

CCQM-GAWG guidance for designing model 1 and model 2 key comparisons

P J Brewer, R I Wielgosz, J Viallon, S Lee, C Cecelski, J van Wijk and A van der Veen

1. Introduction

This document provides guidance on designing key comparisons to critically evaluate and benchmark National Metrology Institutes' (NMI) and Designated Institutes' (DI) claimed competences for reference materials and capabilities for gas composition. It describes two models for organising CCQM-GAWG key comparisons (model 1 and 2) and how to develop a protocol to optimise the performance of the comparison for assessing the compatibility between the measurement results. It also describes the mechanisms for organising and linking to satellite key comparisons. This document supplements the guidance given in CIPM MRA-G-11 [1].

Figure 1 shows the number of model 1 and model 2 key comparisons that have been completed by the CCQM-GAWG since it was formed.

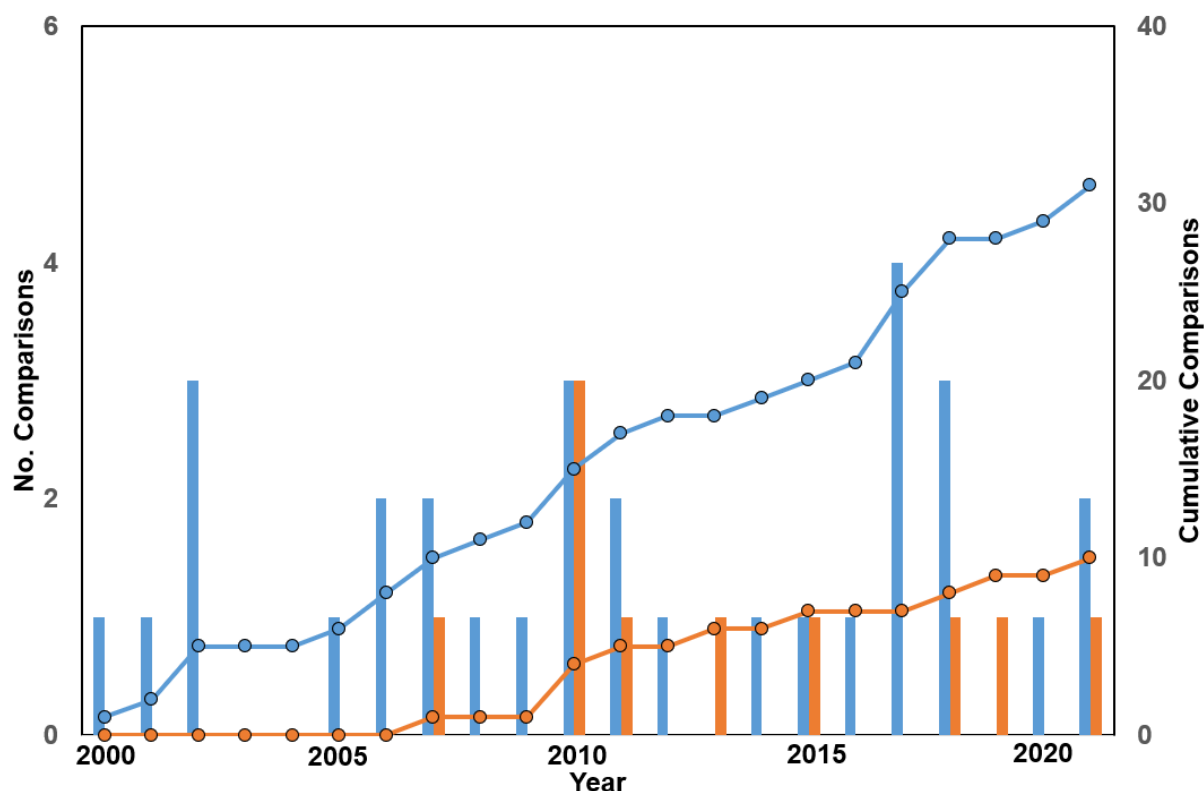


Figure 1 Number of model 1 (blue) and model 2 (orange) key comparisons completed by the CCQM-GAWG since 2000. The bars (corresponding to the left vertical axis) show the number of key comparisons of each model completed in any given year and the lines show the cumulative total (corresponding to the right vertical axis).

