
CIPM MRA-G-11 v1.2

— EXPLANATIONS —

Antonio Possolo (NIST, U.S.A.)

April 29, 2021

1 Introduction

This document explains and justifies recommendations for improvement of version 1.1 of CIPM MRA-G-11, *Measurement comparisons in the CIPM MRA: Guidelines for organizing, participating and reporting*, proposed by the JCRB Task Group on Statistical Methods for Key Comparisons.

This Task Group was formed in response to JCRB Action 42/3 taken during the 42nd meeting, in September 2020. The recommendations are the result of collaborative work of a majority of the members of the Task Group, and did not face sustained opposition from any of them:

AFRIMETS	Ian Veldman	NMISA	South Africa
APMP	Victoria Coleman	NMIA	Australia
APMP	Annette Koo	MSL	New Zealand
COOMET	Anna Chunovkina	VNIIR	Russian Federation
EURAMET	Steve Ellison	LGC	United Kingdom
GULFMET	Abdulrahman R. AlAskar	NMCC	Saudi Arabia
GULFMET	Christos Mitsas	EMI	United Arab Emirates
SIM	Silvina Forastieri	INTI	Argentina
SIM	Blaza Toman	NIST	U.S.A.
TG LEAD	Antonio Possolo	NIST	U.S.A.

The Task Group was given *Measurement comparisons in the CIPM MRA: Guidelines for organizing, participating and reporting* (CIPM MRA-G-11 DRAFT, dated November 5, 2020), as the target for improvement, and submitted justified recommendations to the JCRB on February 15th, 2021, for editorial and technical improvements.

However, the target document underwent changes while the Task Group was doing its work based on that earlier version. For this reason, the references to line numbers that appear in the Task Group’s report are no longer relevant, hence the need for the explanations given here.

All references made to page numbers pertain to version 1.1 of CIPM MRA-G-11, dated January 18th, 2021. The editorial changes are incorporated without comment, and the explanations and justification focus on the technical changes, reformulated as needed to fit in the context of version 1.1. Both kinds of changes are clearly visible as tracked changes in the appendix, as well as in the companion Word file CIPM-MRA-G-11-v1.2-Proposal.docx.

2 Technical Issues

Page 6 The CIPM MRA, Section T.3, states that “in some exceptional cases a Consultative Committee may conclude that for technical reasons a reference value for a particular key comparison is not appropriate.”

If the CC working group responsible for the comparison finds that such technical reasons exist, then the exception applies and no KCRV is calculated. In this case, “the results are then expressed directly in terms of the degrees of equivalence between pairs of standards.”

Page 9 Item (j) has been added to address challenges that often arise in comparisons involving either chemical or radiometric quantities. In particular, in comparisons involving radiometric quantities, participants often make measurements before and after measurements made by the pilot rather than the other way around, hence changes can occur for any participant, not just for the pilot.

Page 11 Data reductions of measurement results from KCs involve statistical models and statistical data analysis. Consistently with generally accepted best practices in statistics, the selection of a model for the data and of a procedure for

data analysis should take the specific data into account, hence ought not to be decided before the data are in hand.

The recommendation is that an agreement be reached about how a method will be selected for the data reduction when the comparison is being planned, but that the actual selection of the specific method to be used will be made only after the measurement results will have been obtained.

The Inorganic Analysis Working Group of CCQM, for example, has recently decided to adopt a particular procedure to select the method that will be used to determine the key comparison reference value, and to reserve the choice of specific method until after the measurement results are in hand. CCT's Working Group for Key Comparisons has made a similar decision for an upcoming comparison involving the triple point of water.

The proposed text gives greater flexibility to the CCs and to their working groups, for exactly when and how they will choose a specific statistical model for the measurement results and a specific method to reduce them.

The proposed text also highlights the need to take into account statistically and substantively significant *dark uncertainty* (Thompson and Ellison, 2011) (Koepke et al., 2017) (Possolo et al., 2021) that may be detected when the mutually consistency of the measurement results is evaluated.

Page 16 The proposal gives the CCs greater latitude for how to express the uncertainty associated with the KCRV, and emphasizes that decisions about which measurement results to include, and which to exclude, from the calculation of the KCRV, should be based on substantive considerations, not merely on statistical criteria. Statistical diagnostics (for example, to indicate whether some results appear to be anomalous) serve to draw attention to results that may require closer examination, but should not be decisional.

Page 17, Top The fourth and fifth paragraphs suggest, but do not mandate, that leave-one-out (or, cross-validated) DoEs be examined as useful diagnostics of the reliability of the KCRV and of the conventional DoEs.

Not only is there already some precedent for such examination (Duewer et al., 2014) (Koepke et al., 2017), it is already a fact that the DoEs for participants whose measurement results are not included in the calculation of the KCRV are leave-one-out (or leave-several-out) estimates.

It should be noted that the leave-one-out estimates of the KCRV (which are the versions of the KCRV when one measurement result is left out of its calculation in turn) enjoy the same statistical properties as the usual KCRV, other than for a slight increase in the associated uncertainty. Also, D_j and $D_{\text{LOO},j}$ estimate the same quantity, for $j = 1, \dots, n$, where n denotes the number of participants.

Furthermore, the leave-one-out estimates of the KCRV provide yet another means to evaluate the uncertainty associated with the KCRV, by application of the statistical *jackknife* (Mosteller and Tukey, 1977, Chapter 8) (Efron and Stein, 1981) (Efron, 1982), provided the number of participants is not too small.

On the one hand, the leave-one-out versions of the DoEs offer two important advantages:

- (i) $D_{\text{LOO},j}$ is uncorrelated with the corresponding *leave-one-out* estimate of the KCRV, thus simplifying the calculation of $U_{95\%}(D_{\text{LOO},j})$; and
- (ii) $D_{\text{LOO},j}$ is likely to be a more realistic estimate of the true difference than D_j because it is a *cross-validated* estimate (Hastie et al., 2009) (Stone, 1978) of this difference.

On the other hand, $U_{95\%}(D_{\text{LOO},j})$ is likely to be larger than $U_{95\%}(D_j)$, and the derivation of DoEs after linking an RMO key comparison with a CIPM key comparison may be more involved when leave-one-out DoEs are used than otherwise.

For this reason, they are proposed as potentially useful diagnostics about the reliability of the KCRV and about the leverage that individual measurement results have upon the KCRV and the degrees of equivalence, not as replacements for the conventional degrees of equivalence.

Page 17, Bottom and Page 18, Top Even though v.1.1 does not mention bilateral degrees of equivalence, they were mentioned in the version of the document that the JCRB Task Group examined. For this reason a passage relating to them was reintroduced.

Page 18, Third Paragraph The suggested rewrite of the passage concerning the possible expression of the unilateral DoEs in relative terms is intended as a clarification of the text that it replaces.

Page 20 The proposed text refrains from suggesting that the coverage factor $k = 2$ should always be used because, consistently with the GUM ([Joint Committee for Guides in Metrology \(JCGM\), 2008](#)), the selection of the coverage factor should take into account the number of degrees of freedom that the corresponding uncertainty evaluation is based on. Also, when a Monte Carlo method is used for the uncertainty evaluation, it is possible to compute the expanded uncertainty for any intended coverage probability directly, instead of as a multiple of the standard uncertainty.


Page 21 The proposed text serves to clarify which outcomes of the supplementary comparison should be reported.

