European Commission

JRC SCIENTIFIC AND POLICY REPORTS

Determination of a reference value associated standard uncertainty and degrees of equivalence

for CCRI(II)
key comparison data

Stefaan Pommé

2012

Report EUR 25355 EN

Joint
Research
Centre

Determination of a reference value, associated standard uncertainty and degrees of equivalence for CCRI(II) key comparison data

Stefaan Pommé

Institute for Reference Materials and Measurements, 2440 Geel, Belgium
stefaan.pomme@ec.europa.eu

April 2012

Executive summary

CCRI(II) key comparison data consist of a measured value of activity concentration, independently obtained, and the associated standard uncertainty for each laboratory participating in a key comparison. A method is proposed for calculating a key comparison reference value (KCRV), its associated standard uncertainty, and degrees of equivalence (DoE) for the laboratories.

The method allows for technical scrutiny of data, correction or exclusion of extreme data, but above all uses an estimator (power-moderated mean, PMM) that can calculate an efficient and robust mean from any data set. For mutually consistent data, the method approaches a weighted mean, the weights being the reciprocals of the variances (squared standard uncertainties) associated with the measured values. For data sets suspected of inconsistency, the weighting is moderated by augmenting the laboratory variances by a common amount and/or decreasing the power of the weighting factors.

The PMM has the property that for increasingly discrepant data sets there is a smooth transition of the KCRV from the weighted mean to the arithmetic mean. It is a good compromise between efficiency and robustness, while providing also a reliable uncertainty.

Before applying the method, the data provided by the key comparison participants should be scrutinised to see whether any appear to be discrepant. Extreme data may also be identified subsequently by the application of a suitable statistical criterion. Such data should be considered for exclusion from the calculation of the KCRV on relevant technical grounds.

Then the KCRV, its associated uncertainty and DoEs can meaningfully be obtained. DoEs are calculated using the uncertainties provided by the key comparison participants, not the augmented uncertainties.

1 Introduction

The CCRI(II) organises key comparisons in which each of N participating laboratories independently provides a measured value of an activity concentration x_i and an associated standard uncertainty u_i. Until now in the CCRI(II), the uncertainties u_i have generally been disregarded for the calculation of a KCRV, the KCRV being calculated as an arithmetic (unweighted) mean.

The CCRI(II) is now considering calculating a KCRV using a method, such as the weighted mean, that accounts for the u_i. However, the CCRI(II) takes into consideration that these uncertainty values are generally imperfect estimates of the combined effect of all sources of variability, and therefore also prone to error.

A method is proposed for calculating a key comparison reference value (KCRV), its associated standard uncertainty, and degrees of equivalence (DoEs) for the laboratories. The method is based on a few fundamental principles:

- The estimator should be efficient, providing an accurate KCRV on the basis of the available data set (x_i, u_i) and technical scrutiny.

- The estimator should give a realistic standard uncertainty on the KCRV.

- The N data are treated on an equal footing, albeit that relative weighting may vary as a function of stated uncertainty. The method shall optimise the use of information contained in the data.

- Before evaluation, all data is scrutinised in an initial data screening by (representatives of) the CCRI(II), which may choose to exclude or correct data on technical grounds from the calculation of the KCRV.

- Extreme data can be excluded from the calculation of the KCRV on statistical grounds as part of the method. The CCRI(II) is always the final arbiter regarding excluding any data from the calculation of the KCRV.

- The estimator should be robust against extreme data, in case such data have not been excluded from the data set. It should also adequately cope with discrepant data sets.

- The method is preferably not complex and conveniently reproducible.

The estimator of choice is the *power-moderate weighted mean* (PMM), an upgrade of the well-established *Mandel-Paule mean* [1-2], incorporating ideas by *Pommé-Spasova* [3]. Its results are generally intermediate between arithmetic and weighted mean. Annex A gives a resume of the *rationale* behind the choice of estimator on the basis of conclusions drawn from simulations. Annex B contains an overview of relevant formulae for the *arithmetic mean*, the classical *weighted mean*, the *Mandel-Paule mean* and the *PMM*. In Annex C, formulae are derived for degree of equivalence and outlier identification.