2. Key comparison models

Model 1

Model 1 key comparisons compare laboratories' abilities to assign a value to a measurand of a travelling standard using their in-house capabilities and standards. A suite of travelling standards is prepared or procured and value assigned by the coordinating laboratory and sent to participants. The measurands of the travelling standards are measured by the coordinating laboratory before and after despatch to evaluate their stability. If required, a function can be developed to make a correction

when the travelling standards are drifting during the key comparison. More guidance on designing stability studies is given in ISO 17034, ISO Guide 35, ISO/IEC 17043, and ISO 13528 [2-5].

Challenges for the coordinating laboratory include:

- Ensuring availability of the technical resources to produce or procure the travelling standards required for the comparison; arranging the logistics for shipment of the travelling standards;
- Ensuring availability of sufficient resources for the assessment of homogeneity and stability of the travelling standards, in particular, if drift needs to be characterised as a function of time.

Benefits of Model 1 comparisons include:

- Direct assessment of the calibration and measurement capabilities at the NMI/DI
- Simplicity of design and data treatment
- Limited additional development work for the coordinating laboratory if there is experience in producing standards like those envisaged as travelling standards.

Examples of Model 1 key comparisons are [CCQM-K117](#), [CCQM-K41.2017](#) and [CCQM-K112](#).

Model 2

Model 2 key comparisons compare the results assigned by laboratories to their own standards using the analytical capabilities of the coordinating laboratory. Travelling standards are prepared by the participating laboratories and sent to the coordinator for analysis. Participants are responsible for making stability measurements on their travelling standards before and after despatch and informing the coordinating laboratory of any changes.

A model 2 key comparison requires analysis of submitted travelling standards by a coordinating laboratory performing measurements with a stable measurement system which is ideally achieved by performing measurements under 'repeatable and reproducible' conditions, over a time period where the response characteristics of the instrument remain stable. It also needs to account for any instability in submitted travelling standards, at least in the uncertainty submitted by the participating laboratory.

Challenges for the coordinating laboratory include:

- Ensuring sufficient validation of their measurement system and uncertainty to be considered fit for purpose by participants.
- Ensuring appropriate comparison design and availability of appropriate data treatment tools, to develop reference values for each compared standard.

Benefits of Model 2 comparisons include:

- Direct assessment of the capabilities of providing a reference material by the NMI/DI
- Preparation costs of standards shared amongst participants
- Direct comparison of travelling standards produced by laboratories, with smaller uncertainties than generally achieved by model 1 comparisons, allowing discrepancies and uncertainties related to preparation of standards to be assessed more readily.

Examples of Model 2 key comparisons are [CCQM-K10.2018](#), [CCQM-K74.2018](#) and [CCQM-K137](#).

3. Selecting the key comparison model

An important stage in designing a key comparison is identifying which model will deliver the desired outcomes with optimal efficiency and minimal resources. Figure 2 shows a checklist of criteria to assist with selection.

Part A - model 2 comparison requirements	
1. Is the comparison aim to assess the comparability of laboratories' preparative capabilities?	Y N
2. Is the composition of one or more components likely to drift and if uncorrected impact results?	Y N
3. If Y to question 2, can coordinators develop a protocol to assist participants monitor drift?	Y N
4. Will the travelling standards contain more than one target component?	Y N
5. Will the travelling standards cause a systematic bias with different analytical techniques?	Y N
6. If Y to question 5, do coordinator(s) have capability to address potential biases?	Y N
7. Is gravimetric information likely to facilitate data analysis?	Y N
8. Can the coordinator achieve a KCRV uncertainty that meets laboratories' requirements for CMCs?	Y N
All green answers - coordinate model 2 or consider model 1 (Q9-15)	
Any amber answers - consider both models and select most suitable model	
Any red answers - consider model 1 (Q9-15) and do not coordinate model 2	
Part B - model 1 comparison requirements	
9. Is the comparison aim to assess the comparability of laboratories' analytical capabilities?	Y N
10. Is the composition of one or more components likely to drift and if uncorrected impact results?	Y N
11. If Y to question 10, does the coordinator have a capability to monitor drift?	Y N
12. Will the travelling standards cause a systematic bias with different analytical techniques?	Y N
13. If Y to question 12, can coordinator(s) provide sufficient information to participants to avoid any potential bias?	Y N
14. Can the coordinator source the travelling standards?	Y N
15. Can the coordinator achieve a KCRV uncertainty that meets laboratories' requirements for CMCs?	Y N
All green answers - coordinate model 1 or consider model 2 (if no red answers to Q1-8)	
Any amber answers - consider both models and select most suitable	
Any red answers - consider model 2 (if no red answers Q1-8) and do not coordinate model 1	
Any red answers in both parts - seek alternative coordinating laboratory	