References

- D. L. Duewer, K. W. Pratt, C. Cherdchu, N. Tangpaisarnkul, A. Hioki, M. Ohata, P. Spitzer, M. Máriássy, and L. Vyskočil. “Degrees of equivalence” for chemical measurement capabilities: primary pH. *Accreditation and Quality Assurance*, 19:329–342, 2014. doi: 10.1007/s00769-014-1076-1.
- B. Efron. *The jackknife, the bootstrap, and other resampling plans*, volume 38 of *CBMS-NSF Regional Conferences in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1982. ISBN 0-89871-179-7.
- B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9:586–596, May 1981. doi: 10.1214/aos/1176345462.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York, second edition, 2009. URL statweb.stanford.edu/~tibs/ElemStatLearn/.
- Joint Committee for Guides in Metrology (JCGM). *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement*. International Bureau of Weights and Measures (BIPM), Sèvres, France, 2008. URL www.bipm.org/en/publications/guides/gum.html. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, JCGM 100:2008, GUM 1995 with minor corrections.

- A. Koepke, T. Lafarge, A. Possolo, and B. Toman. Consensus building for interlaboratory studies, key comparisons, and meta-analysis. *Metrologia*, 54(3): S34–S62, 2017. doi: 10.1088/1681-7575/aa6c0e.
- F. Mosteller and J. W. Tukey. *Data Analysis and Regression*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1977. ISBN 0-201-04854-X.
- A. Possolo, A. Koepke, D. Newton, and M. R. Winchester. Decision tree for key comparisons. *Journal of Research of the National Institute of Standards and Technology*, 126:126007, 2021. doi: 10.6028/jres.126.007.
- M. Stone. Cross-validation: a review. *Series Statistics*, 9(1):127–139, 1978. doi: 10.1080/02331887808801414.
- M. Thompson and S. L. R. Ellison. Dark uncertainty. *Accreditation and Quality Assurance*, 16:483–487, October 2011. doi: 10.1007/s00769-011-0803-0.

Appendix



Measurement comparisons in the CIPM MRA

Guidelines for organizing, participating and
reporting

CIPM MRA-G-11

~~Draft Version 1.1~~ Version 1.2
~~18/01/2021~~ 30/04/2021

Contents

1. Introduction	3
2. Type of comparisons	65
3. Organization	76
3.1 Initiation of comparisons.....	87
3.2 Points for consideration.....	87
4. Technical protocol	109
4.1 Circulation of transfer standards and customs formalities.....	12-10
5. Registration of comparisons in the KCDB and status report	13-11
5.1 Nomenclature of comparisons	13-12
6. Participation.....	14-13
6.1 Associates of the CGPM in comparisons organized by Consultative Committees ...	14-13
6.2 Pilot studies run in conjunction with comparisons.....	15-14
7. Outcomes of comparisons	15-14
8. Reporting of comparison	18-15
8.1 Results of measurements	18-15
8.2 Comparison reports	18-16
8.3 Authorship of comparison reports	21-18
9. Disagreements	22-19
10. Publication in the KCDB	23-20
11. Monitoring the comparison results	23-20
12. Resources related to the CIPM MRA.....	25-22
13. Revision History.....	26-23
Appendix A - Flowchart of CIPM and RMO key comparisons	27-24
Appendix B - Flowchart of supplementary comparisons	28-25

1. Introduction

The technical basis of the CIPM MRA is the set of results obtained over the course of time through scientific key comparisons carried out by the Consultative Committees (CCs) of the CIPM, the BIPM and the RMOs. These results are published by the BIPM following their approval and are maintained in the BIPM key comparison database (KCDB). ~~The comparisons identify the participants and their individual results, and are available to support the CIPM MRA processes.~~ The comparisons identify the participants, provide their measurement results and degrees of equivalence with the comparison reference value, which become available to support the CIPM MRA processes. This document (CIPM MRA-G-11) outlines the requirements for the international comparisons under the CIPM MRA. It supersedes CIPM MRA-D-05 and CIPM MRA-G-04.

Figure 1 illustrates the key comparisons within the framework of the CIPM MRA. Table 1 gives a general overview of comparison types in the CIPM MRA. Appendix A and Appendix B contain flowcharts illustrating the comparison processes.

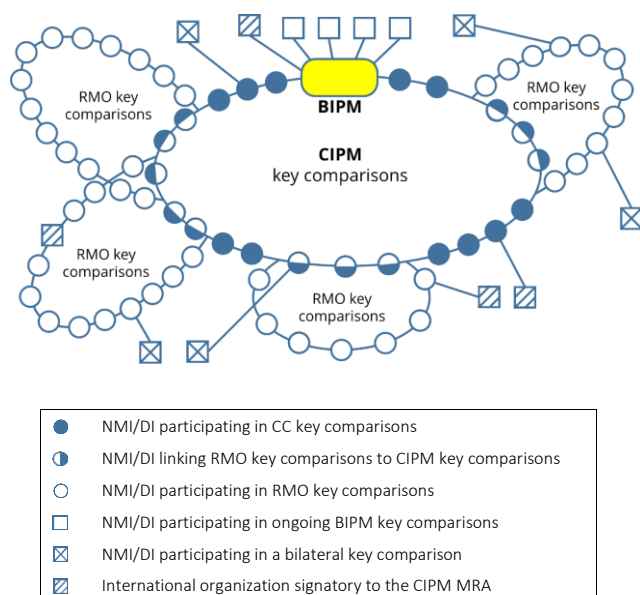


Figure 1. Illustration of relations between comparisons

Table 1. An overview of the comparisons organized within the frame of the CIPM MRA.

Type Activity	CIPM comparisons		RMO comparisons		Pilot studies
	Key	Supplementary	Key	Supplementary	
Objective (Section 2)	To test the principal techniques and methods in the field	To meet specific needs not covered by key comparisons	To extend the coverage of the CIPM key comparisons regionally	To meet specific needs not covered by RMO key comparisons	To establish measurement parameters for a "new" field or instrument, or as a training exercise
Organization (Section 3)	CCs and BIPM		RMO TCs/WGs		BIPM, CCs and RMOs
Technical protocol (Section 4)	Includes the proposal for the method of determination of the key comparison reference value	According to common requirements	Follows the CIPM key comparison and any relevant CC guidelines. Includes the way in which the results will be linked to the CIPM key comparison	According to common requirements	Depends on CCs and RMOs
Registration (Section 5)	Registered in the KCDB				Not registered in the KCDB.
Participation (Section 6)	Open to laboratories having the highest technical competence and experience (CC members). Participation may be restricted (see "2. Type of comparisons Type of comparisons" for details). Associates may participate in special cases		Open to all RMO members and other institutes (including from other RMOs), subject to decision by the organizing RMO		CCs and RMOs
Outcomes (Section 7)	Measured values and measurement uncertainties				Measured values and measurement uncertainties
	Key comparison reference values and degrees of equivalence	May include degrees of equivalence	Degrees of equivalence	May include degrees of equivalence	
Approval of reports (Section 8)	Withdrawal is generally not allowed				According to practice of CCs and RMOs
	Approved by CCs	Approved by CCs	Approved by CCs	Approved by RMOs	
CMC support (Section 8.2)	Draft B may be used to underpin CMCs	Final report needed to underpin CMCs	Draft B may be used to underpin CMCs	Final report needed to underpin CMCs (overseen by CC)	Generally not used to support CMCs
Publication (Section 10)	Published in the KCDB. For up-to-date information, the pilot institute shall report the status of comparisons. Recommended to publish in the Technical supplement of <i>Metrologia</i> or other scientific publications.				Not published in the KCDB. Pilot studies run by the BIPM are available on the BIPM website. Pilot studies by CC or RMO may be available

