the calibrator is a prerequisite: it must be possible to store the material for a long time without change in measurand concentration. “Long time” must be defined. Pharmacopeia reference materials are made for 3-5 years; international standards for 10-15 years. The storage time has an impact on their presentation, i.e. lyophilized versus frozen. One has to distinguish storage stability and in-use stability. Stability can be estimated in a relatively short period with an accelerated stability test but there should always be a formal stability test performed in parallel.

Paper Options

- Literature: has stability of candidate calibrators been investigated?
- Information Manufacturers/Reference Institutions: data specified or at least expiration terms in combination with storage conditions?
- EQA: if candidate calibrators were included in an EQA program along with commutable patient samples in a number of consecutive years.

8. Sustainability

Considerations.

To maintain harmonization over many years, it must be established that calibrators remain available. New batches must have the same properties as the original batches. Comparability of batches should be assessed. An important aspect is also how values will be transferred from batch-to-batch to prevent drift.

Paper Options

- Literature: papers dealing with different batches?
- Information Manufacturers/Reference Institutions: data available?
- EQA: if candidate calibrators were included in an EQA program along with commutable patient samples in a number of consecutive years.

9. Value Assignment

Considerations

A calibrator can only be applied as such when a value has been assigned for the measurand of interest.

Paper Options

- Literature: a reference or designated comparison procedure is known and was used to assign the value

10. Availability

Considerations

For a successful harmonization effort sufficient amounts of calibrator must be available for a long time period.

Paper Options

- Literature: sources for materials are reported
- Manufacturers/Reference Institutions: materials are in their catalogue

11. Costs

Considerations

For a successful harmonization effort, costs for calibrators or for a harmonization process must be affordable. This is true for the supplier of a material (an investment to produce a calibrator must have a financial return in sales of that calibrator) and for the user (the unit cost must be reasonable when considered along with the total cost to perform a harmonization process)

Paper Options

- One or more stakeholders supplies funding for a given harmonization project

12. Will to harmonize

Considerations

Although technically possible, a harmonization effort can fail because there is no will to harmonize. Manufacturers may be reluctant because they fear loss of their unique selling proposition or because costs to recalibrate are high. Clinical chemists as well as clinicians may be reluctant because they do not want to change reference or decision values.

13. Harmonization Achievable

Considerations

Once all information on the 12 essential parameters described above is collected, a final decision can be made regarding the technical feasibility to achieve harmonization. If an initial assessment is being made by a Special Working Group, a recommendation regarding the need to collect additional experimental data can be made.

14. Effort

Considerations

There are two options to conduct a harmonization implementation project for a specific measurand: the passive and the active approach. In the passive approach, the HOG collaborates with another group that is addressing harmonization for a particular measurand to make sure that the tools for harmonization (calibrator and value assignment procedure) are registered in appropriate lists (e.g. JCTLM). The final action of the HOG is that stakeholders (manufacturers, EQA/PT organizers) are informed that harmonization tools are available. In the active approach, a Harmonization Implementation Group is formed to conduct the harmonization project and guides the harmonization process until there is proof of harmonization.

Options

- passive approach through collaboration
- IHP protocol
- SDH protocol

15. Success

Considerations

The success of a harmonization effort is demonstrated by evidence that variability of results from different laboratory measurement procedures are within clinically meaningful limits. Success can be monitored passively or actively. The passive approach is not a task of the HIG but can be initiated and coordinated by a HIG in collaboration with EQA/PT organizers in various countries or regions. In general, the EQA/PT approach will use commutable samples to investigate and monitor the variation among routine measurement procedures as the parameter for success of harmonization. In the active approach the HIG organizes periodic studies with the manufacturers to monitor the success of harmonization. Such studies may be for either a limited period of time to monitor the success of implementation, or for a longer period of time to monitor the maintenance of success. A formal certification approach may be considered.

Options

- passive approach (EQA/PT programs)
- active approach (certification program or other monitoring scheme)

Part 3A: Integrated Harmonization Protocol (IHP)

The IHP describes in detail an initial experimental design that may be used by a HIG to approach the harmonization process. Section A deals with the assessment study and section B with the harmonization effort and the monitoring of success.

Initially a core HIG will be formed of a few members to prepare a protocol for harmonization that will be used for understanding the scope of work to be done and to be used to solicit funding for the project. No experiments will be performed by the core HIG. Consequently the toolbox need is for an assessment experiment that, if needed, would be conducted by the full HIG once funding is committed and the project initiated. This assessment protocol is meant to be used by the full HIG but the HOG and core HIG will find the information useful to know what the full HIG will need to do. A SWG may find guidance on how to evaluate information available regarding the technical feasibility to achieve harmonization. The full HIG will build on data collected by the SWG and core HIG to initiate the harmonization project.

Section A. Assessment Study

A1. The link between required information and experimental design of the IHP

Figure 1 on page 6 shows the parameters to be determined. Figure 2 below links the 13 assessment parameters of figure 1 (left side) with the summarized experimental design of the IHP on the right. It can be seen that the experimental design comprises four sample types:

- Yellow: 32 samples from individual persons healthy and diseased
- Green: 5 mixtures made from the samples of the individual persons
- Amber: a linearity panel of 5 samples made from the samples of individual persons
- Blue: 7 calibrators of which 3 are different candidate calibrators, 2 are candidate calibrators of the same type but of a different batch and 2 are candidate calibrators of the same type stored for some time.

