Guidelines for CCPR Key Comparison Report Preparation

CCPR Key Comparison Working Group

CCPR G2, Rev.3, Draft 2.0, 2011.2.8

These guidelines are prepared by CCPR WG-KC and approved by CCPR to ensure that reports of CCPR Key Comparisons are prepared in fair and uniform manner. This document is to supplement the CIPM guidelines on measurement comparisons [1]. This document does not cover RMO Key and Supplementary Comparisons and bilateral comparisons, for which other Guideline documents are available.

1. Pre-Draft-A Process 1: Verification of reported results

1.1 After the results have been submitted by all the participants and the measurements of the pilot lab have been completed, within two months, the pilot lab sends to each participant, individually, their reported values as received by the pilot lab for verification.

1.1.1 Each participant reviews their reported results received from the pilot lab and examine if there are any errors. If any clerical errors made by pilot lab are found, the participant should correct their results at this stage.

1.1.2 Each participant must respond to the pilot lab within three weeks from receiving the verification data, to confirm that there is no problem in their data or to request any corrections.

1.1.3 After this process period is over, any corrections of participants' reported results due to errors or misinterpretation by pilot lab are not allowed.

2. Pre-Draft-A Process 2: Review of uncertainty budgets

2.1 The uncertainty budget (table of uncertainty components and uncertainty contributions, as well as descriptions of measurement technique and facility) must be submitted from each laboratory together with their results. The overall uncertainty values alone will not be sufficient. If the uncertainty budget received is not complete, the pilot lab contacts the participating lab to request for complete uncertainty budget.

2.1.1 Specific instructions on reporting the uncertainty budget must have been given in the protocol of the comparison that was agreed before the start of comparison.

2.1.2 If a participant fails to provide their uncertainty budget in the required detail within a given deadline, the pilot lab may request WG-KC to approve that the participant be removed from the comparison, or from the calculation of the key comparison reference value, as appropriate. (In this case, the fact will be stated in the final report).

2.2 After all the results with uncertainty budgets from all the participants have been submitted, the pilot lab distributes to all participants the uncertainty budgets of all the participants to allow them to review other labs' uncertainty budgets. This is done within two months from receipt of all information.

2.3 Any participants including the pilot lab can send questions or comments on other participant's uncertainty budgets and ask for further information, for example, if a

participant's uncertainty is considered unusually small, or if some important uncertainty components are missing.

2.3.1 Comments/questions from any participants are accepted within six weeks from distribution of the uncertainty budgets.

2.3.2 Comments/questions should be sent to the pilot lab, who will then forward the comments anonymously to the participant being asked and copied to all other participants. The pilot lab takes the records of all communication.

2.4 The participants who received comments must respond promptly and, if necessary, can revise their uncertainty budget. This, however, does not force the participant to revise it. At this stage, any participants can submit correction of their uncertainty budget, even without receiving comments. However, revision of uncertainty components is allowed only in the direction to increase the overall uncertainty.

2.4.1 Responses to comments and revisions of uncertainty budgets (if any) are accepted within two months from distribution of the uncertainty budgets.

2.4.2 Replies to comments should go to the pilot lab and forwarded.

2.4.3 If any correction or changes of the uncertainty budget is submitted in this stage, the changes of values and the reason will be reported in the appendix of the comparison report.

3. Pre-Draft-A Process 3: Review of Relative Data

3.1 After the results have been submitted from all the participants and the measurements of the pilot lab have been completed, within two months, the pilot lab prepares "Relative Data" of each participant, which are the data reduced to show only the stability of transfer standards for each participant before and after travel and the internal consistency of all the transfer standards measured at each participant lab.

3.1.1 The Relative Data can be obtained by calculating the ratios of values of all transfer standards measured by the participant and by the pilot lab, and normalizing the ratios to their mean. This normalization removes any relationship of the participant's absolute scale to the pilot lab, and leaves only internal consistency information. (For spectral data, the normalization is done at each wavelength.) See Appendix A for an example.