Formatted: Font: (Default) Open Sans Light, 8 pt

CIPM MRA-G-11 – Measurement comparisons in the CIPM MRA

2. Type of comparisons

A **key comparison** is selected by a Consultative Committee to test the principal techniques and methods in the field. Key comparisons may include comparisons of representations of multiples and sub-multiples of SI base and derived units as well as comparisons of artefacts. The key comparisons are essentially of two types:

- **CIPM key comparisons:** of international scope, are organized by Consultative Committees or the BIPM, and are restricted to laboratories of Member States and normally members of the corresponding Consultative Committees. CIPM key comparisons typically deliver a “reference value” for the key quantity chosen. Occasionally, no “reference value” is appropriate or substantively meaningful (CIPM, 1999, T.3): in such cases it is still possible to compute bilateral degrees of equivalence; and in some cases (when a reference value can be computed that is not substantively meaningful) it may also be possible to compute unilateral degrees of equivalence. Any degrees of equivalence that are deemed to be metrologically valid should be reported.
- **RMO key comparisons:** of regional scope, are organized at the scale of a region (though they may include additional participants from other regions) and are open to laboratories of Associates as well as Member States. RMO key comparisons are intended to provide RMO members with the means to link to the reference value established by the corresponding CIPM key comparison. The RMO key comparisons deliver complementary information without changing the reference value derived from the CIPM key comparison. A degree of equivalence derived from an RMO key comparison has the same status as one derived from a CIPM key comparison.

Key comparisons may be extended by **subsequent key comparisons**.

A **supplementary comparison** is intended to cover areas or techniques not addressed by key comparisons. These are complementary to key comparisons and are not intended as second-level comparisons. Their final reports are published in the KCDB, but degrees of equivalence are not necessarily computed.

Pilot studies are a third category of comparison normally undertaken to establish measurement parameters for a “new” field or instrument, or as a training exercise. The results of pilot studies alone are not normally considered sufficient support for calibration and measurement capabilities (CMCs) and the studies are not registered nor published in the KCDB.

3. Organization

For all comparisons, the body carrying out the comparison shall identify a **pilot institute** to take the main responsibility for running, registering and keeping the comparison updated in the KCDB.

Key comparisons

The Consultative Committees are responsible for selecting key comparisons. In each field, a set of key comparisons is identified to test the principal techniques in the field.

The procedures used by Consultative Committees for selecting, conducting and evaluating key comparisons, including their detailed technical protocols and periodicity, are designed to ensure that:

- a) comparisons test all the principal techniques in the field;
- b) results are clear and unequivocal;
- c) results are reliable and reproducible;
- d) results are, ideally, easy to compare with those of corresponding comparisons carried out subsequently by regional metrology organizations;
- e) overall, the comparisons are sufficient in range and periodicity to demonstrate and maintain equivalence between the institutes.

RMO key comparisons carried out by regional metrology organizations shall be linked to the corresponding CIPM key comparisons by means of joint participants. This is mandatory to demonstrate global equivalence. To achieve this, it is recommended that at least two of the participants in the preceding CIPM key comparison also participate in the RMO key comparison.

NOTE 1 Bilateral comparisons of standards with long-term stability carried out by the BIPM may be conducted according to special arrangements not necessarily covered by this document.

Supplementary comparisons

Supplementary comparisons are usually organized by RMOs to address regional needs; for example, measurements of specific artifacts, or measurements of properties that do not fall within the “normal” scope of the CCs. Usually, measurements made in supplementary comparisons have lower accuracy than comparable measurements made in key comparisons.~~These comparisons are normally organized by RMOs to cover regional needs, for instance, measurements of specific artefacts or with lower accuracy~~

~~measurements~~. Although outside of the “normal” scope of the Consultative Committees, they may also organize supplementary comparisons when:

- there are only few participants (none sharing the same RMO) capable of measuring the required quantity;
- no link can be made to an RMO comparison; or the distribution of comparison artefacts or transfer standards¹ to be measured is a constraint (for instance, measurements of radioactive matrix reference materials).

Subsequent comparisons

~~Any eligible national metrology institute or designated institute may request participation in a subsequent comparison involving one or more of the participants in the original comparison. Subsequent comparisons organized for one or several participants.~~ These comparisons should follow the same procedure as ~~the forerunning previous~~ comparison and are normally carried out for one of the following reasons:

- after completing a comparison, an institute considers its result ~~unrepresentative non-representative~~ of its standards/capabilities;
- an institute was not ready to participate at the time a comparison was conducted.

3.1 Initiation of comparisons

CIPM key comparisons are initiated at a Consultative Committee meeting. At each meeting, the Consultative Committee will consider the need for comparisons and may decide, at that meeting, after taking into account the views of RMOs among other things, to initiate new key comparisons.

RMO key comparisons may be initiated by individual RMOs, to allow all institutes belonging to that RMO to participate in key comparisons.

All key comparisons shall be approved in advance by the corresponding Consultative Committee. The mechanism for approval depends on the particular Consultative Committee’s practice.

3.2 Points for consideration

The organization of a CIPM key comparison is the responsibility of the pilot institute, which may be helped by a **coordinating group**. The Consultative Committee can form

¹ Transfer standards should be interpreted as referring to standards, artefacts, instruments, samples, etc.

the coordinating group by nominating one or more institutes to help the pilot institute throughout the process of the comparison. In the planning of the comparisons, the main points to be decided by the group are the following:

- a) In selecting participants, the Consultative Committees should take proper account of regional representation and the need to limit participation (typically no more than three institutes per RMO) when transfer standards are used sequentially.
- b) List of participants with full details of mailing and electronic addresses.
- c) Measurement standard intended to be used in the comparison by each participant.
- d) Transfer standards to be used in the comparison.
- e) Whether or not a pilot comparison or any other preliminary work needs to be carried out among a restricted number of participants to verify the performance of the transfer standard.
- f) Pattern of the full-scale comparison, which ranges from the simple circulation of a single transfer standard around all the participants to the sending of an individual transfer standard directly to each participant from the pilot institute, or from each participant to the pilot institute or some combination of these.
- g) Starting date, detailed timetable, means of transport and itinerary to be followed by each transfer standard. This starting date is subsequently referred to as the starting date for the comparison.
- h) Procedure in the case of failure of a transfer standard.
- i) Procedure in the case of unexpected delay at a participating institute.
- j) Procedure to adjust the measurement results to render degrees of equivalence meaningful.
 - when the values realized in the standards drift in the course of the comparison,
 - when the participants measure essentially different standards (a common occurrence in gas metrology and in chemistry generally), or
 - when a measurement scale used by a participating laboratory changes in the course of the comparison (for example, in comparisons involving radiometric quantities).
- ~~i) Procedure in the case of unexpected delay at a participating institute.~~

jk Customs documents to accompany the transfer standards, either ATA carnet or other for those participants not qualifying for the ATA scheme. If applicable, the *Customs Convention on the temporary importation of scientific equipment* should be considered.

The timetable should be discussed to ensure that the workload of the whole set is not too great for the participating and pilot institutes and that the results will be available for the next meeting of the Consultative Committee, normally in three (or occasionally two) years' time. For this, the total circulation time of the standards should be fixed and not exceed eighteen months unless there are exceptional circumstances.