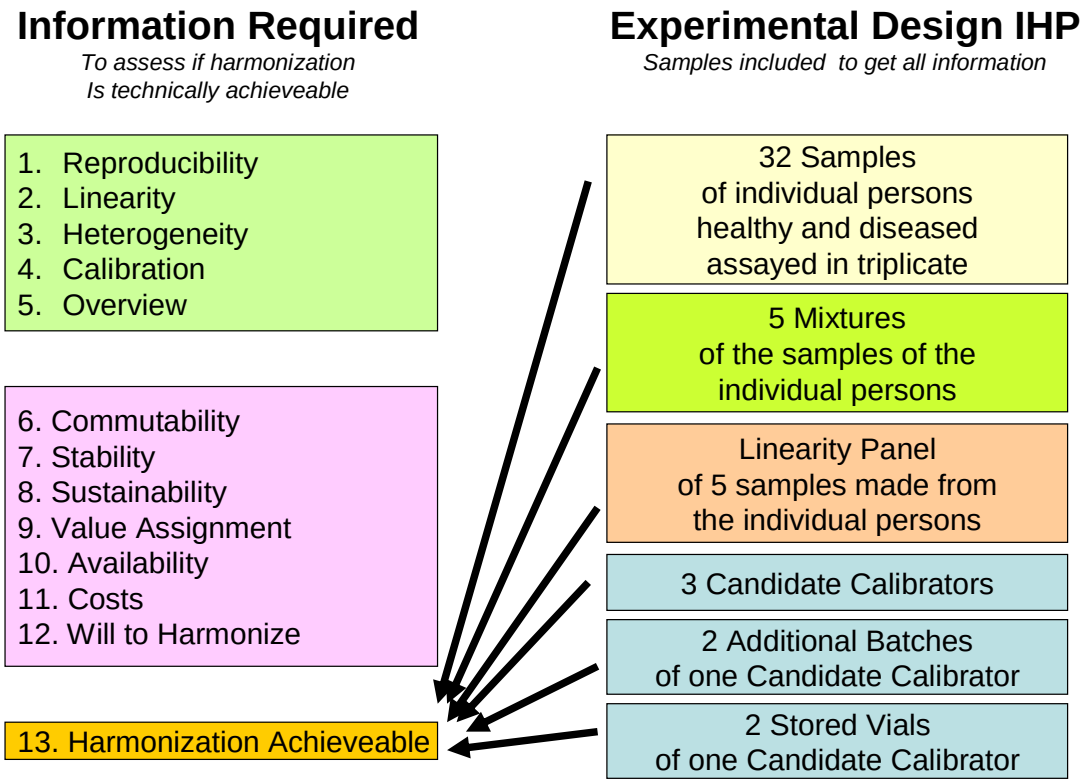


Figure 2. Link between the assessment parameters and the IHP

Remark
It should be stressed that the number of samples is arbitrarily chosen. Depending on the measurand and the limitations in cost of measurements, availability of samples and desired uncertainty in the information, the numbers can be changed: more or fewer samples, single measurements instead of triplicates etc. It is recommended to perform the whole experiment in one run (efficient and reliable), however, it is also possible to do the experiments spread over a longer period of time.

A2. The Experimental Design in Detail

Table 1 shows the experimental design in more detail. The essentials are:

- measure 47 samples in triplicate with each of the relevant measurement procedures
- evaluate the results to generate all required information for the assessment
- and from this information derive whether harmonization is technically achievable (parameter 13)

Table 1. Generic experimental design in detail

Sample Type	Sample Nr.	Description	Results			Results used to assess
			1	2	3	
Individual Persons	1	Person 1				1. Reproducibility 3. Heterogeneity 4. Calibration 5. Overview 6. Commutability 9. Value Assignment 12. Will to harmonize
	2	Person 2				
	…	Person …				
	…	Person …				
	…	Person …				
	31	Person 31				
	32	Person 32				
Mixtures	33	M1/L1 (mixture 1-8)				6. Commutability 10. Availability 11. Costs
	34	M2 (mixture 9-16)				
	35	M3 (mixture 17-24)				
	36	M4/L5 (mixture 25-32)				
	37	M5 (mixture 1-32)				
Linearity Panel	38	L2 (75/25 M1/M4)				2. Relation 5. Overview
	39	L3 (50/50 M1/M4)				
	40	L4 (25/75 M1/M4)				
Candidate Calibrators	41	Candidate 1				5. Overview 6. Commutability 10. Availability 11.Cost
	42	Candidate 2				
	43	Candidate 3				
	44	Candidate 1 Store T1				7. Stability
	45	Candidate 1 Store T2				
	46	Candidate 1 batch 2				8. Sustainability
	47	Candidate 1 batch 3				

A3 Experimental design for a specific measurand

The experimental design for a specific measurand starts with reviewing aspects and making choices. It cannot be stressed enough that a good design of the IHP is half the work. Making such a design consists of considerations, decisions, making draft designs, review/modify of the design to achieve the final design.

Initial Considerations

Measurement procedures: make an inventory

A HIG should make a list of all routine measurement procedures as well as potential designated comparison measurement procedures available for the measurand. For each measurement procedure, the sample volume to do a single, duplicate and triplicate measurement should be known.

Samples: review what is relevant

A HIG should review which samples are relevant. In general, relevant means that the samples should cover the measuring interval and thus be taken from healthy and diseased persons to include concentrations of the measurand characteristic of absence and presence of the disease or diseases for which it is a biomarker.

Samples: estimate limitations for storage

Ideally the 32 patient samples of the protocol will be collected and analyzed on the same day in the same laboratory. In practice this will nearly always be impossible. It may take several days, weeks or even months to get the samples with the desired concentrations. In addition, measurements may not be done immediately or not in the same place (must be shipped to other labs). Thus, samples must be stored and it is required to verify that storage does not affect the samples. The HIG should estimate the limitations of storage. Example: storage at <-70ºC and shipment on dry ice is common but the validity should be confirmed. This can be done quite simply:

a) take a sample and dispense in four vials,
b) store vial 1 in the refrigerator and freeze the three others at <-70ºC,
c) thaw vial 3 and 4 and freeze them again at <-70ºC,
d) thaw vial 4 and freeze again at <-70ºC,
e) thaw vials 2, 3 and 4,
f) assay the four samples in triplicate with the relevant measurement procedures.

The sample storage and measurements should be done within a reasonable timeframe (24-72 hours). When the measured concentration and precision of the four samples is identical, it can be concluded that storage at <-70ºC is very robust: even three times freeze/thaw (sample 4) does not have an impact. From results of samples 2 and 3 it can be concluded if one freeze/thaw or 2 freeze/thaw cycles have an impact. Evaporation of sample during the test period should be prevented. When a measurand is not stable at refrigerator temperature, an alternate or truncated assessment scheme will need to be developed, perhaps with an immediate measurement of the fresh sample and consideration of run to run imprecision in the assessment of results from the frozen samples.

Samples: calculate volume required

For each measurement procedure in the study, the volume for triplicate measurements plus any dead volume in the volumetric process, and the volume needed to make the mixtures and the linearity panel, plus an allowance for loss in the storage container need to be considered in determining the total volume of sample that will be required to perform the study.

Candidate Calibrators

The HIG should investigate if candidate calibrators exist. If yes, determine if there are one or multiple concentrations of a candidate calibrator and if one or multiple batches are available. Potential sources are organizations for reference materials (e.g. NIBSC, WHO, IRMM, NIST, ReCCs, others), manufacturers of measurement procedures, EQA/PT organizations.

Additional Considerations

The exploration of feasibility may disclose aspects that require modification of the IHP. Not all situations can be foreseen but some are listed below:

- Measurements are very expensive and the number of measurements should be limited; an option is to do single or duplicate measurements. The disadvantage is that limited or no information is gained on precision of the tests. Another option is to include fewer than 32 samples; the disadvantage is that the basis for commutability and inter-individual heterogeneity will be less reliable.
- The volume of sample required is higher than can be collected; an option is to do single or duplicate measurements. The disadvantage is that limited or no information is gained on precision of the tests.
- The number of 32 healthy/diseased individuals is not achievable because diseased persons are very rare or it is complicated to get enough sample from diseased persons. An option is to include fewer samples of diseased persons; the disadvantage is that the basis for commutability and inter-individual heterogeneity will be less reliable.
- The number of 32 healthy/diseased individuals is regarded as too low for reliable information on commutability and inter-individual heterogeneity. An option is to increase the number of samples. The disadvantage is that workload and costs increase.
- It is expected that batches of reagents and/or measurement procedure product calibrators are heterogeneous. An option can be to include more than one batch of reagent or product calibrator per measurement procedure. The disadvantage is that workload and costs increase.
- “Quantitative” has a different meaning in Molecular Diagnostics. Quite often results are in log terms. This requires a specific statistical approach (not worked out here).