3.1.2 Any data reflecting the relationship of the absolute scales between participant and pilot are not allowed to be disclosed in this Pre-Draft A process.

3.1.3 The pilot lab distributes the Relative Data of all participants to all participants.

3.2 The participants review the Relative Data and examine the stability of the transfer standards (before and after transportation). If significant changes or drifts in any of the transfer standards are identified, the participant can propose removal of the data of the particular transfer standards, or if necessary, re-measurement of the transfer standards. The participants can also examine the internal consistency of the Relative Data to confirm that their measurements (and pilot lab's measurements) of all samples went well.

3.2.1 Removal of results or re-measurement is discussed and agreed by the participant and the pilot lab, and all participants will be informed of such decisions.

3.2.2 Each participant must respond to the pilot lab within one month from distribution of Relative Data, to confirm that there is no problem in their data or to request any corrections. All the participants should respond, but if no response is received by the deadline, the original data stand.

3.2.3 Re-measurement can be done only when it is absolutely necessary and when it will not delay the schedule of the comparison significantly.

3.2.4 If a participant finds any errors (clerical, technical, or any other reasons) in their reported values, from any anomalous feature in their Relative Data, the participant can submit a correction of the results at this stage. But, it is the participants' responsibility to identify any anomalous feature of their Relative Data that imply errors. If the pilot lab finds obvious anomalous results of any participant that cannot be identified from Relative Data, a warning may be sent to all participants (without specific information).

3.2.5 If any corrections of data are submitted from participants in this stage, the changes of values and the reason will be reported in the appendix of the report of comparison.

3.2.6 If data of any transfer standards are removed, the fact will be stated in the report of comparison.

Note: Processes 1, 2, and 3 above can proceed simultaneously.

4. Pre-Draft-A Process 4: Identification of outliers and consistency check

4.1 If, after calculating a tentative KCRV and results for all participants, obvious outlier(s) are observed in the comparison results, the pilot lab should discuss with all the participants removal of such data from the calculation of the KCRV before distributing Draft A.

4.1.1 The pilot laboratory can propose to all participants exclusion of data (from calculation of KCRV) that are visibly obvious outliers (e.g., the deviation from the KCRV is larger than 3 times of the expanded uncertainty of the deviation with k=2), by providing relative data, e.g, the list of ratios of deviation from KCRV (tentative) to the uncertainty (k=2), of all participants without identification of the laboratories. Note that, in this communication, any data that allows participants to figure out identification of the outlier lab(s) should not be distributed.

4.1.2 The outlying data will be excluded from the calculation of the final KCRV, with agreement by all participants.

4.1.3 If a consensus is not reached and if the pilot lab still believes exclusion is necessary, the pilot can consult WG-KC.

4.2 After removing outlier(s), if any, the pilot laboratory should analyze the data following the recommendations given in section 5 to calculate a tentative KCRV, and perform the consistency check for the weighted mean (calculate the Birge ratio or Chi-square test).

4.2.1 If the data do not show reasonable consistency (e.g., the Birge ratio is considerably higher than 1), the pilot lab should propose to use the Mandel-Paule method [4], applying an additional "interlaboratory variance" s^2 that will force the Birge ratio to 1. (See step 6. in Appendix B for an example of

this calculation.) The origin of this variance, in the pilot lab's interpretation, should also be explained if possible.

4.2.2 For discussion with participants, the original Birge ratio value(s) can be distributed to participants. The use of the Mandel-Paule method should be agreed by all participants before the distribution of Draft A.

4.2.3 If a consensus is not reached and if the pilot lab still believes the use of Mandel-Paule method is necessary, the pilot can consult WG-KC.

4.3 After applying or considering 4.1 and 4.2 and if the pilot lab still finds serious problems on the results using the default data analysis method given in section 5, the use of other methods (e.g., simple arithmetic mean) can be discussed with the participants before distribution of Draft A, and the steps described in 5.3 shall be taken if a different method is to be used.