4. Technical protocol

The technical protocol is an important part of the comparison and specifies in detail the procedure to be followed. The pilot institute draws up the detailed technical protocol and timetable for the comparison and its dispatch. An RMO key comparison should follow the same protocol as the preceding CIPM key comparison.

In those Consultative Committees having permanent working groups or sections responsible for specific areas of activity, the draft protocol shall be sent to the Chair of the relevant working group or section and to the Executive Secretary of the Consultative Committee. Consultative Committees or working groups may decide to publish the draft protocol on their corresponding website. The pilot institute is encouraged to publish the approved technical protocol in the KCDB.

The purpose of a key comparison is to compare the standards/capabilities as realized in the participating institutes, not to require each participant to adopt precisely the same conditions of realization. The protocol shall therefore specify the procedures necessary for the comparison, but not necessarily the procedures used for the realization of the standards being compared. The protocol should include:

- a) Detailed description of the transfer standard, relevant to the comparison: manufacturer, type, serial number, homogeneity and stability of samples, country of origin, size, weight, packaging, etc., and technical data needed for its operation.
- b) The metrological parameters that need to be measured.
- c) A statement indicating which service categories/CMCs can be supported by the comparison, or criteria to identify such categories/CMCs (i.e., 'how far the light shines').

- d) Advice on handling the transfer standards, including unpacking and subsequent packing and shipping to the next participant. This should include a complete list of the contents of the package, including handbooks etc., and the weight and size of the whole package.
- e) Actions to be taken on receipt of the standards in a participating institute.
- f) Any tests to be carried out before measurement.
- g) Conditions of use of transfer standards during measurement.
- h) Instructions for reporting the results.
- i) For CIPM key comparisons, description of [how to select](#) the method to be used to determine the key comparison reference value. [The method that ultimately will be selected should include provisions to detect and handle issues such as:](#)
 - [artifact drift.](#)
 - [measurement of essentially different standards \(a common occurrence in gas metrology and in chemistry generally\).](#)
 - [changes in measurement scales of the participants, or](#)
 - [mutual inconsistency of the measurement results, including statistically and substantively significant dark uncertainty \(Thompson and Ellison, 2011; Koepke et al., 2011\).](#)
- ⚡ RMO key comparison technical protocols should include the method to be used to link to the corresponding CIPM key comparison reference value.
- j) A list of the principal components of the measurement uncertainty budget to be evaluated by each participant and any necessary advice on how measurement uncertainties are estimated, based on the principles laid out in the *Guide to the expression of Uncertainty in Measurement (GUM)*. In addition to the principal components of the measurement uncertainty common to all participants, individual institutes may add any others that they consider appropriate. Measurement uncertainties shall be reported as standard uncertainties, and information shall be given on the number of degrees of freedom.
- k) A timetable for communicating the results to the pilot institute. Early communication helps to reveal problems with the transfer standard during the comparison.
- l) Financial aspects of the comparison including transport and customs charges as well as any damage that may occur, noting that in general each participating institute is responsible for its own costs for the measurements. Overall costs of the organization of the comparison, including the supply of the transfer

standard, are normally borne by the pilot institute. Any other arrangement for sharing costs is accepted, if agreed, by all the participants.

- m) Insurance arrangements for transfer standards are decided by agreement among the participants, taking account of the responsibility of each participant for any damage within its country.

4.1 Circulation of transfer standards and customs formalities

The pilot institute is responsible for organizing the circulation and transport of the standards and ensuring that the participants make proper arrangements for local customs formalities, noting particularly any specific requirements related to biological, chemical or ionizing material.

The transfer standards must be handled with care, i.e. only by qualified metrology personnel. It is desirable, and in some cases essential, that the transfer standards be hand-carried. If this is not deemed essential, certain precautions must nevertheless be taken. As goods are usually delivered to a shipping department in an institute, a warning note should be attached to the package indicating that the package should be opened only by the suitably qualified laboratory personnel. The participating institutes are responsible for transport to the next institute according to the circulation scheme. The method of transport as defined in the instructions shall be respected.

Before dispatching the package, each participant shall inform the next participant and the pilot institute, giving transport details.

In cases where an ATA carnet is needed, it must be used correctly. For each movement of the package, the person organizing the shipment must ensure that the carnet is presented to customs on leaving the country and again on arrival in the destination country. When the package is sent unaccompanied, the carnet must be included with the other forwarding documents so that the handling agent can obtain customs clearance. The carnet shall not be packed inside the package with the transfer standards under any circumstances. In some cases, it is possible to attach the carnet to the package.

When the shipment arrives, the participating institute shall inform the pilot and, if required, the dispatching institute by completing and returning a form included with the package or in the technical protocol. Immediately after receipt, the participating institute shall check for any damage to the transfer standards according to instructions provided in the comparison protocol, and report this to the pilot institute.

If a delay occurs, the pilot institute shall inform all the participants and, if necessary, revise the time schedule or the order of circulation between countries.

5. Registration of comparisons in the KCDB and status report

Registration of comparisons shall be made through the KCDB after having been approved to be carried out by the Consultative Committee or RMO, and before starting the measurements. Only key and supplementary comparisons are registered in the KCDB. Only the institutes participating in the CIPM MRA will be listed in the public website of the KCDB for the comparison.

During the course of a comparison that is registered in the KCDB, it is important that up-to-date information on the progress of the comparison is readily available. On a regular basis, the pilot institute will receive an automatic notification to update the comparison status in the KCDB. The President, the Executive Secretary, and the working group designated by the Consultative Committee shall concurrently be informed on the progress by the pilot. Once the progress of the comparison is reported in the KCDB by the pilot, the updated status will automatically be made public.

5.1 Nomenclature of comparisons

On registration in the KCDB, the KCDB Office assigns a code for each comparison. The pilot institute may suggest a code to the KCDB Office. The format of the comparison nomenclature is normally (square brackets indicating optional elements):

BODY[.]Area[.WG]-TypeX[.ID]

BODY	the operator, e.g. Consultative Committee, BIPM or specified RMO. A separating dot [.] is added to BIPM and the RMO names for clarity.
Area	the corresponding acronym of the metrology area
.WG	sub-field or section, for example .RF for Radio Frequencies or (I) – the latter a roman numeral with brackets as separators instead of a dot.
Type	capital letter: K for key comparison, S for supplementary comparison.
X	number (normally in successive order)
.ID	optional identifier (.a , .b , .Xy-αβγ , .1 , .2 , .year , ...) that may be requested according to practice in metrological fields. It may also be used for subsequent bilateral key comparisons. Dots or hyphens may be used for clarity in this part of the identifier, as required.

Two or more comparisons corresponding to the same description but carried out over two different time intervals shall have unique identifiers. Normally, these comparisons are identified with the successive numbers, in which case the **.ID** part may be kept unchanged. However, it is possible to keep the same number, in which case changing the **.ID** part is mandatory.

Some RMOs use an internal identifier before the comparison is registered. This identifier can be listed in the KCDB and found using the website search engine.

6. Participation

Participation in a CIPM key comparison is open to laboratories having the highest technical competence and experience, normally the member laboratories of the appropriate Consultative Committee. The number of laboratories participating in CIPM key comparisons may be restricted by the Consultative Committee for technical reasons or when transfer standards are used sequentially. An invitation to participate should be sent by the pilot institute to the relevant Consultative Committee members, with copy to the Executive Secretary of the Consultative Committee and the RMO Secretariats.