- Table 2 shows the effect on sample volume and workload/costs resulting from condensation options of the protocol. The order in the table is also the list of prioritization.

Table 2. Prioritization of options to reduce components and their effects.

Limitation	Reduction in Sample Volume	Reduction in Workload/Costs
Duplicates (thus no triplicates)	33%	22%
Single assays (no triplicates)	67%	44%
Limit number of samples	0%	2% per sample
Leave out linearity panel	8%	6%
Leave out mixtures	8%	6%

First estimation of feasibility

With the information collected above a first estimation of feasibility of the IHP can be made. If storage of samples is possible, collection of samples can be spread over time and measurements can be spread over time and multiple places. If not, collection of samples and assays should be concentrated in one place on one day.

The inclusion of measurement procedures should be considered. If there are only two procedures for a measurand, the choice is not difficult; both are included. But what if there are many (>20) measurement procedures on the market; ideally all should be included. However, that may not be technically possible in terms of sample volume required and in terms of logistics and costs. A realistic approach would be to include 10-15 measurement procedures, representing each of the various analytical principles for the measurand and/or representing the major measurement procedures in the market (can be derived from EQA/PT programs). Once the measurement procedures are chosen, the required volume of sample can be calculated. The HIG must determine if it is feasible to collect the required volume of samples and if laboratories (manufacturers) are willing to perform the measurements.

Summary of Considerations and Decisions

Considerations can be summarized in the table below. On the basis of the decisions, the final design is established.

Table 3. Summary of considerations and decisions for the experimental design of the IHP

Measurement Procedures in the Market				
Measurement Procedure	Measuring Interval	Sample volume single/duplicate/ triplicate		
1	……	……ml	……ml	……ml
2	……	……ml	……ml	……ml
n	……	……ml	……ml	……ml
Remarks				
Candidate Designated Comparison Measurement Procedure(s)				
Measurement Procedure		Sample volume single/duplicate/ triplicate		
1		……ml	……ml	……ml
2		……ml	……ml	……ml
n		……ml	……ml	……ml
Remarks				
Relevant Samples				
Concentration Range healthy individuals Concentration Range diseased individuals				
Remarks				

Sample Storage: Impact of (repeated) freeze/thaw				
Measurement Procedure	Fresh Mean (CV)	Once Fr/Thaw Mean (CV)	Twice Fr/Thaw Mean (CV)	Three Fr/Thaw Mean (CV)
1 2 n				
Remark				

Sample Volume required per sample	
Measurement Procedures involved	Volume required
All Methods; n = 5 Selected methods N selected methods; n =	……ml ……ml ……ml
Remarks	

Candidate Calibrators				
Source	concentrations	batches	ml/vial	price/vial
Metrology Institute etc.				
Remark				

Additional Considerations
- Cost of measurements too high to do triplicates - Required volume too high - Samples of diseased persons limited - Need for more than 32 samples - Need to investigate more than one batch of reagents - ……
Remark

A4 Organization

On the basis of the final design, the HIG organizes the IHP.

Time Schedule

The time to perform the IHP will depend on the availability of samples (how much time will it take to determine sample storage stability and to collect the 32 samples), and of candidate calibrators (how long will it take to get them and allowance should be made to do accelerated stability tests). Once the time to manage the logistical components has been determined, the participating manufacturers and

laboratories can be approached to determine when the measurements can be performed. Note that manufacturers typically require substantial lead time to schedule and budget projects. It is recommended to contact and engage manufacturers early in the planning stages.

Acquisition of Candidate Calibrators

Candidate calibrators are ordered from their manufacturers. Note that candidate calibrators may be prepared as pooled patient samples, some of which may be supplemented with a measurand, based on the technical assessment of the HIG for a particular measurand. Some of them are stored at higher temperatures than standard storage conditions to allow an accelerated stability test. The number of vials to be ordered depends on the number of measurement procedures involved and if they are also used for stability experiments. Example: if 16 measurement procedures are involved, 48 vials of candidate calibrator 1 would be required (16 for standard testing = sample 41, 16 for storage at T1 = sample 44, 16 for storage at T2 = sample 45). For candidate calibrator 2, only 16 vials are required for sample 42. Etc.

Collection of Samples

Samples may be collected at one or more centers. In general, this will imply that, following concentration specifications, one or more labs will collect the samples during a period of weeks or months. Each sample is analyzed with an agreed measurement procedure (to know the approximate concentration), aliquotted as needed and frozen at -70ºC or colder (or under specified conditions for a specific measurand).

Required Volume per Sample

The collected sample will be dispensed in vials for the respective measurement procedures and to make the mixtures and the linearity panel. Rule of thumb: the required volume = N x V x 1.5. N is number of measurement procedures that will be included where V is the volume that will be dispensed per vial and the factor 1.5 reflects the surplus of sample needed to make the mixtures and the linearity panel. Example: if 16 measurement procedures are included and vials of 1 mL will be dispensed the required volume at collection is 16 X 1 X 1.5 = 24 mL. However, it is recommended to determine carefully the actual total volume required based on the specific measurement procedures included and the experimental design to be used. It is common that different measurement procures will have different sample volume requirements.

Measurement Procedures (and candidate Designated Comparison Measurement Procedures)

Measurement procedures are made available. There are several options: a) manufacturers involved are asked to perform the measurements, b) manufacturers are asked to nominate a medical laboratory to perform the measurements, c) the HIG invites laboratories with the relevant measurement procedures, d) all measurement procedures are installed in the same laboratory.

Management

The HIG determines that samples, candidate calibrators and measurement procedures are available.

A5 Manufacture of sample sets

Sample sets are typically prepared by clinical laboratories to allow efficient use of limited amounts of sample volume. Residual samples from clinical laboratory testing may be used or it may be necessary to develop a process to identify and collect samples from a group of volunteers with the appropriate concentrations of the measurand of interest.

NOTE: If the measurand in the clinical sample is not stable to freeze/thaw, alternative ways to make mixtures and linearity panels should be developed that can be implemented within a time interval that allows storage without freezing to make the mixtures and linearity panels. Careful planning and timing of shipments to participating laboratories are essential for successful evaluations under these conditions.

