Metrologia 41 (2004) 231–237 PII: S0026-1394(04)78515-1

# Statistical analysis of key comparisons with linear trends

**Nien Fan Zhang**[1], **Hung-kung Liu**[1], **Nell Sedransk**[1] and **William E Strawderman**[2]

[1] Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
[2] Faculty Visitor, Statistical Engineering Division, National Institute of Standards and Technology and Department of Statistics, Rutgers University, New Brunswick, NJ, USA

Received 22 December 2003
Published 7 May 2004
Online at stacks.iop.org/Met/41/231
DOI: 10.1088/0026-1394/41/4/002

**Abstract**
A statistical analysis for key comparisons with linear trends is proposed. The approach has the advantage that it is consistent with the no-trend cases. The uncertainties for the key comparison reference value, which is time dependent in this case, and the degrees of equivalence are also provided. As an example, the approach is applied to key comparison CCEM-K2.

## 1. Introduction

With the recent signing of the mutual recognition arrangement (MRA), national metrology institutes (NMIs) and regional metrology organizations (RMOs) around the world have committed themselves to establishing the equivalence of their measurement standards through key comparisons of national measurement standards. In most key comparisons, the pilot NMI organizes the circulation and transport of standards or artefacts. See appendix F of the 'Mutual Recognition of National Measurements Standards and of Measurement Certificates Issued by National Metrology Institutes' [1]. In some key comparisons, the measurements of the transport standards made by the participating NMIs will show a trend or a drift (a linear drift in this paper). If this occurs, a traditional statistical analysis, which assumes no trend exists, as in the recent publications (see, e.g., Cox [2]), is inappropriate. Within some consultative committees (CCs) of the International Committee for Weights and Measures (CIPM), some uncertainty analyses are performed to accommodate the drift effect. One approach is for each NMI to calculate the difference between the measured value and the predicted value based on the trend, e.g. a linear trend. By combining these differences, e.g. using a weighted mean, the key comparison reference value (KCRV) is then calculated. However, in the process of calculating the uncertainties relating to the KCRV and other quantities, the correlations among the predicted values of the non-pilot NMIs, which are jointly based on the linear regression from the measured values of the pilot NMI, were not accounted for, and thus a more rigorous statistical study was needed. In the paper by Zhang *et al* [3], an approach was proposed and applied to the CCEM-K2 Key Comparison of Resistance Standards at 10 MΩ and 1 GΩ (see Dziuba and Jarrett [4]). The proposed method in the above article improved the uncertainty analysis by considering the correlations (which were ignored in many practices) due to the predictions for non-pilot laboratories jointly based on the pilot laboratory measurements. However, due to the definition of the KCRV in the existing methods dealing with trends, the results cannot be applied to the trivial case when the trend reduces to zero. In this paper, we define the time-dependent KCRV in an appropriate way and propose a new approach that overcomes the shortcomings in [3, 4].

## 2. Models

We assume that the measurements of any particular laboratory have a linear trend in time and the slopes of the linear trends for all the laboratories are the same, while we allow for different intercepts for different laboratories. In this paper, we consider the case of a single artefact circulating through all laboratories. The case of multiple circulating artefacts will be considered in a forthcoming paper. Without loss of generality, we assume that the pilot laboratory is the first one among all $I$ laboratories. Denote the time and the result of the $k$th measurement by the pilot laboratory by $t_{1k}$, $k = 1, \ldots, K$ (which are taken to have negligible associated uncertainty), and $X_{1k}$, $k = 1, \ldots, K$, with $K > 2$, respectively. In practice, $t_{1k}$ can be an average value of the time when the measurements were made in the $k$th period, and then $X_{1k}$ is the average value of the corresponding measurements in that period. We assume that for each non-pilot laboratory, say the $i$th laboratory ($i \neq 1$), a measurement of the travelling standard can be expressed as

$$X_i = X_{i,\mathrm{A}} + X_{i,\mathrm{B}}. \tag{1}$$

Here the components $X_{i,\mathrm{A}}$ and $X_{i,\mathrm{B}}$ are statistically independent from each other, and their corresponding standard uncertainties are $\sigma_{i,\mathrm{A}}$ and $u_{i,\mathrm{B}}$, which are the Type A and Type B evaluations of uncertainty of the measurements made by the $i$th laboratory, respectively. The notation $\sigma_{i,\mathrm{A}}$ also means the standard deviation of $X_{i,\mathrm{A}}$. By the same token, we do not distinguish the notations of a variance and the square of a standard uncertainty in this paper. Obviously, $u_i$, the standard uncertainty of the measurements made by the $i$th laboratory, has the property