Section 2 discusses the use of the PMM estimator for the KCRV and its uncertainty. Section 3 considers the identification and treatment of data regarded as statistically extreme. Section 4 describes the determination of degrees of equivalence. Section 5 summarises the proposed method. Section 6 gives conclusions.

2   The PMM estimator

The *PMM estimator* combines aspects of the arithmetic mean, the weighted mean and the Mandel-Paule mean. The logical steps leading to this procedure can be read in Annex B.

In this section, the mathematical steps are shown in order of execution:

---

1   Calculate the *Mandel-Paule mean*

$$x_{\mathrm{mp}} = \sum_{i=1}^{N} \frac{x_i}{u_i^2 + s^2} \bigg/ \sum_{i=1}^{N} \frac{1}{u_i^2 + s^2} \tag{1}$$

using $s^2=0$ as initial value, conform to the weighted mean.

---

2   Calculate the *modified reduced observed chi-squared value*

$$\widetilde{\chi}_{\mathrm{obs}}^2 = \frac{1}{N-1} \sum_{i=1}^{N} \frac{(x_i - x_{\mathrm{mp}})^2}{u_i^2 + s^2} \tag{2}$$

---

3   If $\widetilde{\chi}_{\mathrm{obs}}^2 > 1$, increase the variance s^2 and repeat steps 1-2 until $\widetilde{\chi}_{\mathrm{obs}}^2 = 1$ is obtained.

---

4   Assess the reliability of the uncertainties provided. Choose a value for the *power α* of the uncertainties in the weighting factors:

power	reliability of uncertainties
$\alpha = 0$	uninformative uncertainties (arithmetic mean)
$\alpha = 0$	uncertainty variation due to error at least twice the variation due to metrological reasons (arithmetic mean)
$\alpha = 2\text{-}3/N$	informative uncertainties with a tendency of being rather underestimated than overestimated (intermediately weighted mean)
$\alpha = 2$	informative uncertainties with a modest error; no specific trend of underestimation (Mandel-Paule mean)
$\alpha = 2$	accurately known uncertainties, consistent data (weighted mean)

---

5   Calculate a *characteristic uncertainty per datum*, based on the variance associated with the arithmetic mean, $\bar{x}$, or the Mandel-Paule mean, x_{mp}, whichever is larger[1].

---

[1] Both variances are equal when $\widetilde{\chi}_{\mathrm{obs}}^2 = 1$.

$$S = \sqrt{N \cdot \max(u^2(\bar{x}), u^2(x_{mp}))} \tag{3}$$

in which

$$u^2(\bar{x}) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N(N-1)}, \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad u^2(x_{mp}) = \left(\sum_{i=1}^{N} \frac{1}{u_i^2 + s^2} \right)^{-1} \tag{4}$$

6   Calculate the reference value and uncertainty from a *power-moderated weighted mean*

$$x_{\text{ref}} = \sum_{i=1}^{N} w_i x_i \qquad \frac{1}{u^2(x_{\text{ref}})} = \sum_{i=1}^{N} \left[\left(\sqrt{u_i^2 + s^2} \right)^\alpha S^{2-\alpha} \right]^{-1} \tag{5}$$

in which the normalised weighting factor is

$$w_i = u^2(x_{\text{ref}}) \left[\left(\sqrt{u_i^2 + s^2} \right)^\alpha S^{2-\alpha} \right]^{-1} \tag{6}$$

3   Treatment of extreme data

Statistical tools may be used to indicate data that are extreme. An extreme datum is such that the magnitude of the difference e_i between a measured value x_i and a candidate KCRV x_{ref} exceeds a multiple of the standard uncertainty $u(e_i)$ associated with e_i:

$$|e_i| > k u(e_i), \quad e_i = x_i - x_{\text{ref}}, \tag{7}$$

where k is a coverage factor, typically between two and four, corresponding to a specified level of confidence.