EXAMPLE: The process to prepare samples sets from frozen individual samples is illustrated by the following example which assumes the measurand and sample are stable to freeze/thaw, 16 measurement procedures are involved and 1 mL is needed for each vial.

Step 1: Thaw the collected samples and pipet into tubes.

All 32 collected samples are thawed, mixed and the contents of each sample is pipetted into the tubes shown in Table 4. From these tubes, the generic sample types in Table 1 are prepared. There are 32 tubes for each of the individual samples marked I1 to I32. There are 5 tubes for each of the mixtures (marked M1 to M5) and 3 tubes for each of the linearity panel samples (marked L2 to L4).

For example from Table 4:
Sample 1 is pipetted: 16 mL in tube I1, 2 mL in tube M1, 0.5 mL in tube M5, 1.5, 1.0, 0.5 mL in the tubes L2, L3 and L4.

Sample 2 is pipetted: 16 mL in tube I2, 2 mL in tube M1, 0.5 mL in tube M5, 1.5, 1.0, 0.5 mL in the tubes L2, L3 and L4.

Sample 9 is pipetted: 16 mL in tube I9, 2 mL in tube M2, 0.5 mL in tube M5.

Sample 17 is pipetted: 16 mL in tube I17, 2 mL in tube M3, 0.5 mL in tube M5.

Sample 32 is pipetted: 16 mL in tube I32, 2 mL in tube M4, 0.5 mL in tube M5, 0.5, 1.0, 1.5 mL in tubes L2, L3, L4.

Step 2: Dispense the contents of the tubes into vials

For example from Table 4:
From tube I1, 16 vials labeled 1 are dispensed; 1 mL in each vial

From tube I2, 16 vials labeled 2 are dispensed; 1 mL in each vial

From tube I32, 16 vials labeled 32 are dispensed; 1 mL in each vial

From tube M1, 16 vials labeled 33 are dispensed; 1 mL in each vial

From tube M5, 16 vials labeled 37 are dispensed; 1 mL in each vial

From tube L2, 16 vials labeled 38 are dispensed; 1 mL in each vial

From tube L4, 16 vials labeled 40 are dispensed; 1 mL in each vial

Step 3: Making the sample sets
16 sets, each with 1 vial of the samples 1-40 are made and stored (frozen <-70 in most cases).

Step 4: Make sets with candidate calibrators. Note that the candidate calibrators must be prepared according to the manufacturers' instructions for use at the time the measurements are made. In most cases, these candidate calibrators cannot be thawed or reconstituted and then dispensed and refrozen or otherwise stored prior to distribution. Candidate calibrators must be stored and shipped under conditions specified by their manufacturer.

These are the samples 41 to 47 of Table 1. These samples are relabeled 41 to 47 and 16 sets, each with 1 vial of the samples 41-47 are made and stored.

Step 5: Shipment
One set 1–40 and one set 41–47 are shipped (under appropriate conditions) to the laboratories participating in the study.

Table 4. Example of preparing the set of samples derived from the 32 patient samples

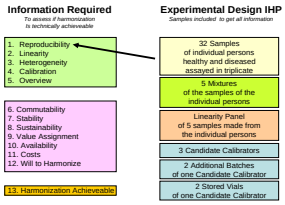
Collected Sample Number	Tube I1–I32 mL	M1 mL	M2 mL	M3 mL	M4 mL	M5 mL	L2 mL	L3 mL	L4 mL
1 - 8	16	2				0.5	1.5	1.0	0.5
9 - 16	16		2			0.5			
17 - 24	16			2		0.5			
25 - 32	16				2	0.5	0.5	1.0	1.5
Dispensed Sample Number	1-32	33	34	35	36	37	38	39	40

A6 Measurement of the samples by participating measurement procedures

Each of the samples of the sample set is measured in triplicate by each of the measurement procedures. Note that instructions for thawing, mixing and handling for each sample must be provided along with any precautions to control evaporation, position effects, order of samples, inclusion of controls, etc. Results are collected in an Excel file and sent to the HIG

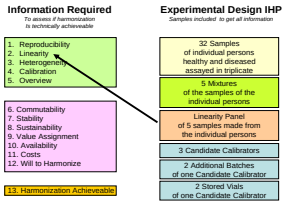
A7 Data Evaluation

Reproducibility



Reproducibility is calculated from the triplicates of the 32 samples. In general, ANOVA can be used to estimate imprecision components. However, in cases when the samples include extreme low/high concentrations that may have inappropriate influence on the estimates of imprecision, one should consider if ANOVA will result in relevant estimates of the reproducibility. In such situations, a biostatistician should be consulted.

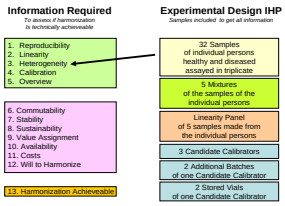
Linearity



For each of the methods the mean of the results for each member of the linearity panel are plotted with the mixture ratio (0 – 25 – 50 - 75 – 100% high) on the x-axis and the mean result on the y-axis. There are several ways to express the relationship:
- visual inspection
- calculation of r and y = ax + b (the most common approach)

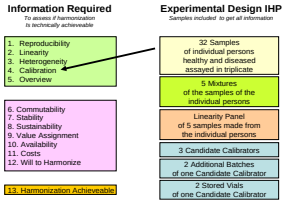
The protocol in CLSI guideline EP6 applies.
A linear relation allows convenient harmonization. Any non-linear relation will make harmonization more complicated. Non-linearity can be a) concentration-dependent (not linear starting from concentration x), b) S- shaped, convex, concave, or any other relation.

Heterogeneity



X-Y plots (difference plots may be used if needed) of all measurement procedure combinations are made (for example if there are 4 measurement procedures: procedure A against procedures B, C and D; procedure B against procedures C and D; procedure C against D; and r and y = ax + b calculated. Note that Deming regression should be used to account for imprecision in the values on both the x and y axis. The greater the dispersion of the results, the more heterogeneous is a measurement procedure. It could occur that results for normal samples show less scatter than for disease samples.

Calibration

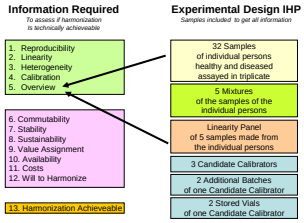


No need for additional calculations. Data calculated under heterogeneity can be interpreted:

- r <0.9: heterogeneity or non-linearity or non-reproducibility are limitations of one or both of the measurement procedures. However, calibration differences may also be evaluated from the slope but with less confidence than when r is larger. Excessive heterogeneity may preclude the ability to achieve harmonization.
- r >0.9 and the confidence interval for the intercept includes zero: if the slope indicates a difference in calibration between the two measurement procedures, harmonization with one point calibration is possible.
- r >0.9 and the confidence interval for the intercept does not include zero: if the slope indicates a difference in calibration between the two measurement procedures, harmonization with multiple point calibration is possible.