5. Preparation and Distribution of Draft A

5.1 After the Pre-Draft A processes are complete, the pilot lab prepares and distributes Draft A to all the participants, which discloses the absolute results of the comparison with identification of all the participating labs. The Draft A should tabulate all the results as well as present them in graphical form as necessary. It is recommended that the pilot lab also distribute the data of the analyses in a spreadsheet file. The Draft A should be distributed within six months after completion of all the measurements of the comparison.

5.2 Draft A should give the designation of the comparison as CCPR-Kx.YEAR (e.g., CCPR-K6.2009). YEAR is the year of registration to KCDB.

5.3 The default method for calculating KCRV is the weighted mean with cut-off. Use of other methods can be discussed only when the pilot lab finds serious problems in using the default method, and should be discussed before distribution of Draft A. Other method may be used with consensus of all the participants and subsequent approval of WG-KC.

5.3.1 The cut-off value for the uncertainty, as a default, is determined as the average of the uncertainty values of those participants that reported uncertainties smaller than or equal to the median of all the participants. (For example, if there are 10 participants, the cut-off value will be the average of the 5 smallest values of uncertainty.)

5.3.2 The use of a cut-off value other than the default, if necessary, should be discussed and agreed by all participants before Draft A is distributed. Follow also 5.3.4.

5.3.3 The weights are determined based on the participants' reported uncertainties adjusted by the cut-off, combined with the transfer uncertainty of the comparison (reproducibility of measurements at the pilot lab and other components associated with difference in measurement conditions between pilot and participants, etc.).

5.3.4 When discussing use of other methods, the pilot lab must be careful not to disclose the results of the comparison while providing some data for discussion. (For example, it is acceptable to disclose standard deviation of

the results and the average of the stated uncertainties, or Birge ratio, etc. The plots of absolute results with uncertainty bars even with anonymous lab identification must not be distributed for the discussion because lab identification might be inferred from such data. The ratios of deviation from KCRV and stated uncertainty (k=2) of each lab (without identification) can be plotted).

5.4 The data analysis should be as simple as possible, and the calculation process should be made transparent so that the final results can be reproduced by others, without difficulty, from NMIs' reported measurement results included in Draft A. The data analysis program and intermediate results should be made available for all participants. The approach used should be agreed by all participants before the publication of Draft A.

5.4.1 An example of a commonly used data analysis as described in 5.3 is provided in Appendix B, which is a step-by-step approach. The calculation process may be elaborated as necessary for each comparison.

5.4.2 Alternative calculation techniques based on the least-square model approach (e.g., [2]) may also be used if they comply with 5.3 and 5.4.

5.5 The authors of the reports of CCPR KCs will normally be the staff members of the Pilot lab who worked to conduct the comparison, and the decision will be made by the pilot lab.

6. Review of Draft A by participants

Each participating lab carefully reviews all the data presented in Draft A, and reports to the pilot lab if they find any clerical errors made by the pilot lab or send any other comments. Comments should be sent within two months from distribution of Draft A.

6.1 After Draft A has been distributed, correction of the results (reported values and uncertainty values) due to errors by participating labs, for any reason, cannot be accepted.

6.2 Once Draft A has been distributed, the whole or any part of participant's results cannot be withdrawn even if they are found in error. Under special circumstances, it might be allowed if it is approved by CCPR.

6.3 If a participating lab has found error(s) that they made in their measurements or in data analysis that affected the reported results, the fact should be reported to the pilot lab. The corrections are documented in the appendix of the report. In this case, under the pilot lab's decision, the lab's results (or part of the results) may be excluded from the KCRV calculation, with the fact stated in the report.

6.4 A change of the method for calculating KCRV, if necessary, should be discussed and agreed in Pre-Draft A stage (section 5.3). However, if such discussion did not take place and Draft A shows serious problems, it can be changed with consensus of all the participants and subsequent approval by WG-KC.