Irrespective of the type of mean, the variance of the difference is conveniently calculated from the modified uncertainties through the normalised weighting factors (see Annex C):

$$u^2(e_i) = u^2(x_{\text{ref}})(\frac{1}{w_i} - 1) \qquad (x_i \text{ included in mean}) \tag{8}$$

$$u^2(e_i) = u^2(x_{\text{ref}})(\frac{1}{w_i} + 1) \qquad (x_i \text{ excluded from mean}) \tag{9}$$

Preferably, the identification and rejection of extreme data is kept to a minimum, so that the mean is based on a large subset of the available data. A default coverage factor of $k=2.5$ is recommended.

After exclusion of any data, a new candidate KCRV and its associated uncertainty are calculated, and on the basis of test (7) possibly further extreme values are identified. The process is repeated until there are no further extreme values to be excluded. The CCRI(II) is always the final arbiter regarding excluding any data from the calculation of the KCRV.

In this way, the KCRV can be protected against extreme values that are asymmetrically disposed with respect to the KCRV, and the standard uncertainty associated with the KCRV is reduced. The approach of using the modified uncertainties limits the number of values for which the inequality in expression (7) holds.

4 Degrees of equivalence

The degrees of equivalence, DoE, for the ith laboratory has two components $(d_i, U(d_i))$, where, assuming normality,

$$d_i = x_i - x_{\text{ref}}, \quad U(d_i) = 2u(d_i). \tag{9}$$

$u(d_i)$ is the standard uncertainty associated with the value component d_i, and $U(d_i)$, the uncertainty component, is the expanded uncertainty at the 95 % level of confidence.

Given a KCRV x_{ref} and the associated standard uncertainty $u(x_{\text{ref}})$ obtained from expressions (5-6), the corresponding DoEs are determined from the generally valid expression for any kind of weighted mean[2] (see Annex C):

$$u^2(d_i) = (1 - 2w_i)u_i^2 + u^2(x_{\text{ref}}). \tag{10}$$

The DoEs for participants whose data were excluded from the calculation of the KCRV are given by essentially the same expression, applying $w_i = 0$:

$$u^2(d_i) = u_i^2 + u^2(x_{\text{ref}}). \tag{11}$$

The variance u_i^2 associated with x_i is *not* augmented by s^2 for the calculation of the DoE, since it is the measurement capability of laboratory i, including proper uncertainty statement, that is being assessed.

5 Summary of the method

1. Carry out a careful examination of the participants' data. If necessary, correct or exclude erroneous data on technical grounds.

2. Form the weighted mean and the associated standard uncertainty of the remaining data, (x_i, u_i), $i = 1, \ldots, N$, using Eq. (1) with $s^2 = 0$.

3. Test for consistency of the data with the weighted mean by calculating $\widetilde{\chi}^2_{\text{obs}}$ using Eq. (2), regarding the data as consistent if $\widetilde{\chi}^2_{\text{obs}} \leq 1$.

4. If $\widetilde{\chi}^2_{\text{obs}} > 1$, calculate the Mandel-Paule mean of the remaining data. That is, the variance s^2 in the weighted mean (Eq. 1) is chosen such that $\widetilde{\chi}^2_{\text{obs}}$ (Eq. 2) is unity.

5. Choose a value for the power α based on the reliability of the uncertainties and the sample size.

6. Calculate the PMM and its standard uncertainty from Eqs. (3-6).

7. Use the statistical criterion in Eq. (7) to identify any further extreme values, applying the normalised weighting factors (Eqs. 8-9). Should the CCRI(II) exclude such data from the calculation of the KCRV, repeat steps 2-6 on the remaining data set.

8. Take the PMM as the KCRV and its associated standard uncertainty as the standard uncertainty associated with the KCRV.

9. Form the DoEs for all participating laboratories (Eqs. 10-11).

---

[2] In the particular case of the classical weighted mean, the expression for $u^2(d_i)$ is identical to $u_i^2 - u^2(x_{\text{ref}})$, because $w_i = u^2(x_{\text{ref}})/u_i^2$. The more general expression has to be used for the Mandel-Paule and PMM ($s^2 > 0$ or $\alpha < 2$).

6 Conclusions

The method proposed here for calculating a KCRV and its uncertainty is based on a weighted mean, in which the relative weighting factors are adjusted to the level of consistency in the data set. It applies when the measured values provided by the participants in the key comparison are mutually independent.