Figure 2 Selection criteria for model 1 and 2 key comparisons.

The complexity of the key comparison can be classified by considering: (i) the stability of the component(s) of interest, and (ii) how many components will be present in the mixture, and whether the techniques that will be used for their analysis will impose any systematic bias in the measurement results (e.g. isotopic composition of the component of interest where instrumentation is used that does not discriminate different isotopocules or impact of differences in the composition of the matrix gas due to pressure broadening when using spectroscopy). Both of these issues were present and addressed in CCQM-K120.

An essential element in choosing the most appropriate model for the key comparison is to consider the requirements and capabilities of the participants. They will require the comparison to demonstrate analytical or preparative capabilities (or both). For key comparisons where a drift in the composition is expected, model 1 is preferred as the coordinating laboratory will be able to set up a measurement system to assess this under repeatability conditions for the majority of the time period of the comparison and apply appropriate correction. This assumes that the long- and short-term stability of the travelling standards is the same. For instances where the stability of standards retained by the coordinating laboratory is not representative of those sent to participants, further provision is required.

Considering the capabilities of the potential coordinator is another important step. This includes the performance of the comparison methods, access to metrological and statistical expertise, financial resources, and access to appropriate travelling standards (for a model 1 comparison). For comparisons where different techniques are likely to result in systematic biases, coordinators will either need to have capability to correct for them (model 2) or provide the appropriate information to participants (model 1).

It is also important to minimise the effort and resource required to achieve the desired outcome. The capabilities of the participating laboratories to be assessed and the statistical treatment of the data should be considered. Any comparison design that relies on the results of the participants for the Key Comparison Reference Value (KCRV) (relevant to both models) requires significant statistical knowledge to progress to a satisfactory outcome. Access to the gravimetric data may be beneficial for determining the KCRV. A benefit of a model 2 comparison is that the preparation costs of travelling standards are shared amongst the participants. This however often comes at the expense of additional (and substantial) effort surrounding the laboratory analysis work and data treatment.

4. Developing a key comparison protocol

After identifying the coordinator(s), defining the objective of the key comparison, and selecting a model for the key comparison, a technical protocol shall be prepared, specifying in detail the procedures necessary, including requirements for the standards being compared. It shall also include a template for reporting results which underpins their use in support for CMCs, for example inclusion of the data from the purity analysis of source gases used for preparing reference materials to support purity claims. The sections below provide guidance for model 1 and model 2 key comparisons.

Model 1

The duties of the coordinating laboratory in a model 1 key comparison can be summarised as follows:

- 1) to provide a comparison protocol
- 2) to prepare or procure a suite of travelling standards that are sufficiently homogeneous and stable to be used in the key comparison
- 3) to assign values and uncertainties to the measurand(s) to serve as KCRV(s) based on

- a. preparation, or,
 - b. a (potentially) primary method, or
 - c. a consensus value based upon the results submitted by the participants
- 4) to assess the proposed KCRV for bias with respect to the submitted results by the participants
 - 5) to provide appropriate packaging and labelling for shipping
 - 6) to provide the necessary shipping documents (e.g., MSDS) and permits (e.g., for customs clearance) to have the travelling standards shipped to the participant and back
 - 7) to perform the analysis of the data, provide a proposal for the KCRV(s) and calculate the degrees of equivalence
 - 8) to provide the data for guidance note for support of CMCs by the key comparison.