Participation in key comparisons organized by an RMO is open to all RMO members and to other institutes that meet the rules of the regional organization (including institutes invited from outside the region) and that have technical competence appropriate to the particular comparison. Participation in RMO key comparisons is decided by the appropriate committee of the RMO.

The results for participants non-signatory to the CIPM MRA should be considered as evidence of metrological competence for any future CMC submissions in the event that the laboratory becomes a signatory to the CIPM MRA. Note, that this would not apply to laboratories participating in a measurement comparison under less stringent rules than the signatory laboratories (for example, as a 'pilot study' participant for a measurement comparison in chemistry).

The rules for the participation in CIPM and RMO key comparisons also apply to CIPM and RMO supplementary comparisons.

6.1 Associates of the CGPM in comparisons organized by Consultative Committees

The participation of Associates in comparisons organized by Consultative Committees shall be carefully considered by the relevant Consultative Committee on a case-by-case

basis. Specifically, and in exceptional circumstances, Associates may be invited to take part in comparisons organized by Consultative Committees and pilot studies where:

- this adds scientific or other value to the work or to the results obtained by other participants;
- reference samples are only produced for the purposes of comparisons organized by Consultative Committees, and no linked RMO comparisons are possible; and
- their participation increases the efficiency or adds effectiveness to the relevant activity.

Reports of comparisons organized by Consultative Committees where Associates take part may be published in the KCDB. These reports shall make clear which results come from Associates. Their results shall not contribute to the key comparison reference value unless it can be shown to be of significant scientific value to other participants.

Associates invited to take part in a key comparison organized by a Consultative Committee may be invited to attend working group meetings at which the results from that comparison are discussed.

6.2 Pilot studies run in conjunction with comparisons

It is important to note that an institute that has never taken part in a comparison may wish to acquire a benchmark of its performance before participating in a comparison. This can be achieved by running pilot studies in conjunction with a comparison or by participating in a comparison in “pilot study” mode. The results of participants seeking to benchmark their performance are not to be used to compute reference values, and the name of those institutes will not be published in the KCDB. Participation in “pilot studies” running in conjunction with comparisons shall be agreed before the comparison measurements start. Results from pilot studies are not considered sufficient support of CMCs. Such exercises should be organized such that any risk of delay is minimized for the publication of CMCs in the KCDB.

7. Outcomes of comparisons

~~The key comparison reference value is the reference value resulting from the measurements taken in a CIPM key comparison, accompanied by its measurement uncertainty (normally the standard uncertainty). Only CIPM key comparisons (carried out by a Consultative Committee or the BIPM) generate a key comparison reference value. Each key comparison reference value is considered to be a close approximation of the~~

~~true value. The method used to determine the key comparison reference value is part of the protocol of the comparison and is agreed by the Consultative Committee or by the appropriate working group to which the Consultative Committee has delegated this task.~~

~~The key comparison reference value (KCRV) is a consensus value that results from blending the measurement results (measured values and reported uncertainties) obtained by selected participants in a CIPM key comparison.~~

~~The KCRV must be qualified with an evaluation of the associated uncertainty, ideally expressed as a probability distribution on the set of the values of the measurand, or alternatively and summarily as a coverage interval with specified coverage probability, or as a standard uncertainty qualified with the effective number of degrees of freedom that it is based on.~~

~~The results obtained by some participants may be excluded from the calculation of the KCRV and associated uncertainty for substantive and documented causes, but not merely for their appearing to be outliers according to a statistical criterion. However, degrees of equivalence should be computed for all the participants, regardless of whether their results have been included in or excluded from the calculation of the KCRV.~~

~~Only CIPM key comparisons (carried out by a Consultative Committee or the BIPM) generate a key comparison reference value. Each key comparison reference value is considered to be a close approximation of the true value of the measurand. The approach used to select a method to determine the key comparison reference value is part of the protocol of the comparison and is agreed by the Consultative Committee or by the appropriate working group to which the Consultative Committee has delegated this task.~~

For RMO key comparisons and other subsequent key comparisons, the link to the comparison reference value is obtained by reference to the results from those institutes that have also taken part in the initial CIPM key comparison. A linkage procedure and associated uncertainty should be available in the final report. A degree of equivalence derived from such comparisons has the same status as one derived from an original CIPM key comparison. Participating institutes shall be listed only with one degree of equivalence per measurand. When an institute has acted as a linking institute, the degrees of equivalence obtained in the original comparison shall remain.

~~The degrees of equivalence of national measurement standards are understood as the degrees to which those standards are consistent with key comparison reference values and hence consistent with other national standards. A degree of equivalence is expressed quantitatively by two terms: a deviation from the key comparison reference value and an expanded uncertainty in that deviation, evaluated at a 95 % level of confidence (in practice, this is often approximated by using a coverage factor k equal to 2). The relationship between different degrees of equivalence and the key comparison reference~~

value could be represented in a so-called ‘graph of equivalence’. The (unilateral) degrees of equivalence (DoEs) gauge the consistency between national measurement standards and a KCRV. When there are n participants, the DoEs are n pairs $(D_1, U_{95\%}(D_1)), \dots, (D_n, U_{95\%}(D_n))$, where D_j denotes an estimate of the difference between the value measured by participant j and the KCRV, and $U_{95\%}(D_j)$ denotes an expanded uncertainty evaluated so that the interval $D_j \pm U_{95\%}(D_j)$ includes the true value of that difference with 95 % probability, for $j=1, \dots, n$.

The values of both D_j and $U_{95\%}(D_j)$ should be reported for each participant, regardless of whether the participant's measurement result was included in the calculation of the KCRV or not. The DoEs should also be depicted graphically to facilitate ascertaining by visual inspection whether the intervals of the form $D_j \pm U_{95\%}(D_j)$ include the value of magnitude 0.

Each expanded uncertainty $U_{95\%}(D_j)$ should express:

- the uncertainty associated with the KCRV;
- the reported uncertainty associated with the measured value;
- the correlation that typically exists between this measured value and the KCRV;
- contributions from any sources of uncertainty uncovered during the comparison, for example artifact drift, changes in measurement scales used by the participants, or mutual inconsistency between the measurement results.

The reliability of the KCRV and of the DoEs shall be corroborated using generally accepted diagnostics for the results of statistical procedures, for example *Leave-One-Out* versions of the degrees of equivalence [Duewer et al., 2014] [Koepke et al., 2017, §6], which are a form of *cross-validation* [Stone, 1974] [Mosteller and Tukey, 1977, 8E].

Leave-One-Out versions of the DoEs, $(D_{100,1}, U_{95\%}(D_{100,1})), \dots, (D_{100,n}, U_{95\%}(D_{100,n}))$, comprise the difference, $D_{100,j}$, between the value measured by participant j and the KCRV derived from all the selected measurement results except the result from participant j , and the associated expanded uncertainty, for $j=1, \dots, n$.

~~The Consultative Committees may decide that the degrees of equivalence can be expressed in relative values after normalization to the key comparison reference value or the nominal value of the measurand. At the discretion of the CC organizing the comparison, the results may also include the degrees of equivalence between pairs of measurement standards or between two measurement results (also known as “bilateral degree of equivalence”). For participants j_1 and j_2 , the bilateral degree of equivalence comprises the difference $D_{j_1, j_2} = D_{j_1} - D_{j_2}$ and the expanded uncertainty $U_{95\%}(D_{j_1, j_2})$, for j_1 and j_2 such that $1 \leq j_1 < j_2 \leq n$. When the KCRV and the unilateral degrees of equivalence~~

cannot be meaningfully computed, the $\{D_{j1,j2}\}$ are calculated as the differences between pairs of values measured by the participants.