The method foresees in protection against erroneous and extreme data through the possibility for correction or exclusion of data. Further discrepancy of the data, most typically caused by understatement of the uncertainty, is generally well taken into account by the estimator. This is established by augmenting the uncertainties and reducing the power of uncertainties in the weighting factors. This is done purely for the calculation of the KCRV, as the laboratory data remain unaltered when obtaining degrees of equivalence.

For consistent data with correctly determined uncertainties, the KCRV approaches the classical weighted mean. For highly discrepant data with uninformative uncertainties, the KCRV approaches the arithmetic mean. There is a smooth transition from the weighted mean to the arithmetic mean as the degree of data inconsistency increases. For CCRI(II) intercomparison results, typically slightly discrepant data with informative but imperfectly evaluated uncertainties, the KCRV is intermediate between the Mandel-Paule mean and the arithmetic mean.

The resulting KCRV should in general combine good efficiency and robustness properties. The approach will not perform well if a majority of the data values have significant bias of the same sign. The associated uncertainty will in most cases be realistic. The method remains vulnerable to mainly small, seemingly consistent data sets with systematically understated (or overstated) uncertainties.

References

1. J. Mandel, R.C. Paule, Interlaboratory evaluation of material with unequal number of replicates. Anal Chem 42 (1970) 1194–1197

2. R.C. Paule, J. Mandel, Consensus values and weighting factors. J Res Nat Bur Std 87 (1982) 377–385

3. S. Pommé, Y. Spasova, A practical procedure for assigning a reference value and uncertainty in the frame of an interlaboratory comparison. Accred Qual Assur 13 (2008) 83–89

Annex A. Rationale behind the choice of estimator

Measurement results show deviations from the "true" value of the measurand. The same is also true for the reported uncertainty, which in general is only a rough estimate of the combined effect of all sources of variability.

Computer simulations were performed to evaluate the performance of estimators of the mean of data sets, in particular of discrepant data sets for which the variation of the data x_i exceeds expectation from the stated uncertainties u_i. Important criteria were *efficiency*, a measure for the accuracy by which the true value was approached, *robustness* against extreme data, and *reliability*, a measure for the accuracy of the uncertainty value provided.

Some conclusions were as follows:

1. If none of the u_i is informative, the *arithmetic mean* is the most efficient. In practice, the arithmetic mean is a good choice for data with poorly known uncertainty, i.e. if the variation of uncertainties u_i due to error in the uncertainty assessment is twice or more the variation due to metrological origin.

2. If the uncertainties u_i are informative, an estimator that uses them can be employed to improve the efficiency. Some approaches using the u_i are better than others.

3. The classical *weighted mean*, using the reciprocal of the variance as weighting factor, is the most efficient estimator for normally distributed data, only in absence of unrepresentative data. Even modest contamination by such data, in particular those having extreme values and/or understated uncertainties, results in a too low uncertainty estimate.

4. The *Mandel-Paule mean* provides a good combination of efficiency and robustness for discrepant data that are approximately symmetrically disposed with respect to the KCRV. It degrades little with increasing contamination, is only slightly dependent on the level of information carried by the u_i, and is superior to the classical weighted mean when the u_i are not very informative.

5. The *PMM* yields more reliable uncertainties for discrepant data sets in which uncertainties are likely to be underestimated. It uses moderate weighting for small data sets, thus being less influenced by unidentified outliers with underestimated uncertainties.

6. In the presence of extreme values, the *M-P and PMM estimators* approach the result of the arithmetic mean. They can easily be complemented with an outlier rejection mechanism, which improves their efficiency. For a consistent data set, the *M-P mean and PMM* approach the classical weighted mean.