The criteria above are somewhat arbitrary but are useful for an initial assessment of technical feasibility for harmonization. Specific criteria may be different for different measurands based on the clinical requirements for using the biomarker.

Overview



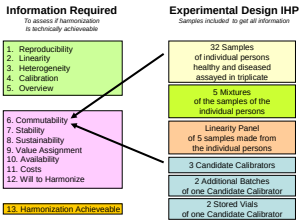
The degree of overall equivalence of results among the measurement procedures is expressed as the inter-method CV calculated from the means of the results for the 32 individual patient samples for each measurement procedure. This CV represents the “current equivalence.”

Next, the results of the respective measurement procedures are virtually harmonized by assigning a value to each of the 32 individual patient samples. If a designated comparison measurement procedure has been agreed, the value from that procedure can be used as the assigned value for each sample. In the absence of a designated comparison procedure, the mean (possibly trimmed) of the results from each measurement procedure for each sample may be used. An x-y plot is then made with the assigned value for each sample on the x-axis and the results of each individual measurement procedure on the y-axis. With the regression relation for each measurement procedure, the results of the individual patient samples are recalculated. Then the inter-method CV is calculated using these virtually harmonized results. This CV is the “achievable equivalence.”

If the achievable equivalence is much lower than the current equivalence, differences in calibration between the measurement procedures is the major factor for non-equivalence and harmonization is in principle possible. If not, one or more of the other test-related parameters (reproducibility, relationship, heterogeneity) are the major reason for non-equivalence. Which one can be derived from the calculations made.

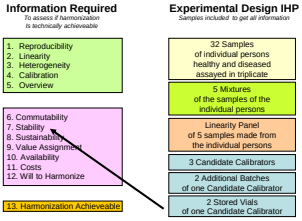
An alternative for determination if harmonization is achievable is (retrospective) calibration with the candidate calibrators (when these have been shown to be commutable) and recalculation of the results based on the alternative calibrators.

Commutability



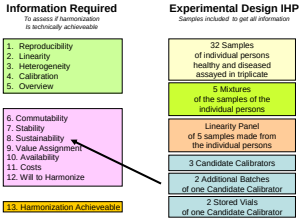
The results of the 32 individual patient samples and the candidate calibrators are evaluated according to CLSI EP30-A (formerly C53-A). The outcome is “commutable” or “not commutable”. However, the criteria in EP30 are based only on statistical distributions of data. There are no guidelines relating the statistical criteria to fitness for purpose of calibration of measurement procedures. Consequently, a work group will need to give consideration to the criteria for commutability in relation to the clinical use of results for a given measurand.

Stability



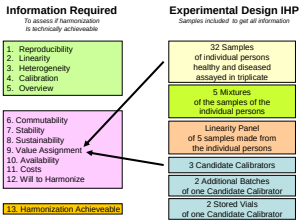
Measured concentration and precision of the candidate calibrators stored (e.g. 3 months) at higher temperatures for accelerated ageing are compared with the measured concentration of the same candidate calibrator stored at standard temperature. No difference is indicative for robust stability. Note that the conditions to be evaluated for accelerated ageing should be done in consultation with the manufacturer of the candidate calibrators.

Sustainability



Results of several batches of the same candidate calibrator and results of the 32 individual patient samples are evaluated for commutability according to CLSI EP30-A (formerly C53-A). If all are commutable, it can be concluded that it is possible to manufacture the candidate calibrator with reproducible commutability and thus the candidate calibrator is likely to be “sustainable”.

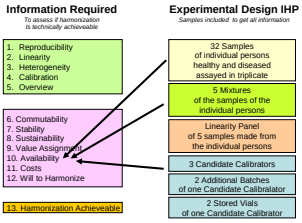
Value Assignment



Samples and candidate calibrators have been assayed by the respective measurement procedures and (optionally) also with candidate designated comparison measurement procedures. Data can be used to decide whether it is valid/acceptable to use an “all measurement procedures mean or trimmed mean” or the value of the candidate designated comparison method as the way to assign the value to

a calibrator. In the case of commercially available candidate calibrators, the manufacturer will have assigned a value to the calibrator. In this case, the data can be used to determine the suitability of that value or if another value assignment process is needed.

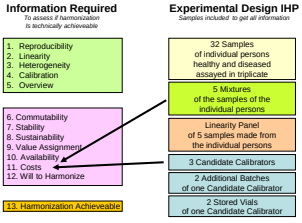
Availability



Mixture 1 is prepared from equal volumes of samples 1 to 8. The mean of samples 1 to 8 is calculated for the results of each of the measurement procedures. Similarly the means of samples 9-16 (mixture 2), 17-24 (mixture 3), 25-32 (mixture 4) and 1-32 (mixture 5) are calculated. When the means of the individual samples are the same as the measured value in the mixture it can be concluded that mixtures are commutable with individual samples. Then it is possible to make calibrators of pools of patient samples. Implication: such a calibrator is available for it is relatively easy to collect pools of patient samples. This capability is important if no candidate calibrators are available or the commutability of candidate calibrators is not acceptable for use with the routine measurement procedures.

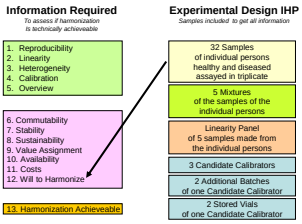
If the candidate calibrators are shown to be commutable, stable and sustainable; then use of such calibrators to achieve harmonization is warranted.

Costs



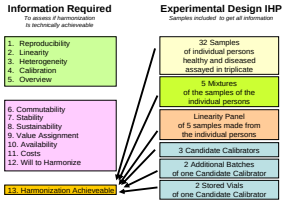
Costs of the tested (and approved) candidate calibrators allow estimating if these costs are acceptable for a harmonization effort. If no candidate calibrators are available (and thus should be developed) it is important to know if mixtures of patient material are suitable. If yes, then costs may be acceptable.

Will to harmonize



During the assessment study there will have been contacts with the manufacturers and other stakeholders. From this collaboration, the HIG will get an impression if the stakeholders have the will to harmonize when harmonization proves to be technically achievable. It is also recommended to contact regulatory agencies to solicit their cooperation and advice in working with IVD manufacturers to achieve recalibration of measurement procedures.