Each expanded uncertainty $U_{95\%}(D_{j1,j2})$ should express (i) the uncertainties associated with D_{j1} and D_{j2} , hence, and in particular, significant contributions from any sources of uncertainty uncovered during the comparison, for example artifact drift, changes in measurement scales used by the participants, or mutual inconsistency between the measurement results (including "dark uncertainty"), and (ii) any correlation between D_{j1} and D_{j2} .

In special cases, the Consultative Committees may decide to express the unilateral degrees of equivalence in relative terms as $(D_j/M, U_{95\%}(D_j)/M)$, where M denotes either the KCRV or the nominal value of the measurand. Similarly for the bilateral degrees of equivalence.

8. Reporting of comparison

8.1 Results of measurements

The participating institutes shall report measurements results to the pilot institute as soon as possible after the measurements have been completed, and no later than six weeks after the participating institute's measurement period ends. The measured values, together with the associated measurement uncertainties and any additional information required, shall be reported in the format given in the instructions as part of the protocol, usually by completing standardized forms provided with the protocol instructions.

A result from a participant is not considered complete without its associated measurement uncertainty; measured values are not included in the draft report unless they are accompanied by a measurement uncertainty supported by a complete measurement uncertainty budget. Measurement uncertainties are estimated following the guidance given in the technical protocol.

If, upon examination of the complete set of results, the pilot institute finds results that appear to be anomalous, the corresponding participating institutes are invited to check their results for numerical errors but without being informed of the magnitude or sign of the apparent anomaly. If no numerical error is found, the result stands and the complete set is sent in a report to all participants according to Section 8.2

8.2 Comparison reports

Measurement comparison reports should be written to reflect the experiment that was actually performed, including summary results from all participants.

The report should include, or give reference to, most of the information specified in the Technical protocol. It should also include:

- a) measurement results identified for the individual participants;
- b) the key comparison reference value (reference value for supplementary comparisons) with a description how it was calculated (if applicable), or how the linking to the key comparison reference value was carried out;
- c) the degrees of equivalence and how these were evaluated (not mandatory for supplementary comparisons).

The pilot institute is responsible for writing the report of the comparison with assistance from the coordinating group (where such a group has been established). The report passes through three stages before publication, referred to as Draft A, Draft B and Final Report. The stages are differentiated by:

- Draft A being available only to the participants in the comparison;
- Draft B being available to the relevant Consultative Committee;
- Final Report being publicly available.

The first draft, Draft A, is prepared as soon as all the results have been confirmed by the participants according to Section 8.1. In the case of any outliers, the results are not communicated until the participants concerned have been contacted to ensure that no arithmetic, typographical or transcription errors are present. Draft A includes the results transmitted by the participants, identified by name, including the degrees of equivalence and, in the case of CIPM key comparisons, the proposed key comparison reference value.

The participants in the comparison may make comments on their own results and these may be modified if there were errors in the report of the result (typographical errors, different prefixes of units, transcription errors from the institute report to the Draft A report). In the case of results that are discrepant with the reference value or are not consistent with their published CMCs, the participants are not allowed to withdraw their results from the report unless a reason not attributable to the performance of the laboratory can be assigned (for example, if an excessive drift or a malfunction is detected in the transfer standard). Individual values and measurement uncertainties may be changed or removed or the complete comparison abandoned only with the agreement of all participants and on the basis of a clear failure of the transfer standard or some other phenomenon that renders the comparison or part of it invalid.

There may be several successive versions of a report (A1, A2, etc), but the Draft A stage will not be complete until all participants have agreed on the report. Draft A shall be

considered confidential and distributed among the participants only. As results can change, Draft A reports shall not be used to support CMC claims.

In calculating a key comparison reference value, the pilot institute will use the method considered most appropriate for the particular comparison (normally that proposed in the protocol), subject to confirmation by the participants and, in due course, the key comparison working group and the Consultative Committee, and will explain the reasons for the choice of method in the Final Report.

After selecting a procedure for the computation of the key comparison reference value and for the evaluation of the associated uncertainty, the degrees of equivalence (DoEs) will be computed consistently with the underlying statistical model and assumptions.

The expanded uncertainty that is part of each DoE may be computed as a suitable multiple (coverage factor, k) of the corresponding standard uncertainty, or in some other way.

In the former case, k depends both on the intended coverage probability, 95 %, and on the probability distribution that describes the uncertainty associated with the difference, D_j or $D_{j1,j2}$ (or with their LOO counterparts).

In the latter case, for example when $U_{95\%}(D_j)$ or $U_{95\%}(D_{j1,j2})$ are evaluated employing a Monte Carlo method, these expanded uncertainties will be derived directly from the sample produced by the Monte Carlo method, for example as one half the length of the shortest, symmetrical interval centered at D_j or at $D_{j1,j2}$ that includes 95 % of the sample values.

~~After deciding on the key comparison reference value and its uncertainty, the deviation from the reference value and the expanded uncertainty are deduced for each of the individual results.~~

Once the final version of Draft A is approved by the participants, the report becomes Draft B, which shall be submitted for approval by the corresponding Consultative Committee. The Draft B report should explain any changes that the reported uncertainties, or the degrees of equivalence, may have undergone since the approval of the Draft A.

The Draft B report of CIPM / RMO key comparisons can be used to support CMCs. At this stage, the measurement values are not considered confidential and may be used for presentations and publications. However, the key comparison reference value and the degrees of equivalence shall be considered confidential until they are approved by the Consultative Committee and published in the KCDB.

The working group on key comparisons is normally charged with examining a Draft B report. Unless the working group has been delegated full responsibility, it shall then

distribute the Draft B to all members of the Consultative Committee to ensure that it meets all the requirements set by the Consultative Committee.

Entry of the key comparison results, including the degrees of equivalence, into the KCDB must wait until Draft B has been approved by the Consultative Committee. At that stage, the “Draft B” in the title or contents should be replaced by “Final Report” and the report saved in portable document format (pdf). Each Consultative Committee will set its own procedures for approving the results of key comparisons in the most efficient and timely way possible. When the comparison report has been approved, the KCDB Office will be informed by the Executive Secretary or by the working group Chair concerned.

Supplementary comparison reports should follow the same three-stage process to approval: Draft A, Draft B, Final Report. The differences compared to key comparison reports are:

- a) approval for RMO supplementary comparisons is given by the corresponding RMO committee;
- b) degrees of equivalence relative to a supplementary comparison reference value can be computed, but this is not mandatory; it suffices to report the measured values and the standard uncertainties associated with them;
- c) Final Reports of supplementary comparisons shall be published in the KCDB in order to support CMCs.

The Final Report of RMO supplementary comparisons, approved by the RMO, shall be forwarded to the Consultative Committee Executive Secretary and the Chair of the relevant working group of the Consultative Committee for a six-week period of comment and editorial control. If no objections have been raised within this working group by the end of the period, the RMO TC Chair shall inform the KCDB office with a statement that the report has been approved. Those Consultative Committees that wish to discuss RMO supplementary comparison reports and formally approve them at the meetings of their relevant working groups may do so as an alternative.