$$u_i^2 = \sigma_{i,\mathrm{A}}^2 + u_{i,\mathrm{B}}^2 \tag{2}$$

for $i \neq 1$. For the pilot laboratory, each of the $K$ measurements can be expressed as

$$X_{1k} = X_{1k,\mathrm{A}} + X_{1,\mathrm{B}} \tag{3}$$

for $k = 1, \ldots, K$. We assume that $X_{11,\mathrm{A}}, X_{12,\mathrm{A}}, \ldots, X_{1K,\mathrm{A}}$ are statistically independent. The average of $X_{1k}$ is denoted by $X_1$. Specifically,

$$X_1 = \frac{\sum_{k=1}^{K} X_{1k}}{K}. \tag{4}$$

The corresponding uncertainty for $X_1$, $u_1$, is given by

$$u_1^2 = \frac{\sigma_{1,\mathrm{A}}^2}{K} + u_{1,\mathrm{B}}^2, \tag{5}$$

based on the uncertainties of the pilot laboratory. We assume that a simple linear regression model holds for the measurements made by the pilot laboratory,

$$X_{1k} = \alpha_1 + \beta t_{1k} + \varepsilon_{1k} \tag{6}$$

for $k = 1, \ldots, K$. We assume that the random error, $\varepsilon_{1k}$, has a zero mean and uncertainty of $u_1$. When $i \neq 1$, each laboratory takes one measurement at time $t_i$ and the corresponding model is

$$X_i = \alpha_i + \beta t_i + \varepsilon_i, \tag{7}$$

where the random error, $\varepsilon_i$, has a zero mean and standard uncertainty of $u_i$ for $i = 2, \ldots, I$. In a key comparison study, NMIs will provide $x_i$s, i.e. the values of $X_i$s, and either $u_i$ or $\sigma_{i,\mathrm{A}}$, and $u_{i,\mathrm{B}}$. In the case of a single circulating artefact as discussed in this paper, the separate uncertainties, $\sigma_{1,\mathrm{A}}$ and $u_{1,\mathrm{B}}$, for the pilot laboratory are needed. $\sigma_{1,\mathrm{A}}$ is usually estimated based on $X_{1,k}$, $k = 1, \ldots, K$. For a non-pilot laboratory, $u_i$ ($i > 1$) is sufficient. However, in the case of multiple circulating artefacts, separate uncertainties, $\sigma_{i,\mathrm{A}}$ and $u_{i,\mathrm{B}}$, for non-pilot laboratories are needed when a trend exists.

## 3. Parameter estimators and their variances

For the pilot laboratory, by (3) and (6), the least squares estimators of the regression parameters are given by

$$\hat{\beta} = \frac{\sum_{k=1}^{K} (t_{1k} - t_1)(X_{1k} - X_1)}{\sum_{k=1}^{K} (t_{1k} - t_1)^2} = \frac{\sum_{k=1}^{K} (t_{1k} - t_1)(X_{1k,\mathrm{A}} - X_{1,\mathrm{A}})}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}, \tag{8}$$

where $X_{1,\mathrm{A}}$ is the average of $X_{1k,\mathrm{A}}$ and

$$\hat{\alpha}_1 = X_1 - \hat{\beta} t_1 = X_{1,\mathrm{B}} + X_{1,\mathrm{A}} - \hat{\beta} t_1, \tag{9}$$

where $X_1$ is defined in (4) and $t_1$ is the average of $\{t_{11}, \ldots, t_{1K}\}$. Equation (8) can be found in Gunst and Mason [5, p 69]. For other laboratories, i.e. $i = 2, \ldots, I$, the corresponding intercepts can be estimated from (7) by

$$\hat{\alpha}_i = X_i - \hat{\beta} t_i, \tag{10}$$

where $t_i$ is the time when the $i$th ($i \neq 1$) laboratory made its measurement. By the second equality in (8),

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}. \tag{11}$$

From (8),

$$\mathrm{Cov}[X_1, \hat{\beta}] = \frac{\sum_{k=1}^{K} (t_{1k} - t_1)\,\mathrm{Cov}[X_1, (X_{1k} - X_1)]}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}.$$

By (4),

$$\mathrm{Cov}[X_1, (X_{1k} - X_1)] = \frac{\mathrm{Var}[X_{1k}]}{K} - \mathrm{Var}[X_1] = 0.$$