Annex B. Formulae for KCRV estimators and associated standard uncertainties

1.   Arithmetic mean

The arithmetic mean is calculated from

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad\qquad (B.1)$$

and its uncertainty, applying the propagation rule, is

$$u(\bar{x}) = \frac{1}{N}\sqrt{\sum_{i=1}^{N} u_i^2} \ . \qquad\qquad (B.2)$$

As the arithmetic mean is of particular interest when the u_i are not informative, one can replace the individual variances u_i^2 by an estimate of the sample variance

$$u^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x}) \qquad\qquad (B.3)$$

resulting in

$$u(\bar{x}) = \frac{1}{N}\sqrt{Nu^2} = \frac{u}{\sqrt{N}} = \left[\sum_{i=1}^{N}\frac{(x_i - \bar{x})^2}{N(N-1)}\right]^{1/2} \qquad\qquad (B.4)$$

As the dispersion of data is determined by chance, the calculated uncertainty of the mean can sometimes be unrealistically low, in particular with small data sets showing almost no scatter. If the u_i are informative with respect to the uncertainty scale, one could take the maximum value from both approaches:

$$u(\bar{x}) = \max\left(\frac{1}{N}\sqrt{\sum_{i=1}^{N} u_i^2} , \sqrt{\sum_{i=1}^{N}\frac{(x_i - \bar{x})^2}{N(N-1)}}\right) \qquad\qquad (B.5)$$

2.   Weighted mean

The classical weighted mean uses the reciprocal variances as weighting factor. The weighted mean x_{ref} of the data set and the associated standard uncertainty $u(x_{\mathrm{ref}})$ are obtained from

$$x_{\mathrm{ref}} = \sum_{i=1}^{N}\frac{x_i}{u_i^2} \Big/ \sum_{i=1}^{N}\frac{1}{u_i^2} \quad \text{or} \quad x_{\mathrm{ref}} = u^2(x_{\mathrm{ref}})\sum_{i=1}^{N}\frac{x_i}{u_i^2} \qquad\qquad (B.6)$$

and

$$\frac{1}{u^2(x_{\mathrm{ref}})} = \sum_{i=1}^{N}\frac{1}{u_i^2} \quad \text{or} \quad u(x_{\mathrm{ref}}) = \left[\sum_{i=1}^{N}\frac{1}{u_i^2}\right]^{-1/2} \qquad\qquad (B.7)$$

The weighted mean and its uncertainty are particularly inadequate when applied to discrepant data with understated uncertainties. One can look for indications of discrepancy by calculating the *reduced observed chi-squared value*

$$\tilde{\chi}_{\mathrm{obs}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(x_i - x_{\mathrm{ref}})^2}{u_i^2} \qquad\qquad (B.8)$$

A $\widetilde{\chi}^2_{\mathrm{obs}}$-value (significantly) higher than unity (is) may be an indication of inconsistency.[3]

3. Mandel-Paule mean

The M-P mean was designed to deal with discrepant data sets, having a *reduced observed chi-squared value* $\widetilde{\chi}^2_{\mathrm{obs}}$ larger than unity. For the purpose of establishing a more robust mean, the laboratory variances u_i^2 are incremented by a further variance s^2 to give *augmented variances* $u^2(x_i) = u_i^2 + s^2$. The value of the unexplained variance s^2 is chosen such that the *modified reduced observed chi-squared value*,

$$\widetilde{\chi}^2_{\mathrm{obs}} = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(x_i - x_{\mathrm{ref}})^2}{u_i^2 + s^2}, \tag{B.9}$$

equals one.

The calculation of the M-P mean and its uncertainty proceeds through the same equations as for the weighted mean, replacing the stated variances u_i^2 by the augmented variances $u^2(x_i)$

$$x_{\mathrm{ref}} = u^2(x_{\mathrm{ref}})\sum_{i=1}^{N}\frac{x_i}{u_i^2 + s^2}, \qquad \frac{1}{u^2(x_{\mathrm{ref}})} = \sum_{i=1}^{N}\frac{1}{u_i^2 + s^2}. \tag{B.10}$$

As the M-P mean x_{ref} occurs in the equation for $\widetilde{\chi}^2_{\mathrm{obs}}$, an iterative procedure is applied to find the appropriate value of the variance s^2.

For data sets with $\widetilde{\chi}^2_{\mathrm{obs}}$ smaller than 1, the variances are not augmented, $s^2 = 0$, and the result is identical to the weighted mean. For an extremely inconsistent set, s^2 will be large compared with the u_i^2 and the Mandel-Paule mean will approach the arithmetic mean. For intermediate cases, the influence of those laboratories that provide the smallest uncertainties will be reduced and the standard uncertainty associated with the KCRV will be larger compared with that for the weighted mean.