6.5 Removal of partial results should be discussed in Pre-Draft A stage, and it is not allowed at this stage except when the problem in transfer standard(s) was not clearly shown in Pre-Draft A stage and with consensus of all the participants.

6.6 If comments are made by one or more participants, these comments should be

circulated to all participants, and if they are significant, the Pilot lab can discuss with participants whether and how changes are to be made for the next Draft A version. If necessary, further data can be distributed as Supplement to Draft A. When changes are made to address comments, the revised draft will be called Draft A-2 and will be distributed again to all the participants for approval. In this case, the revised draft should be distributed within two months from closing comments. If further comments are made to the revised draft, the process can be repeated (Draft A-3, ….) or the Pilot lab can consult WG-KC in case of dispute. When all the participants approve the Draft A-x, it will become Draft B (see section 7).

6.7 Draft A is considered as confidential for only the participants. The data in Draft A shall not be distributed or presented to general public.

7. Preparation of Draft B

When the final version of Draft A has been agreed by all participants, it becomes Draft B. The Pilot lab submits Draft B to WG-KC for approval, within four months from distribution of Draft A (if no further version of Draft A need to be prepared).

7.1.1 Draft B (the final version of Draft A-x) must include tables of unilateral Degrees of Equivalence. Tables of Bilateral DoE are not required. The tables can be in the main body or an Appendix of the report.

7.1.2 Draft B will be reviewed by WG-KC (and no longer by participants). As the result of review, changes in Draft B may be requested to the Pilot lab. If a revision is produced, it is called Draft B-2 (B-3, … if repeated) and reviewed again by WG-KC. Participants do not participate in this process unless some major revision is proposed by WG-KC. When Draft B-x is approved by WG-KC, it will be submitted to CCPR. When it is approved by CCPR, the approved version of Draft B becomes the Final Report.

7.1.3 Any versions of Draft B are not considered confidential, and may be the subject of a publication with the exception of the proposals for the reference value and degrees of equivalence.

8. Publication of Final Report

The final reports of Key Comparisons will be published in the *Technical Supplement of Metrologia* (electronic media on the website). If the Pilot lab chooses to do so, the reports can also be published in a printed journal.

8.1 After the final report has been published, the Pilot lab sends to all participants, with copies to the RMO P&R TC chairs, a reminder to check the consistency of their CMCs with the KC results and to report to the participant's RMO TC chair (with a copy to the pilot) about their evaluation and any proposed actions in case of inconsistency, within two months from the reminder.

Recommended Time Line

Month 0: Pilot Lab receives all the results with uncertainty budgets and finishes all the

measurements.
Month 2: Pilot Lab communicates with each participant to verify the received measurement results (section 1.1). Responses due in 3 weeks.
Month 2: Pilot Lab distributes the uncertainty budgets of all the participants to all participants (section 2.2). Comments due in 6 weeks.
Month 2: Pilot Lab distributes Relative Data to all participants. (section 3.1) Responses due in one month.
Month 3: Responses to Relative Data from all participants due (Note if no response is received by pilot lab by this deadline, original reported values stand).
Month 3.5: Comments on the uncertainty budgets closed.
Month 4: Responses to comments on uncertainty budgets and revision of uncertainty closed.
Month 6: Draft A is completed. In case outliers need to be removed following the procedure described in section 4, this step shall be completed within one month, for which the time of completion of Draft A may be adjusted.
Month 6: Draft A distributed (section 5.1). Approval/comments due in two months.
Month 8: Comments on draft A due.
Month 10: Draft B submitted to WG-KC (section 6) (approval due in 6 weeks). Or, Draft A-2 distributed to participants (comments due in one month).
Month 11.5: Draft B approved by WG-KC. (Or, comments sent to Pilot, requesting revision in one month.)
Month 11.5: Draft B submitted to CCPR (approval due in 6 weeks.).
Month 13: Final Report published.

Due date for comments after revision of Draft A or Draft B may be adjusted depending on the degree of changes.