The coordinating laboratory is responsible for setting up a stable measurement system for the mixture composition(s) of interest over the time of the key comparison. It should determine the number of analytical techniques required to show up any systematic bias in the measurement system. This facility will be used to assign reference value(s) to each travelling standard with all standards analysed using the same techniques maintained under repeatability conditions.

The coordinating laboratory should monitor any changes in composition after preparation (set a frequency and interval for measurements). In some cases, the most significant changes may occur immediately after preparation, and hence an increase in the frequency of measurements in the first few days would provide a more detailed study of the initial behaviour of the mixture. A scheme for subsequent measurements and stability analysis should aim for the centre of gravity of the data to coincide with the measurement date(s) of the participants. The protocol should also include a schedule to measure the stability of the travelling standards after return from the participating laboratories. When required, a function of the drift in each travelling standard should be set up to allow a correction to be made at the time when participating laboratory made their measurements.

Maintaining sufficient cylinder pressure throughout the key comparison is essential as it impacts the composition of the gas in the travelling standards. The coordinator should communicate to the participants a minimum pressure limit.

Minimising the timescale of the comparison is important as it will reduce the time in which the coordinating laboratory needs to maintain a stable measurement system and monitor the drift in the travelling standards. Participants are required to follow the agreed time schedule so that the travelling standards experience as close to the same conditions as possible and stability measurements are performed over a similar time period. Failure to do so may result in being removed from the key comparison.

The coordinating laboratory and participants should identify if commutability of the travelling standards prepared by the coordinating laboratory is an issue that needs to be addressed in the comparison protocol. If any characteristics (e.g., isotopic composition or matrix composition) of the travelling standards are expected to cause a systemic bias in measurement systems at the participating laboratories, the protocol should have sufficient information to allow participants to follow their own procedures in this respect.

The coordinating laboratory must provide adequate information of appropriate handling measures of travelling standards to be carried out by participants (e.g., temperature equilibration and sampling).

Model 2

The duties of the coordinating laboratory in a model 2 key comparison can be summarised as follows:

- 1) to prepare a measuring system that has sufficient repeatability, reproducibility and sensitivity so that it can compare the travelling standards received from the participants with a sufficiently small uncertainty
- 2) to anticipate drift in the measuring system and design the comparison measurements in such a way that they are not affected by a drift in these results and that drift can be duly assessed
- 3) to assign values and uncertainties to the measurand(s) to serve as KCRV(s) based on
 - a. a consensus value when all participants measure the same nominal value, or a set of consensus values determined by a function when measurements are performed over a range of nominal values (minimum of 5) , or,
 - b. a value and uncertainty obtained from a (potentially) primary method
- 4) to return the travelling standards to the participants
- 5) to perform the analysis of the data, make a proposal for the KCRV(s) and calculate the degrees of equivalence.
- 6) to provide a comparison protocol
- 7) to provide the data for a guidance note for support of CMCs by the key comparison.

The coordinating laboratory is responsible for setting up a stable measurement system for comparing the travelling standard for the measurands. This facility will act as a comparator of all the travelling standards prepared by the participating laboratories, meaning that all will be analysed with the same technique(s) maintained under repeatability (or within-laboratory reproducibility) conditions and can be considered a priori as a valid calibration set. The coordinating laboratory should take sufficient measures for quality control and quality assurance to enable the robustness of stability measurements to be demonstrated. The measurement system(s) should be validated, and their performance presented by the coordinating laboratory to the extent reasonable and agreed by the participants prior to the start of the comparison.

For comparisons performed over a specified interval of nominal values with a central facility, the coordinating laboratory shall provide sufficient instruction to enable the development of a calibration function using the participants' mixtures as calibration standards. There are several examples where participants have produced travelling standards with different nominal compositions, while remaining within the order of magnitude of interest. In these instances, the protocol stated the tolerance on the required nominal compositions to allow appropriate spacing of the measurand(s) of the travelling standards over the specified interval(s) for, e.g., regression. In CCQM-K68.2019, the range of compositions was split into two sub-intervals above and below a median to ensure the values from each institute were appropriately distributed over the analysis function.