Though much more robust than the weighted mean, the M-P procedure tends to underestimate its uncertainty for data sets with predominantly understated uncertainties.

4. The PMM

The M-P does not counteract possible errors in the relative uncertainties when the data set appears to be consistent, this is when $\widetilde{\chi}^2_{\mathrm{obs}}$ is not (much) larger than unity. Data with understated uncertainty have a negative effect on the robustness and the calculated uncertainty. The PMM estimator allows moderating the relative weighting also for seemingly consistent data sets.

For the M-P mean as well as the classical weighted mean, uncertainties u_i are used with a power of 2 in the weighting factor. By lowering this power, the influence of understated uncertainties can be moderated. A smooth transition from weighted to arithmetic mean can be realised by intermixing the uncertainties associated with both.

Like with the M-P mean, the variances are increased by an unexplained amount s^2 to ascertain that $\widetilde{\chi}^2_{\mathrm{obs}}$ is not larger than one. Then a variance per datum is calculated for an

---

[3] The mean (or expectation) of a variable having a chi-squared distribution with $N-1$ degrees of freedom is $N-1$. Under normality conditions, the expected value of $\widetilde{\chi}^2_{\mathrm{obs}}$ is unity.

unweighted mean, taking the larger value between the sample variance and the combined augmented uncertainties:

$$S^2 = N \max\left[\left(\sum_{i=1}^{N} \frac{1}{u_i^2 + s^2}\right)^{-1}, \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N(N-1)}\right] \tag{B.11}$$

In the expressions (B.10), the weighting factor $1/(u_i^2 + s^2)$ is replaced by

$$\left[\left(\sqrt{u_i^2 + s^2}\right)^\alpha S^{2-\alpha}\right]^{-1}, \tag{B.12}$$

in which the power α ($0 \le \alpha \le 2$) is the leverage by which the mean can be smoothly varied between arithmetic mean ($\alpha=0$) and M-P mean ($\alpha=2$). The *Mandel-Paule method* can be regarded as a subset of the *PMM method*. Reducing α has a similar effect to the KCRV and its uncertainty as does augmenting the laboratory variances in the M-P method.

The choice of α can be made to reflect the level of trust in the stated uncertainties. For data sets with a predominance of understated uncertainties, one obtains a more realistic uncertainty on the KCRV by reducing the power α. This is particularly recommended with small data sets. Larger data sets have a better defined $\tilde{\chi}^2_{obs}$, facilitating the identification of extreme data and the level of reliability of the u_i. As a practical rule, for data sets in which the uncertainties u_i are informative but frequently understated, one can make the power α depend on the number of data N via a heuristic formula

$$\alpha = 2 - \frac{3}{N} \tag{B.13}$$

Table 1. Estimators of mean

Estimator $x_{\text{ref}} = \sum_{i=1}^{N} w_i x_i$	Normalised weight w_i	Mean x_{ref}	Associated standard uncertainty $u(x_{\text{ref}})$
Arithmetic mean	$\dfrac{1}{N}$	$\dfrac{1}{N}\sum_{i=1}^{N} x_i$	$\left[\sum_{i=1}^{N}\dfrac{(x_i - \bar{x})^2}{N(N-1)}\right]^{1/2}$
Weighted mean	$\dfrac{u^2(x_{\text{ref}})}{u_i^2}$	$u^2(x_{\text{ref}})\sum_{i=1}^{N}\dfrac{x_i}{u_i^2}$	$\left[\sum_{i=1}^{N}\dfrac{1}{u_i^2}\right]^{-1/2}$
Mandel-Paule mean	$\dfrac{u^2(x_{\text{ref}})}{u_i^2 + s^2}$	$u^2(x_{\text{ref}})\sum_{i=1}^{N}\dfrac{x_i}{u_i^2 + s^2}$	$\left[\sum_{i=1}^{N}\dfrac{1}{u_i^2 + s^2}\right]^{-1/2}$
PMM	$\dfrac{u^2(x_{\text{ref}})}{\left(\sqrt{u_i^2 + s^2}\right)^\alpha S^{2-\alpha}}$	$u^2(x_{\text{ref}})\sum_{i=1}^{N}\dfrac{x_i}{\left(\sqrt{u_i^2 + s^2}\right)^\alpha S^{2-\alpha}}$	$\left[\sum_{i=1}^{N}\left(\sqrt{u_i^2 + s^2}\right)^{-\alpha} S^{\alpha-2}\right]^{-1/2}$