The progress of each CCPR comparison will be monitored by WG-KC and reminders will be sent to Pilot lab if schedule is significantly delayed from the recommended time line.

References

1. CIPM MRA-D-05, Measurement comparisons in the CIPM MRA, Version 1, available at http://www.bipm.org/utils/common/CIPM_MRA/CIPM_MRA-D-05.pdf
2. Appendix B “A Guide to the Analysis Approach”, CCPR K1.a Report (2005).
3. Consensus Values and Weighting Factors, R.C. Paule, J. Mandel, J. Res. NBS, Vol.87, n° 5, 1982, 377-385.

Appendix A: An example of Relative Data

An example is given here for an intercomparison of spectral responsivity, where three detectors (NIST04, 08, 10) were used as transfer standards. The detectors were measured by the pilot lab, then by a participant, then by the pilot lab again. Figure A1 shows the plots of the absolute ratios of responsivity values of the three detectors measured by a participant (Lab-1) and Pilot Lab (before and after transportation). So, there are six points at each wavelength. From this, the pilot lab sees an obvious anomaly for detector NIST10 at 900 nm. However, the pilot lab does not know yet if it is a numerical error by Lab-1 or some problem caused by the detector. These absolute results, of course, must not be sent to participants before Draft A. Instead, Relative Data is sent to the participant to let them identify the problem.

The Relative Data can be calculated as below. The Lab/Pilot ratio in Fig. A1 at each wavelength is denoted $R_{i,j}$ for transfer standard i (i=1 to 3 in this example) and round j (j=1 for Before, j=2 for After). The Relative Data $R'_{i,j}$ are calculated by

$$R'_{i,j} = \frac{R_{i,j}}{mean(R_{i,j})}$$

Figure A2 shows the plots of the Relative Data for this example. The six values at each wavelength are normalized in such a way that the average of the six values at each wavelength is always 1. Therefore, the relationship of the scales between Lab-1 and Pilot lab is removed at each wavelength. Only the internal consistency of measurements of three transfer standards is presented. The consistency between rounds is also indicated.

By examining the Relative Data, Lab-1 finds the anomaly at 900 nm, but confirms that all other data are

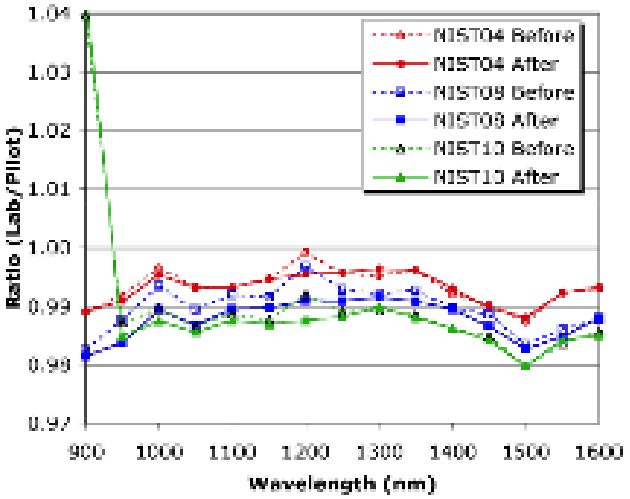

Figure A1. Plots of absolute results in the ratio (Lab-1/Pilot lab)