The coordinator can provide instructions to the participants to monitor the stability of the travelling standards and allow participants to submit an adjusted measurement result. It is important that the participants can provide the coordinating laboratory with a value at any stage in the comparison if they require it for stability. If necessary, the coordinating laboratory can provide instructions on how participating laboratories should monitor any changes in composition (frequency and interval of measurements). If participants do not follow those instructions, the participant may be excluded as

not to tarnish the design and operation of the key comparison. The coordinator should provide instructions in the protocol for reporting these stability monitoring data.

Specifying a sufficient cylinder pressure throughout the key comparison is essential as it impacts the composition of the gas in the travelling standards.

Setting up a feasible timescale for the comparison is important as it will reduce the time in which the coordinating laboratory needs to maintain a stable measurement system and monitor the drift in the travelling standards. An attempt should be made to ensure participants follow the same time schedule so that the travelling standards experience similar conditions and stability measurements are performed over a similar time. Where possible the protocol should impose the same start date for all participants and consider a plan of action if the deadlines are not met.

The coordinating laboratory and participants should identify if commutability of standards submitted to the coordinating laboratory is an issue which needs to be addressed in the comparison protocol. If any characteristics (e.g., isotopic composition or matrix composition) of submitted standards are expected to cause a systemic bias in measurement system of the coordinating laboratory, the protocol should have information on how this will be dealt with, including as appropriate measures for mitigating such effects.

The coordinating laboratory should use validated procedures to ensure appropriate handling and equal treatment (e.g., temperature equilibration and sampling) of travelling standards upon receipt. The order of the measurements should also be considered and the number of separate batches of measurements if required.

5. Satellite key comparisons

Track A key comparisons [6] are designed to test the core skills and competencies required in gravimetric preparation, analytical verification, and purity analysis of gas reference materials. The CCQM-GAWG supports and interacts with RMOs via satellite Track A comparisons, to work with efficiency while ensuring the needs of all metrology institutes are met. Figure 3 shows the four different models in which satellite RMO comparisons may be performed from a CCQM Track A key comparison. The first example (CCQM-K111) employed model 1-1, where the coordinating laboratory sent out travelling standards to participants in the CCQM comparison. When the key comparison was complete, selected participants sent out travelling standards for the RMO satellite comparisons. In this example, the coordinator of the CCQM comparison also used the same suite of gas mixtures to coordinate the RMO comparison. To avoid collusion, the results of the key comparison were embargoed until the RMO comparison was complete.