Annex C. Formulae for 'degree of equivalence' and 'outlier identifcation'

1. Degrees of equivalence

The degrees of equivalence *between pairs of NMIs* are not influenced by the estimator used for the KCRV. The degree of equivalence of laboratory data *with respect to the KCRV* involves calculation of the difference

$$d_i = x_i - x_{ref}$$

$$= x_i - \sum_{j=1}^{N} w_j x_j \tag{C.1}$$

and its expanded uncertainty. In the expression C.1, the factor w_i is:

  - the normalised weighting factor (see Table 1) for data included in the mean

  - zero for data excluded from the calculation of the KCRV

Data excluded from the calculation of the mean are not correlated with it and the variance associated with d_i is, in this case, the sum of two variances:

$$u^2(d_i) = u_i^2 + u^2(x_{ref}) \qquad (x_i \text{ excluded from mean}) \tag{C.2}$$

The data that have been included in the calculation of the mean are correlated with it, and the variance of the difference is calculated from

$$u^2(d_i) = (1 - w_i)u_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^{N} w_j u_j^2 \qquad (x_i \text{ included in mean})$$

$$= (1 - 2w_i)u_i^2 + u^2(x_{ref}) \qquad (x_i \text{ included in mean}) \tag{C.3}$$

2. Outlier identification

For a consistent data set with reliable standard uncertainties, one could apply a weighted mean and use the reciprocal of the variances as weighting factors. Data not complying with this consistent set can be recognised if their difference e_i from the mean exceeds the associated uncertainty by a factor k or more.

$$u^2(e_i) = (1 - 2w_i)u_i^2 + u^2(x_{ref}) \qquad (x_i \text{ included in mean})$$

$$= (1 - 2w_i)\frac{u^2(x_{ref})}{w_i} + u^2(x_{ref}) \qquad (\text{weighted mean})$$

$$= u^2(x_{ref})(\frac{1}{w_i} - 1) \qquad (\text{weighted mean}) \tag{C.4}$$

A similar equation, with opposite sign, holds for data not included in the mean.

$$u^2(e_i) = u_i^2 + u^2(x_{ref}) \qquad (x_i \text{ excluded from mean})$$

$$= u^2(x_{ref})(\frac{1}{w_i} + 1) \qquad (\text{weighted mean}) \tag{C.5}$$

Typical CCRI(II) intercomparison data contain understated uncertainties, and normal criteria for outlier identification would reject many data as extreme. Preferably, the identification and rejection of extreme data is kept to a minimum, so that the mean is based

on a large subset of the available data. This is easily achieved in the philosophy of the M-P mean, even for discrepant data sets, by using the augmented uncertainties

$$u^2(e_i) = (1 - 2w_i)(u_i^2 + s^2) - u^2(x_{ref})$$ (Mandel-Paule mean)   (C.6)

Similarly, one can apply the power-moderated uncertainty for the PMM method. In all cases, the variance equations reduce to the same elegant solutions as in Eqs. (C.4-5), expressed as a function of the weighting factor. If the method reduces to an arithmetic mean (α=0), the weighting factors are equal for all data, irrespective of the stated uncertainty. Extreme data are then identified from their difference with the mean only.

Abstract

CCRI(II) key comparison data consist of a measured value of activity concentration, independently obtained, and the associated standard uncertainty for each laboratory participating in a key comparison. A method is proposed for calculating a key comparison reference value (KCRV), its associated standard uncertainty, and degrees of equivalence (DoE) for the laboratories.

The estimator has the property that for increasingly discrepant data sets there is a smooth transition of the KCRV from the weighted mean to the arithmetic mean. It is a good compromise between efficiency and robustness, while providing also a reliable uncertainty. A suitable statistical criterion is provided to identify extreme data..

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.