8.3 Authorship of comparison reports

It is recognized that publications are measurable output of the work done by the participating institutes and of the contributions by the different participants in the research and/or comparison project carried out. Publication of original scientific research and method development, which may have occurred as part of comparison activity, is recommended in a separate peer-reviewed scientific journal, thereby reaching a wider audience.

In order to qualify as an author of the project/comparison report every individual shall have made a substantial intellectual contribution in at least one of the following activities:

- conception, experimental design and evolution of the project/comparison;
- original contribution to scientific research, having executed at least one or more significant aspects of the project/comparison;
- original contribution to analysis, interpretation and calculations of the measurement data;
- original contribution to authorship of the manuscript and documenting the project with all of its data and results.

The application of these criteria means that at least one person from every participating institute will qualify as an author, because at least one person will have carried out measurements and thereby contributed substantially to the execution of the comparison. Authors should be able to present, explain, and defend their contribution on their own, ~~to the project/comparison self-reliant~~ to outside experts and at scientific/technical conferences and workshops.

All co-authors should have been able to review the content of the article and have given consent for its release. They are jointly responsible for the quality and content of the publication. Authors need to have the authority from their managers to act as an (co-)author.

If desirable, a special section of the article may be used to acknowledge other people who have contributed to the project/comparison. This could pertain to important technical assistance, data collection, review of the manuscript or funding of the project/comparison (this may include heads of laboratories having made available the necessary means and having approved the execution of the project/comparison). General supervision of the project/comparison (for example, by laboratory managers) does not qualify as authorship.

This practice is not intended to overrule the internal rules and criteria for co-authorship that apply in the different participating institutes.

9. Disagreements

An institute that considers its result ~~unrepresentative~~ non-representative of its standards may request a subsequent separate bilateral comparison with the pilot institute or one of the participants. This should take place as soon as possible after the completion of the comparison in progress. The subsequent bilateral comparison is considered as a new and distinct comparison.

In the event of disagreement about the comparison results, or over the interpretation of results, which cannot be resolved by the participants themselves, the corresponding RMO TC/WG, the key comparison working group, or the Consultative Committee, the matter will be referred to the CIPM for a decision.

10. Publication in the KCDB

The Final Report shall be made available by the pilot via the KCDB web platform. In the KCDB, the graphs of equivalence and degrees of equivalence (when available) shall include results only from the institutes participating in the CIPM MRA.

It is recommended that the Final Reports of all comparisons ~~are~~ be published in a technical journal such as the Technical Supplement of *Metrologia* or other publicly available publication. If the report is to be published in the Technical Supplement of *Metrologia*, a dedicated form available on the KCDB web platform shall be completed and uploaded. For key comparisons, a list of degrees of equivalence and final results shall be provided according to the templates available on the KCDB website.

The details of the comparison publication process can be found on the KCDB website.

11. Monitoring the comparison results

If the results of a comparison are inconsistent with CMCs already declared in the KCDB, appropriate action shall be taken with these CMCs according to CIPM MRA-G-13.

If a participant in a comparison detects a discrepancy between its result in a comparison and related CMCs published in the KCDB after Draft B has been approved, that institute shall send a communication to the corresponding RMO technical committee and to the Chair of the RMO technical committee/working group responsible for approval of quality management systems.

If the pilot institute or any other participant detects a discrepancy between the results of an institute in a comparison and published CMCs, the pilot institute shall write to the institute alerting them to any potential problems in their results for the comparison. The communication shall be ~~copied to~~ shared with the participant's RMO technical committee and the Chair of the RMO technical committee/working group responsible for approval of institutes quality management systems.

In both cases, the communication shall also be copied to the Consultative Committee working group on CMCs with jurisdiction over the comparison, the JCRB Executive Secretary and the President of the Consultative Committee.

Within ninety days, the RMO shall write to the Consultative Committee's working group on CMCs, the JCRB Executive Secretary and the President of the Consultative Committee (with copy to the institute) stating the action plan for correcting any potential problems. In the next annual RMO quality management systems report to the JCRB, the results of the corrective actions should be included. In cases where the action plan fails to resolve the problems within six months of its detection, the RMO shall request the JCRB Executive Secretary to grey-out the existing CMCs from the KCDB. The RMO should request reinstatement according to the process detailed in CIPM MRA-G-13: "CMCs in the context of the CIPM MRA: Guidelines for their review, acceptance, and maintenance".

The Consultative Committee should inform the CIPM of the incident as part of its annual report.

12. Resources related to the CIPM MRA

CIPM-D-01, *Rules of procedure for the Consultative Committees (CCs) created by the CIPM, CC working groups and CC workshops.*

CIPM MRA (<https://www.bipm.org/en/cipm-mra/cipm-mra-text/>)

CIPM MRA-P-11, *Overview and implementation of the CIPM MRA*

CIPM MRA-P-12, *Coordination within the CIPM MRA: Consultative Committees, Regional Metrology Organizations, JCRB*

CIPM MRA-P-13, *Participation in the CIPM MRA: National Metrology Institutes, Designated Institutes, International organizations*

CIPM MRA-G-11, *Measurement comparisons in the CIPM MRA: Guidelines for organizing, participating and reporting*

CIPM MRA-G-12, *Quality management systems in the CIPM MRA: Guidelines for monitoring and reporting*

CIPM MRA-G-13, *CMCs in the context of the CIPM MRA: Guidelines for their review, acceptance and maintenance*

JCGM 100:2008, *Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM)*

ILAC-CIPM guidance on the accreditation of NMIs, *Joint ILAC–CIPM Communication regarding the Accreditation of Calibration and Measurement Services of National Metrology Institutes*

PG0128E1, *Customs Convention on the temporary importation of scientific equipment* (<http://www.wcoomd.org/en/about-us/legal-instruments/conventions.aspx>)

JCRB directory (<https://www.bipm.org/en/committees/jc/jcrb/>)

KCDB web portal (<https://www.bipm.org/kcdb>)

RMO websites (<http://www.afrimets.org>; <http://www.apmpweb.org>; <https://www.coomet.net>; <https://www.euramet.org>; <https://www.gulfmet.org>; <https://sim-metrologia.org>)

13. References

Comité International des Poids et Mesures (CIPM). *Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes*. Bureau International des Poids et Mesures (BIPM), Pavillon de Breteuil, Sèvres, France, October 14th 1999. URL www.bipm.org/en/cipm-mra/. Technical Supplement revised in October 2003.

D. L. Duewer, K. W. Pratt, C. Cherdchu, N. Tangpaisarnkul, A. Hioki, M. Ohata, P. Spitzer, M. Máriássy, and L. Vyskocil. “Degrees of equivalence” for chemical measurement capabilities: primary pH. *Accreditation and Quality Assurance*, 19:329–342, 2014. doi: [10.1007/s00769-014-1076-1](https://doi.org/10.1007/s00769-014-1076-1).

A. Koepke, T. Lafarge, A. Possolo, and B. Toman. Consensus building for interlaboratory studies, key comparisons, and meta-analysis. *Metrologia*, 54(3): S34–S62, 2017. doi: [10.1088/1681-7575/aa6c0e](https://doi.org/10.1088/1681-7575/aa6c0e).