Harmonization Achievable

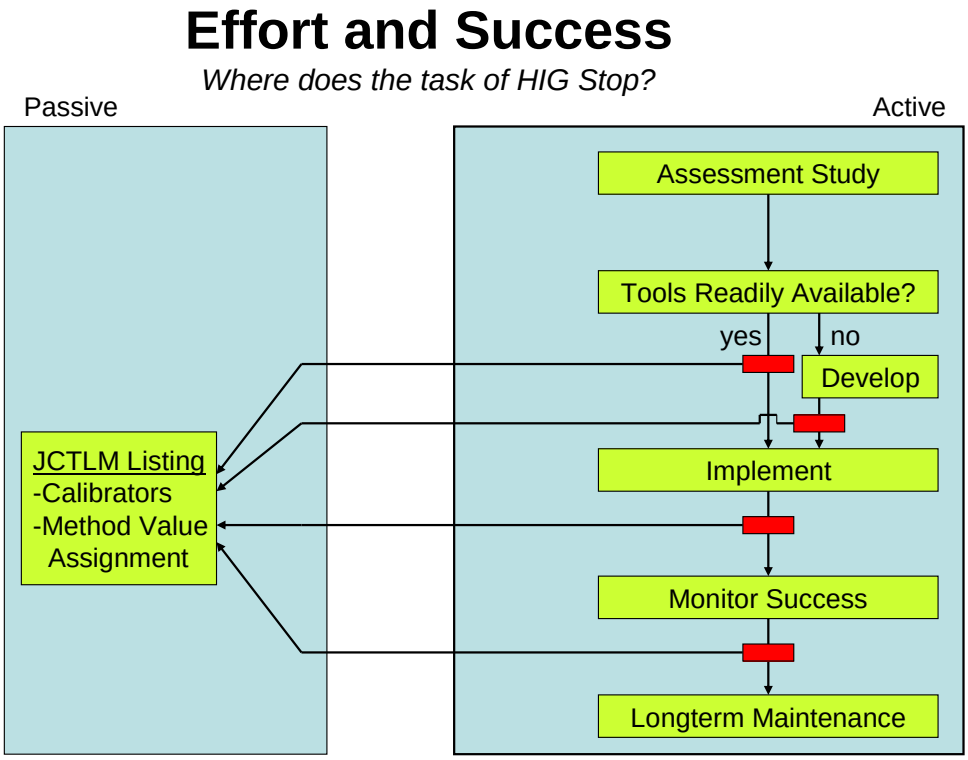


Results of the IHP will supply information to determine if harmonization is in principle possible (parameters 1 to 5), and if tools for harmonization are available or can be developed (parameters 6 to 12).

Section B. Effort and Success

When dealing with the harmonization effort and the monitoring of success the HIG should establish how far they want to be involved. This is summarized in Figure 4

Figure 4. Effort and Success



First of all, there is the difference between active harmonization and passive harmonization. Passive Harmonization (left of the figure) means that essential tools (calibrators and method for value assignment) are on JCTLM lists and it is assumed that the harmonization process will proceed by itself. Active Harmonization implies that the HIG is actively steering the process (right side figure). However, it is possible that data will be collected that demonstrates, for example, that a JCTLM listed or other international reference material used as a calibrator is in fact not commutable for the intended measurement procedures. In such a case, the HIG will need to have an active role in correcting the situation.

Second, the HIG can stop acting at several points in the process as indicated with the red bars in the figure.

1. The first moment to stop is when the assessment study has shown that appropriately validated tools are available. Calibrators and a method for value assignment are listed at JCTLM and the HIG is disbanded.

2. When the assessment reveals that tools are not available, the HIG actively develops them. Once this is done, calibrators and a method for value assignment are listed at JCTLM and the HIG is disbanded.

3. After implementation of a harmonization scheme, the HIG monitors success once or twice, develops appropriate mechanisms by collaboration with suitable organizations to continue periodic assessment and is then disbanded.

The process below describes all these stages; HIGs for specific measurands (in collaboration with the HOG), decide which part they do and which part they do not include in their activities.

Tools available

The first step of the implementation is that the HIG must warrant that the tools for harmonization (once more: the calibrators and the method to assign a value) are indeed available. If not, the HIG has to develop them. Parts of the assessment study can be used to develop the tools.

To create a calibrator the results of mixtures are important. If mixtures are commutable, the basic material for a calibrator can be obtained relatively easily. Then conservation options can be investigated (lyophilization, freezing) and the candidate calibrators can be evaluated according to the process described in the assessment study (commutability and stability). If mixtures do not work, the HIG may consider surrogate materials or panels of patients' samples.

Similarly, if there is no procedure to assign values to a calibrator, such a procedure should be developed. Information from the assessment study can be of help. It may be that a candidate value assignment procedure can be developed to a level that is acceptable for all parties involved. The mean or trimmed mean of results from all measurement procedures may be acceptable to assign a value. Other options include: a) an arbitrarily assigned value, b) a value assigned with the standard addition method after spiking with the pure analyte to the biological matrix. The HIG should ensure that the value assignment is sustainable: when a new lot of calibrator is manufactured a number of years after the first batch it should be possible to assign the value in the same way.

It cannot be stressed enough that calibrators and value assignment are the core of the harmonization process. These materials and procedures must be available and must be accepted as such by all stakeholders; if not the harmonization will never be successful. But at the same time one should realize that development of these tools can be complicated; there is no general recipe to create them, and pragmatic solutions and compromises may be necessary. The overall guiding principle is to have tools that are fit for the intended clinical use.

Implementation, once the assessment is completed, confirms that harmonization is possible and tools for harmonization are available.

Meeting Stakeholders

Implementation should start with a meeting of the stakeholders. The HIG should explain the reason for harmonization and also that it is possible (tools available). The outcome of the meeting should be that the HIG has a commitment for:

- a pilot study with the manufacturers to test if harmonization is technically achievable,
- implementation of harmonization in daily routine once the pilots show the process to be successful.

Results of a pilot study may be sensitive. To prevent that manufacturers may be reluctant to participate, it may be considered in advance of the pilot: a) if results will be presented anonymously in the meeting with stakeholders to review the pilot; and b) if results will only be presented when the manufacturer agrees to share them. However, all parties should discuss in advance of the pilot if results will be published or not. If the desire is to publish the results, the final manuscript will need to identify all participating manufacturers (this requirement is a policy of clinical laboratory journals). Note that JCTLM requires a peer reviewed journal report of the validation studies for harmonization materials and value assignment. Consequently, the results from one or more of the pilot studies will need to be published and all participating measurement procedures identified.

Pilot 1

The HIG makes a set of samples consisting of n calibrators and m blind patient samples. Values are assigned to all these samples using an appropriate procedure on which the stakeholders agreed (e.g. designated comparison measurement procedure or other process such as trimmed mean of all routine procedures). The sample sets are shipped to the respective manufacturers (or laboratories in case of lab developed tests) including a statement of the values assigned to the calibrators (but not to the blind patient samples). Manufacturers use the calibrators to calibrate their measurement procedures, according to protocols suitable for their procedures, and report the results for the blind samples based on the new calibrators (dataset a). Manufacturers also measure the blind samples using their established calibration and report the results (dataset b). The HIG calculates the interlab CV of dataset a (achieveable equivalence) and dataset b (current equivalence). Results are discussed in an

evaluation meeting with the stakeholders. Major points for discussion: a) did the logistics for recalibration and assessment using the individual samples work, b) is the achieveable equivalence acceptable for the clinical use of the measurand, and c) is there committment for a second pilot if needed. Lessons learned in the pilot are the basis to organize the second pilot.