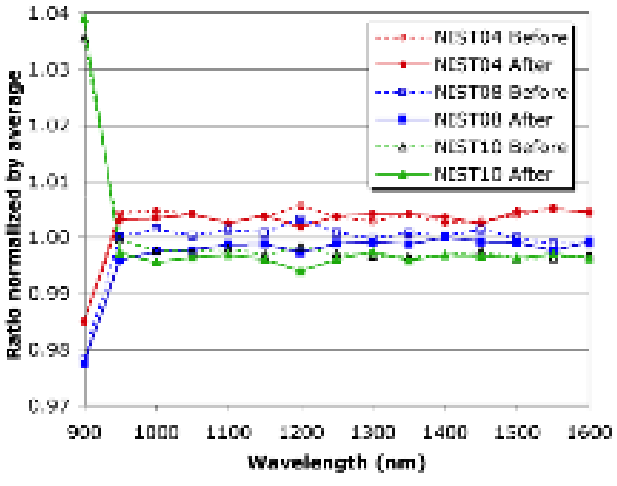

Figure A2. Plots of Relative Data of Lab-1

fairly consistent. It can also be seen that all detectors reproduced well before and after transportation. Lab-1 checks their results at 900 nm. If they find any error (numerical or technical) on this point, they can correct this value. If not, they might suspect some problem of detector NIST10 at this wavelength, and can request removal of the data. Or, if the pilot lab sees some common problems, pilot lab can propose to all participants removal of some detector at particular wavelengths, then participants can look at their Relative Data to see if it is reasonable or how it may affect their results. Such request and/or discussion can be done in a fair manner using the Relative Data without disclosing absolute results.

Appendix B. An example of a commonly used data analysis for an intercomparison

Below is an example of commonly used data analysis for an intercomparison of a spectral quantity. Measurements at each wavelength are taken as each separate comparison. The same analysis will apply to the results at all wavelengths. This example assumes a case as follows. Three lamps of the same type were prepared by each NMI and measured by the NMI, then measured at Pilot, then measured at the NMI again. The two measurements at each NMI (before and after Pilot) are referred to as round 1 and round 2. The total uncertainty of measurement for each lamp at each round is reported. The total uncertainty and reproducibility of Pilot lab measurements for each lamp are reported. In this method, simple arithmetic means are taken in all the intermediate steps for the results from three lamps and two rounds within each NMI, then weighted mean with cut-off is applied at the last step as agreed by CCPR.

The following notations are used:

N	Number of participant NMIs, not counting the Pilot lab.
$E_{i,j,r}$	Spectral irradiance of lamp j (=1 to 3) of NMI i, measured by the NMI in round r (=1 to 2).
$u_{\mathrm{rel}}(E_{i,j,r})$	Total relative uncertainty of $E_{i,j,r}$ reported by the NMI.
$E^{\mathrm{P}}_{i,j}$	Spectral irradiance of lamp j (=1 to 3) of NMI i, measured by the Pilot.
$u_{\mathrm{rel}}(E^{\mathrm{P}}_{i,j})$	Total relative uncertainty of $E^{\mathrm{P}}_{i,j}$.
$u_{\mathrm{rel}}(E^{\mathrm{PR}}_{i,j})$	Reproducibility of Pilot measurements of lamp j of NMI i, including the stability of the comparison scale during the period of comparison and repeatability of the transfer lamp.

1. For each NMI i for each lamp j, the NMI measurements of two rounds are averaged:

$$\overline{E}_{i,j} = \frac{1}{2}\sum_{r=1}^{2} E_{i,j,r} \tag{1}$$

and its uncertainty by

$$u_{\mathrm{rel}}(\overline{E}_{i,j}) = \frac{1}{2}\sum_{r=1}^{2} u_{\mathrm{rel}}(E_{i,j,r}). \tag{2}$$

Note1: This uncertainty calculation is an approximation, assuming that the results from the two rounds of the same lamp measured by the same NMI are nearly fully correlated. This is normally the case when the uncertainty of transfer measurements (random components) is much smaller than the uncertainty of the scale.

Note2: If the uncertainty of measurements are reported separately for the uncertainty of the scale $u_{\mathrm{rel}}(E^{\mathrm{S}}_{i})$ of the NMI and the transfer uncertainty $u_{\mathrm{rel}}(E^{\mathrm{T}}_{i,j,r})$ for the particular measurement, the uncertainty of the average of M rounds (M=2 in example above) is given with correlation taken into account:

$$u_{\mathrm{rel}}(\overline{E}_{i,j}) = \sqrt{u^2_{\mathrm{rel}}(\overline{E}^{\mathrm{S}}_{i}) + \frac{1}{M^2}\sum_{r=1}^{M} u^2_{\mathrm{rel}}(E^{\mathrm{T}}_{i,j,r})} \tag{3}$$