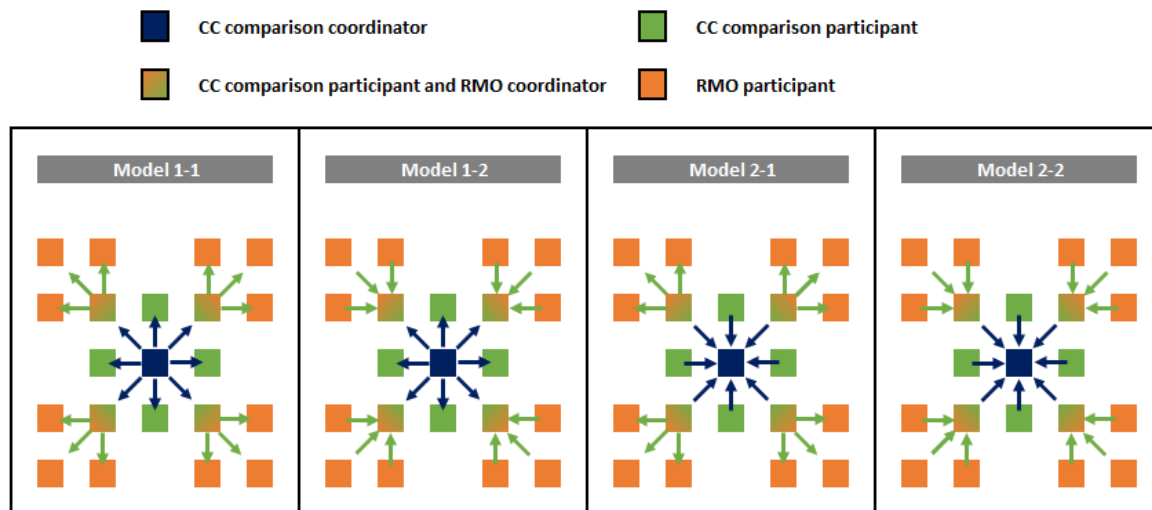


Figure 3 The four models for operating RMO satellite comparisons under the Track A scheme.

The following criteria impact the choice of model:

- Capabilities to be assessed by the satellite comparison
- Effort/burden on the key comparison and RMO comparison coordinators
- Timescale for running the comparison
- Robustness of linking the RMO comparisons to the CCQM comparison
- Extent of preparative capabilities required by coordinators and participants
- Data to inform the calculation of the KCRV

The selection of operating model will depend on the factors above as well as the individual needs of the coordinating and participating laboratories. Model 1-1 will require the preparation of travelling standards by both the CCQM and RMO comparison coordinators. The other three models feature at least one comparison where the participants will prepare the travelling standards. Model 2-1 should require only one suite of travelling standards per RMO as the coordinators can use those prepared for the CCQM comparison in the subsequent model 1 RMO comparison. Models 1-2 and 2-2 have the potential for some parallel running.

For Model 1-2, it may be more challenging to provide the link between the CCQM comparison and the RMO comparisons with two different models being applied (one assessing analytical capabilities and the other assessing preparative capabilities). Although 2 different models are used in model 2-1, linking should be robust as the RMO coordinators are preparing the reference materials in both parts. The travelling standards sent into the CCQM comparison link the set that will be sent out for the RMO comparison in each case. Models 1-1 and 2-2 are also well linked as they involve all preparative capabilities.

An example of model 2-2 is CCQM-K120b and the satellite comparison COOMET.QM-K120b. It was important that the reference values in the satellite comparison were linked to CCQM-K120b. The key requirement is that the result from the RMO coordinating laboratory in CCQM-K120b should be reproducible over the timescale of both comparisons and uncertainty stated associated with linking (to reproduce the same value in the original key comparison).

Until we have examples of the other 2 models put into practice in the CCQM-GAWG, it is difficult to comment further on the merits and drawbacks of each.

6. On-demand (on-going) key comparisons

For several gas components, comparisons of standards can be most effectively and efficiently achieved by comparing to a common and stable reference. This comparison service is provided by the BIPM laboratories. It is organised as an on-going series of bilateral comparisons against a stable reference facility also maintained at the BIPM laboratories. The on-going series of bilateral comparisons that are currently provided or in validation stage are BIPM.QM-K1 (ozone in air), BIPM.QM-K2.a and b (CO₂ in air/ nitrogen), BIPM.QM-K3 ($\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ in CO₂), BIPM.QM-K4 ($\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ in CO₂ in air), BIPM.QM-K5 (CO₂ in air, scales) and BIPM.QM-K6 (NO₂ in nitrogen). The results of the comparisons are reviewed and approved within the CCQM-GAWG. As the on-going comparison capabilities are continually available, it allows NMIs and DIs to request a comparison when they require one, enabling an on-demand assessment of their measurement capabilities and support of calibration and measurement capabilities.

7. References

- [1] Measurement comparisons in the CIPM MRA, CIPM MRA-G-11, BIPM – Pavillon de Breteuil, F-92312 Sèvres Cedex, 2021.
- [2] General requirements for the competence of reference material producers, ISO 17034, 2016.
- [3] Reference materials - guidance for characterization and assessment of homogeneity and stability, ISO Guide 35, 2017.
- [4] Conformity assessment - general requirements for proficiency testing, ISO/IEC 17043, 2010.
- [5] Statistical methods for use in proficiency testing by interlaboratory comparison, ISO 13528, 2015.
- [6] CCQM/GAWG/19-41 CCQM-GAWG strategy for comparisons and CMC claims.