Pilot 2 (Pilot 3 ... n)

The HIG organizes a second pilot on basis of the results/lessons of pilot 1. Outcome of the evaluation can be that results are reassuring and warrant starting implementation of recalibration to achieve harmonization. Or that results are not good enough yet, further improvement is needed and a third pilot is needed, etc., until the conclusion is that the harmonization system works (or does not work). If the harmonization is successful, the calibrators and method for value assignment should be registered at JCTLM.

Implementation in routine laboratory practice

Here communication and commitment are of utmost importance. Thus, the HIG should organize again a meeting of the stakeholders (do not forget clinicians!). Outcome of the meeting should be:

- commitment of all stakeholders to harmonization
- time schedule (when to start; if a period of double reporting of old and new values is needed)
- publicity campaign to educate laboratories and clinical users of the laboratory results

Monitoring success

The HIG has at least two options to monitor success: a) a specific study with blind samples (a similar approach as the pilot studies), or b) results of EQA organizers using commutable samples. Both options can also be used in parallel. The HIG continues monitoring until harmonization is achieved. Monitoring via EQA/PT programs is the preferred option as this does not require additional work and results from many laboratories performing measurements under typical clinical laboratory conditions are a more reliable assessment of harmonization. But suitability of the scheme, especially the use of commutable samples, should be assured.

Long-term monitoring of success and maintenance of harmonization

At the completion of the harmonization project, the HIG will be disbanded. However, it may be necessary for long term surveillance and maintenance of the harmonization scheme, that an existing organization assumes responsibility, or a new organization is formed, to continue to monitor the harmonization and actively intervene to maintain appropriate harmonization. Examples are the NGSP/IFCC network for HbA1c and the CDC network for lipids.

Part 3B: The Step-Up Design for Harmonization

Foreword

The step-up design is intended to establish harmonization of measurements in situations when there is not a reference measurement procedure nor a suitable commutable reference material. The approach is based on the concept of using a statistically valid target as a surrogate reference measurement procedure applied to a panel of clinical samples that become a set of international conventional calibrators. Commutability is an inherent characteristic of these samples by taking care in the origin of the clinical samples and the way they are collected and handled.

As suggested by the name, the design comprises a sequence of phases that should enable a decision that it is appropriate to step-up to the next phase. In essence, each phase consists of a method comparison study, in which measurement procedures that are candidates for harmonization measure a selected set of commutable samples. The harmonization concept itself has been described in general terms (1). For the concept to be successful, it is essential that: a) as many procedures/manufacturers as possible are involved in the method comparison leading to the harmonization target, b) that the procedures' performance is shown to be consistent over the covered measurement interval (including physiological and pathological concentrations of the analyte), and c) that the procedures sufficiently correlate with each other to indicate they are measuring the same quantity. Only under these conditions will it be possible to estimate a statistically valid target value for each sample to serve as a set of international conventional calibrators for harmonization. This target value can be the all procedures' mean, weighted mean, median or another statistical locator. Statistical methods have been elaborated to estimate a harmonization target value, among them principal component analysis (PCA) (2-4). It is the intention to investigate the feasibility of using PCA to estimate the "all procedure trimmed mean (APTM)" for harmonization of TSH immunoassays that participated in a method comparison conducted in the framework of the IFCC project for standardization of thyroid function test (5).

"Step-Up" approach in a nutshell

-Phase 1: "Familiarization phase" that provides: (i) a general picture of the intrinsic quality and comparability of assays by use of high-volume single donation samples from apparently healthy volunteers (note: the intrinsic quality of an assay is reflected by performance attributes such as imprecision, within-run stability, between-run differences, calibration consistency, etc.); and (ii) that allows a decision to "step-up" to phase 2 that uses lower volumes of normal and clinical samples.

-Phase 2: "Step-Up" phase that provides: (i) a detailed insight of assay quality and comparability by use of "normal" and "diseased" clinical samples; (ii) a decision that harmonization is feasible; and (iii) sets preliminary target values to the panel of clinical samples for harmonization.

-Phase 3: "Harmonization phase" that provides: (i) production of a panel of clinical samples for harmonization; and (ii) a protocol for sustainability to include transfer of target values to follow-up panels. The protocol in (ii) will maintain traceability to the same harmonization target values and thus panels that will be suitable for new measurement procedures that appear on the market.

"Step-up" phases in more detail

Phase 1: "Familiarization phase"

Donor Selection
Select 40 high-volume single donations (~200 mL) from a panel of apparently healthy persons (maybe 200 or more) that has been screened for the target analyte.

Note. This approach was applied for serum free thyroxine in (5). A panel of 40 sera was purchased from Solomon Park Research Laboratories (Kirkland, WA). The donors were selected based on

screening 200 subjects for their serum free thyroxine (FT4) concentrations using a commercially available immunoanalyzer.

Sample production
Define the protocol for sample production, for example, according to the CLSI C37-A protocol (6).

Note. In the T4 study, the panel was produced according to the C37-A protocol (6) from the Clinical and Laboratory Standards Institute except that the serum was not filtered nor pooled (5).

Measurement protocol
Define a measurement protocol to assess performance among participating manufacturers that addresses imprecision, reagent lot variations, and quality control.

Note. In the T4 study, manufacturers received 6 vials (1.0 mL) per serum, respectively. They were requested to store them at -70°C until analysis, on which occasion, thawed samples had to be used within a reasonable timeframe (ca. 8 hours), without refreezing. The protocol prescribed that the manufacturers should do the measurements in the samples in duplicate in 3 separate runs. The duplicates per run should be obtained by measuring the samples in ascending (#1 - #40) and descending order (#40 - #1), respectively. Internal quality control samples (preferably at 3 concentrations) should be analyzed at the start, in the middle and at the end of each run. Preferably 3 different reagent lots should be used, however, if not possible, 3 different lots of calibrators or a combination of different reagent/calibrator lots (5).

Data treatment (see also Ref. 5, for example)
- "Contract" a statistician for guidance on the treatment of the data, comprising:

- Data presentation, for example, by scatter-plot, difference-plot, and residuals plot.

- Calculation of the APTM as target values. For statistical approaches for the calculation of an APTM, see Refs. 2-4.

- Assessment of the between-assay comparability by an adequate regression procedure (for example, weighted Deming, polynomial regression) using the APTM and an adequate "location measure" (for example mean or median) of all data from a given procedure.

- Assessment of assay quality parameters from imprecision and comparison of singlicate results versus the APTM using predefined quality specifications for imprecision, bias, and total error. If necessary, other specifications may be defined, for example, the limit of quantitation.