F. Mosteller and J. W. Tukey. *Data Analysis and Regression*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.

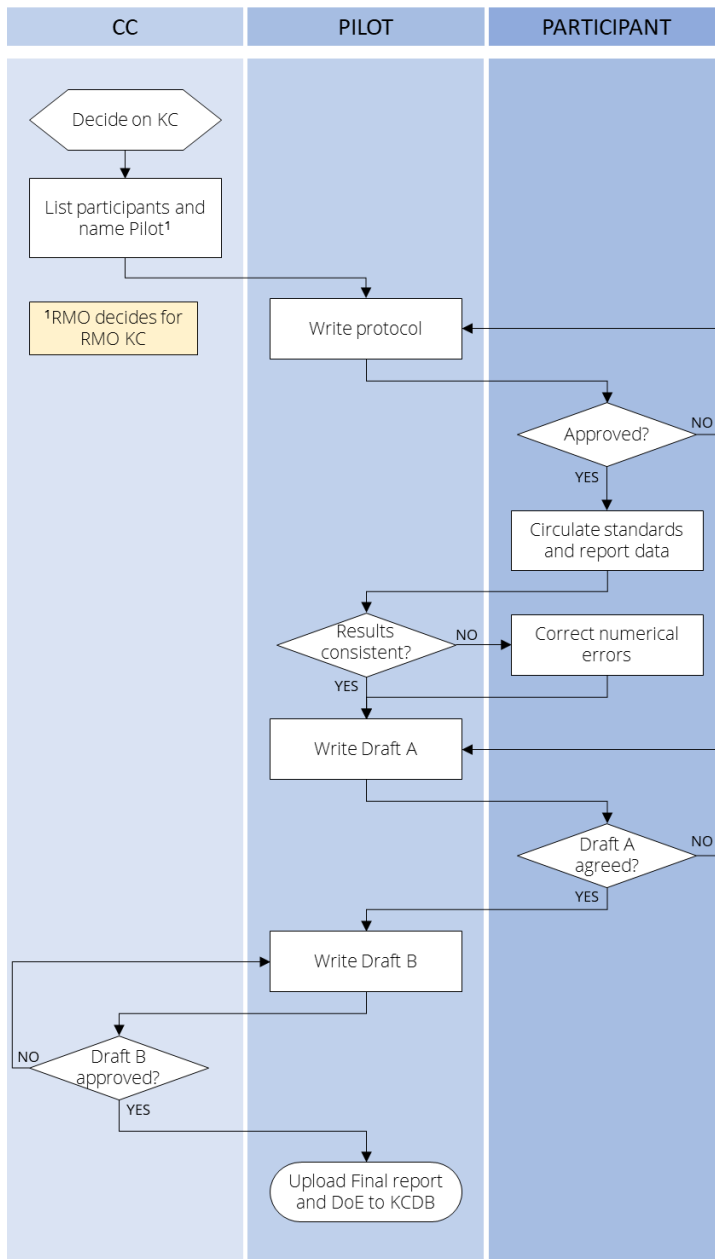
M. Stone. Cross-validated choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974. doi: [10.2307/2984809](https://doi.org/10.2307/2984809).

M. Thompson and S. L. R. Ellison. Dark uncertainty. *Accreditation and Quality Assurance*, 16:483–487, October 2011. doi: [10.1007/s00769-011-0803-0](https://doi.org/10.1007/s00769-011-0803-0).

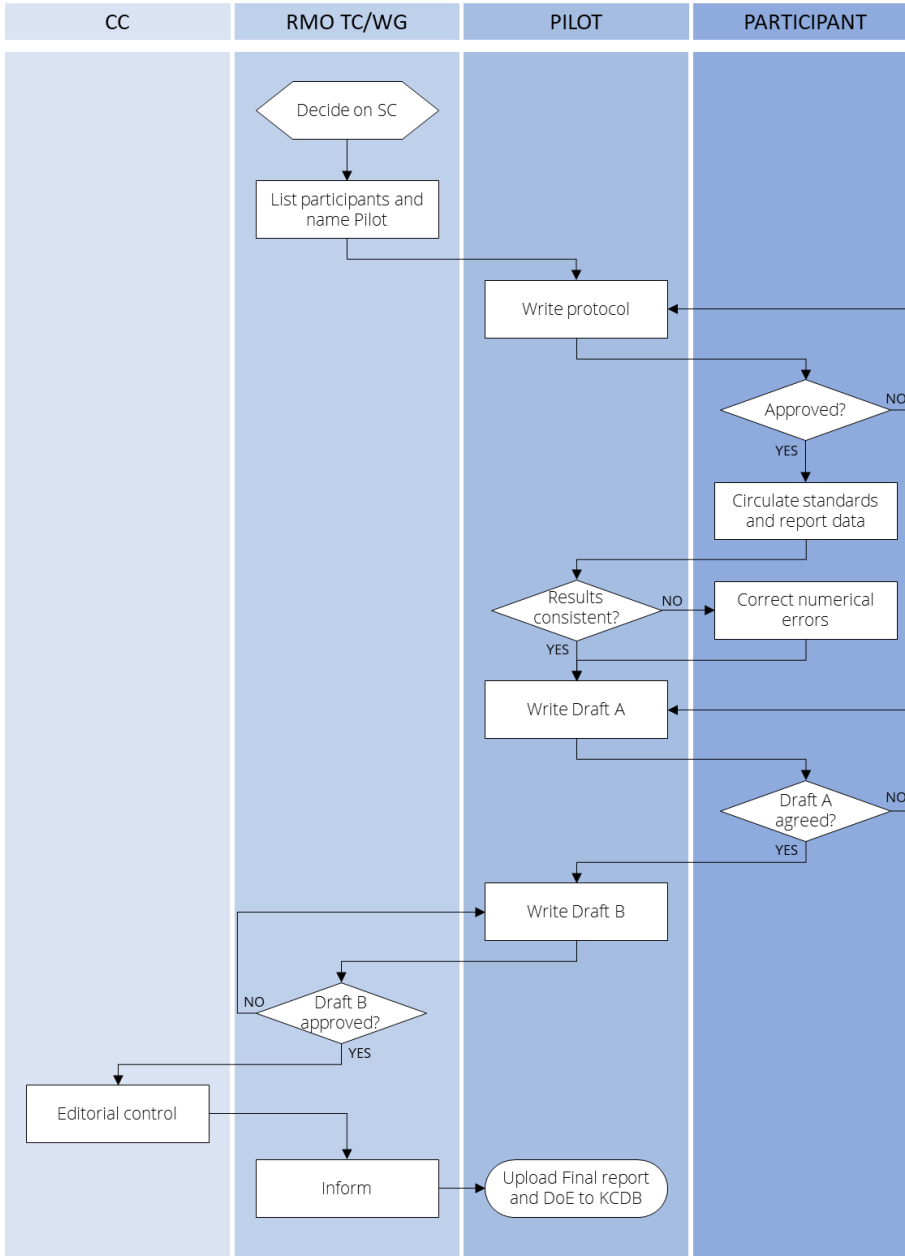
13.14. Revision History

Document and Version number	Date of Issue/ last review	Summary of change
CIPM MRA-G-11 V 1.0	11 January 2021	New document following the CIPM MRA review
V1.1	18 January 2021	Corrected flowcharts
V1.2	30 April 2021	Incorporation of suggestions from JCRB Task Group on Statistical Methods for Key Comparisons

Appendix A - Flowchart of CIPM and RMO key comparisons



Appendix B - Flowchart of supplementary comparisons



CIPM MRA-G-11 – Measurement comparisons in the CIPM MRA