2. For each NMI *i* for each lamp *j*, the relative difference $\Delta_{i,j}$ between NMI measurement (as an average of two rounds) and Pilot measurement is given by,

$$\Delta_{i,j} = \frac{\overline{E}_{i,j}}{E^{\mathrm{P}}_{i,j}} - 1 \tag{4}$$

and its uncertainty by

$$u(\Delta_{i,j}) = \sqrt{u_{\mathrm{rel}}{}^2(\overline{E}_{i,j}) + u_{\mathrm{rel}}{}^2(E^{\mathrm{PR}}_{i,j}) + u_{\mathrm{rel,add}}{}^2(E_{i,j})}\,. \tag{5}$$

where $u_{\mathrm{rel,add}}(E_{i,j})$ is an additional uncertainty in the comparison of lamp *j* of NMI *i*, arising from those components such as changes of the artifact due to transportation (if identified) and different measurement conditions between Pilot and participants that affected comparison results (if applicable) – often related to characteristics of the artifacts.

Note: The term $u_{\mathrm{rel}}{}^2(E^{\mathrm{PR}}_{i,j})$ rather than $u_{\mathrm{rel}}{}^2(E^{\mathrm{P}}_{i,j})$ is used for Pilot lab uncertainty because Pilot measurements $E^{\mathrm{P}}_{i,j}$ are strongly correlated with each other, and only uncorrelated components in Pilot measurements contribute when $\Delta_{i,j}$ are further reduced to calculate DoE.

3. For each NMI *i*, the relative differences Δ_i(average of the three lamps) is obtained by

$$\Delta_i = \frac{1}{3}\sum_{j=1}^{3}\Delta_{i,j} \tag{6}$$

and its uncertainty by

$$u(\Delta_i) = \frac{1}{3}\sum_{j=1}^{3}u(\Delta_{i,j}). \tag{7}$$

Note: This uncertainty calculation is an approximation, assuming that the results from the three lamps measured by the same NMI are nearly fully correlated.

For Pilot lab (*i* = 0 is used hereinafter),

$$\Delta_0 = 0 \quad \text{and} \quad u(\Delta_0) = u_{\mathrm{rel}}(\overline{E}^{\mathrm{P}}) \tag{8}$$

where $u_{\mathrm{rel}}(\overline{E}^{\mathrm{P}})$ is the average total uncertainty of all measurements at Pilot lab:

$$u_{\mathrm{rel}}(\overline{E}^{\mathrm{P}}) = \frac{1}{3N}\sum_{i=1}^{N}\sum_{j=1}^{3}u_{\mathrm{rel}}(E^{\mathrm{P}}_{i,j}) \tag{9}$$

4. The relative uncertainty of measurements of NMI *i*, averaged for all lamps, is determined by

$$u_{\mathrm{rel}}(\overline{E}_i) = \frac{1}{3}\sum_{l=1}^{3}u_{\mathrm{rel}}(\overline{E}_{i,j}) \tag{10}$$

For convenience of calculation hereinafter,

$$u_{\mathrm{rel}}(\overline{E}_0) = u_{\mathrm{rel}}(\overline{E}^{\mathrm{P}}) \tag{11}$$

5. The KCRV is calculated using weighted mean with cut-off. The cut-off value $u_{\text{cut-off}}$ is calculated by

$$u_{\text{cut-off}} = \text{average}\{u_{\text{rel}}(\overline{E}_i)\} \quad \text{for} \quad u_{\text{rel}}(\overline{E}_i) \le \text{median}\{u_{\text{rel}}(\overline{E}_i)\} \; ; i = 0 \text{ to } N \tag{12}$$