- "Mathematical" recalibration. Note that it might be necessary that manufacturers perform recalibration under guidance of the statistician and with consideration of the characteristics of their particular measurement procedures.

Data interpretation
Make decisions to either:
- Stop the process, recommend improvement of the measurement procedures, and repeat the process in a suitable time period; or

- "Step-up" to the use of clinical samples that include healthy and diseased persons (phase 2). This process should be experience-based. Note that such a step up process is in active investigation by the IFCC Committee for Standardization of Thyroid Function Tests.

Phase 2: "Step-up"

Donor Selection
Select ~30 samples from presumably healthy persons and ~100 samples from persons with clinical disease that cover the intended diagnostic applications of the measurand (note the sample volume may be limited due to availability and ethics board requirements: perhaps 15 mL).

Sample production
Define the protocol for sample production, for example, samples collected and handled as typically used in the routine laboratory (for example, gel separator tubes).

Measurement protocol
Use a protocol with a reduced number of measurements, for example, 1 instrument, 1 lot, duplicates. This may require quality specifications for lot-to-lot and instrument-to-instrument variation; include the manufacturers' master calibrators in this and all further phases. If a candidate reference material is under consideration, it can be included as well to evaluate its performance.

Data treatment
See phase 1.
- Manufacturers provide "usual data" and recalibrated data.
- Set provisional target values to the panel of patient samples for harmonization.

Data interpretation
Make decisions to either:
- Stop the process, recommend improvement of the measurement procedures, and repeat the process in a suitable time period; or

- "Step-up" to the harmonization phase (phase 3);

- Select a reduced number of measurement procedures to be used for target value setting of panel 2 in the harmonization phase. Note that the first panel should be measured by all procedures that were included in the harmonization effort from the first phase on, whereas the second panel is intended to sustain harmonization and to be made available for measurement by new procedures that enter the harmonization effort and/or are launched on the market after the first harmonization effort (for more details, see below).

This process should be experience-based. Note that such a step up process is in active investigation by the IFCC Committee for Standardization of Thyroid Function Tests.

Phase 3: "Harmonization phase"

Note in advance
Depending on the number of measurement procedures to be harmonized, 2 panels may need to be produced because of the volume constraints when obtaining clinical samples from diseased individuals.

Donor Selection
Select [2 sets] of dedicated donations for harmonization: cover the typical measurement interval of the procedures; do not use "problematic samples" (i.e., clinical samples for which it is known they give aberrant results with many procedures, e.g. because of a defined cross-reactivity or interference); aim to optimize the sample volume.

Sample production
Define the protocol for sample collection and handling, for example, samples used typically in the routine laboratory (for example, gel separator tubes).

Measurement protocol
Set 1 (for harmonization of participating procedures/manufacturers) Use a protocol with a reduced number of measurements, for example, 1 instrument, 1 lot, duplicates. This may require quality specifications for lot-to-lot and instrument to instrument variation; include the manufacturers' master calibrators in this phase. If a candidate reference material is under consideration, it can be included as well to evaluate its performance.

Set 2 (for sustaining harmonization and for making available to new procedures that enter the harmonization effort or are newly launched on the market). Use a selected group of procedures for value assignment to have sufficient volume left for sustaining harmonization and for new manufacturers. Note, the criteria for selecting this group of measurement procedures is not fixed but the group will likely represent commonly used procedures that meet quality specifications derived from the data obtained in phase 2.

Note: it is important that in this phase manufacturers measure in parallel their own internal samples (pools or master lots of calibrators) for ensuring long-term stability of value assignment protocols in their manufacturing processes!

Data treatment
See phase 2.

Data interpretation
Set target values for harmonization of participating procedures for set 1.

Set target values for sustaining harmonization of participating procedures and harmonization of new procedures for set 2.

Define protocol for new procedures. This process should be experience-based. Note that such a step up process is in active investigation by the IFCC Committee for Standardization of Thyroid Function Tests.

Strengths and limitations of the step-up design in the harmonization concept

The strengths of the step-up design comprising consecutive phases of method comparison, each of them fulfilling specific requirements in terms of, e.g., sample nature, sample number, measurement protocol, are:
(i) it allows to decide after each step whether the step-up to the next phase can be made or is premature. In the latter case, the process can be stopped temporarily until sufficient technical progress has been made to continue the harmonization activity;

(ii) the technical process of method comparisons with application of adequate measurement protocols allows to also look into the intrinsic quality of the participating procedures, more in particular, whether the procedures' performance is commensurate with their intended use;

(iii) the target values statistically inferred from a method comparison average out individual procedure effects;

(iv) it maintains any previously established traceability of the examined procedures, for example to the IU of a WHO standard. This traceability is by virtue of the fact that in the process of determining the APTM, the measurement results from the existing calibration traceability transfer the units to the panel target values. Harmonization against that APTM then only requires application of a so-called master equation that can be applied to each manufacturer's existing calibration traceability procedure.

Potential limitations of this design are:
(i) the need to recruit clinical samples from dedicated supply sources that can deliver with good collection rates and accommodate varying specifications (adequate quality in terms of blood collection and further processing, representativeness of the donors for the application, compliance with exclusion criteria, availability of in-depth patient information such as age, gender, ethnicity, relevant medical history, co-morbidities, current and past medication, etc.).

(ii) The high cost associated with (i).

(iii) Assurance of the sustainability of the harmonization process in the continuum. However, as described in the above outline, solutions have been foreseen, primarily consisting of target setting of a 2nd set of samples by selected procedures only. This approach is a compromise to ensure that sufficient sample material is left for new procedures introduced to the market, and that the units and target values of the first panel can be transferred to all follow-up panels. This sustainability requires the set-up of a stable structure/body and agreement on protocols to prevent set to set drift over longer time periods.

References: Step Up Design for Harmonization

(1) Traceability to a common standard for protein measurements by immunoassay for in-vitro diagnostic purposes". Thienpont LM, Van Houcke SK. Clin Chim Acta 2010;411:2058-61.
(2) Rymer JC, et al. A new approach for clinical biological assay comparison and standardization: application of principal component analysis to a multicenter study of twenty-one carcinoembryonic antigen immunoassay kits. Clin Chem 1999;45:869-81.
(3) Lawton WH, et al. Statistical comparison of multiple analytic procedures - application to clinical-chemistry. Technometrics 1979;21:397-409.
(4) Carey RN, Wold S, Westgard JO. Principal component analysis: an alternative to "referee" methods in method comparison studies. Anal Chem 1975;47:1824-9.
(5) Thienpont LM, et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests; part 1: thyroid-stimulating hormone. Clin Chem 2010;56:902-11. And general Supplement to Parts 1 – 3.
(6) CLSI. Preparation and validation of commutable frozen human serum pools as secondary reference materials for cholesterol measurement procedures; approved guideline. CLSI document C37-A. CLSI; 1999.