The reported uncertainty $u_{\text{rel}}(\overline{E}_i)$ of each NMI i is adjusted by the cut-off,

$$u_{\text{rel,adj}}(\overline{E}_i) = u_{\text{rel}}(\overline{E}_i) \quad \text{for} \quad u_{\text{rel}}(\overline{E}_i) \ge u_{\text{cut-off}}$$
$$u_{\text{rel,adj}}(\overline{E}_i) = u_{\text{cut-off}} \quad \text{for} \quad u_{\text{rel}}(\overline{E}_i) < u_{\text{cut-off}} \qquad i = 0 \text{ to N} \tag{13}$$

The transfer uncertainty component in $u(\Delta_i)$ is separated by

$$u_{\text{T}}(\Delta_i) = \sqrt{u^2(\Delta_i) - u^2_{\text{rel}}(\overline{E}_i)} \tag{14}$$

The uncertainty of Δ_i after cut-off is given by

$$u_{\text{adj}}(\Delta_i) = \sqrt{u^2_{\text{rel,adj}}(\overline{E}_i) + u^2_{\text{T}}(\Delta_i)} \tag{15}$$

The weights w_i for NMI i is determined by

$$w_i = u_{\text{adj}}^{-2}(\Delta_i) / \sum_{i=0}^{N} u_{\text{adj}}^{-2}(\Delta_i) \tag{16}$$

The KCRV, Δ_{KCRV}, is determined by

$$\Delta_{\text{KCRV}} = \sum_{i=0}^{N} w_i \Delta_i \tag{17}$$

The uncertainty of the KCRV (weighted mean with cut-off) is given by

$$u(\Delta_{\text{KCRV}}) = \sqrt{\sum_{i=0}^{N} \frac{u^2(\Delta_i)}{u_{\text{adj}}^4(\Delta_i)}} \Bigg/ \sum_{i=0}^{N} u_{\text{adj}}^{-2}(\Delta_i) \tag{18}$$

6. The Birge ratio R_{B} is calculated for consistency check. i=0 represents the pilot lab.

$$R_{\text{B}} = \sqrt{\frac{1}{N} \sum_{i=0}^{N} \frac{\{\Delta_i - \Delta_{\text{KCRV}}\}^2}{u_{\text{adj}}^2(\Delta_i)}} \tag{19}$$

If $R_{\text{B}} \le 1$, consistency is satisfied. Move to step 7.

If $R_{\text{B}} > 1$, consistency fails. In this case, following the Mandel-Paule method, add the s^2 term in eq. (15) as

$$u_{\text{adj}}(\Delta_i) = \sqrt{u^2_{\text{rel,adj}}(\overline{E}_i) + u^2_{\text{T}}(\Delta_i) + s^2} \tag{20}$$

and recalculate eqs. (16) to (19). The value of s can be determined by iterative process so that R_{B} in eq. (19) equals 1.

7. The unilateral DoE of NMI *i* is given by

$$D_i = \Delta_i - \Delta_{\mathrm{KCRV}} \tag{21}$$

$$U_i = k\sqrt{u^2(\Delta_i) + u^2(\Delta_{\mathrm{KCRV}}) - 2\left(\frac{u^2(\Delta_i)}{u_{\mathrm{adj}}^{\ 2}(\Delta_i)} \Big/ \sum_{j=0}^{N} u_{\mathrm{adj}}^{-2}(\Delta_j)\right)} \quad ; k=2 \tag{22}$$

Note: Eq.(22) takes into account the effect of correlation between Δ_i and Δ_{KCRV}. For any labs that are excluded from KCRV calculation, a simpler form applies:

$$U_i = k\sqrt{u^2(\Delta_i) + u^2(\Delta_{\mathrm{KCRV}})} \tag{23}$$

8. The bilateral DoE between NMI *i* and NMI *m* is given by

$$D_{i,m} = \Delta_i - \Delta_m \tag{24}$$

$$U_{i,m} = k\sqrt{u^2(\Delta_i) + u^2(\Delta_m)} \quad ; k=2 \tag{25}$$

In this equation it is assumed that Δ_i and Δ_m are not correlated.