Thus,

$$\mathrm{Cov}[X_1, \hat{\beta}] = 0. \tag{12}$$

From (9), (5), (11), and (12), the uncertainty of $\hat{\alpha}_1$ is given by

$$u_{\hat{\alpha}_1}^2 = u_{1,\mathrm{B}}^2 + \frac{\sigma_{1,\mathrm{A}}^2}{K} + \frac{t_1^2 \sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K} (t_{1k} - t_1)^2} = u_1^2 + \frac{t_1^2 \sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}. \tag{13}$$

An unbiased estimator of $\sigma_{1,\mathrm{A}}^2$ in (13) is given by the 'mean-square residual', i.e.

$$\hat{\sigma}_{1,\mathrm{A}}^2 = \frac{\sum_{k=1}^{K} (X_{1k} - \hat{\alpha}_1 - \hat{\beta} t_{1k})^2}{K - 2}. \tag{14}$$

For the covariance of $\hat{\alpha}_i$ and $\hat{\beta}$, from (9)–(12) for $i = 1, \ldots, I$,

$$\mathrm{Cov}(\hat{\alpha}_i, \hat{\beta}) = \frac{-t_i \sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}. \tag{15}$$

Equations (11) and (14) can be found in [5, p 185]. From (10), for $i = 2, \ldots, I$,

$$u_{\hat{\alpha}_i}^2 = u_i^2 + \frac{t_i^2 \sigma_{1,\mathrm{A}}^2}{\sum_{i=1}^{K} (t_{1k} - t_1)^2}. \tag{16}$$

The uncertainty of $\hat{\alpha}_i + \hat{\beta} t$ is given by

$$u_{\hat{\alpha}_i + \hat{\beta} t}^2 = u_i^2 + \frac{(t_i - t)^2 \sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}. \tag{17}$$

We note that when $i \neq k$,

$$\mathrm{Cov}[\hat{\alpha}_i, \hat{\alpha}_k] = \frac{t_i t_k \sigma_{1,\mathrm{A}}^2}{\sum_{j=1}^{K} (t_{1j} - t_1)^2}. \tag{18}$$

## 4. Key comparison reference value

One approach to calculating the KCRV at any time $t$ (denoted by $\mathrm{KCRV}_t$) is to use a weighted mean of $\hat{\alpha}_i + \hat{\beta}t$ over the laboratories $i = 1, \ldots, I$,

$$\mathrm{KCRV}_t(w) = \sum_{i=1}^{I} w_i(\hat{\alpha}_i + \hat{\beta}t) = \sum_{i=1}^{I} w_i X_i - \hat{\beta}\sum_{i=1}^{I} w_i(t_i - t) \qquad (19)$$

with weights $w = (w_1, \ldots, w_I)$ satisfying $\sum_{i=1}^{I} w_i = 1$. $\mathrm{KCRV}_t(w)$ is a function of $t$ as well as the weights. The variance or uncertainty of $\mathrm{KCRV}_t(w)$ can be calculated using the second equality of (19).

$$\mathrm{Var}[\mathrm{KCRV}_t(w)] = \sum_{i=1}^{I} w_i^2\,\mathrm{Var}[X_i] + \left[\sum_{i=1}^{I} w_i(t_i - t)\right]^2 \mathrm{Var}(\hat{\beta})$$
$$-2\mathrm{Cov}\left[w_1 X_1, \hat{\beta}\sum_{i=1}^{I} w_i(t_i - t)\right], \qquad (20)$$

since the measurements from non-pilot laboratories, $X_i$ for $i \neq 1$ are statistically independent of $\hat{\beta}$, which are calculated based on measurements made by the pilot laboratory. Since $\mathrm{Cov}[X_1, \hat{\beta}] = 0$,

$$\mathrm{Var}[\mathrm{KCRV}_t(w)] = \sum_{i=1}^{I} w_i^2 \mathrm{Var}[X_i] + \left[\sum_{i=1}^{I} w_i(t_i - t)\right]^2 \mathrm{Var}(\hat{\beta}). \qquad (21)$$

The uncertainty is given by

$$u^2_{\mathrm{KCRV}_t(w)} = \sum_{i=1}^{I} w_i^2 u_i^2 + \frac{\sigma^2_{1,\mathrm{A}}[\sum_{i=1}^{I} w_i(t_i - t)]^2}{\sum_{k=1}^{K}(t_{1k} - t_1)^2}, \qquad (22)$$

where $u_i$ for $i = 1, \ldots, I$ are given by (2) and (5). The uncertainty given in (22) depends on $t$ as well as the weights $w$. For any given $t$, a reasonable criterion for choosing optimal weights is the minimum variance or minimum standard uncertainty of $\mathrm{KCRV}_t(w)$. The weights that give the minimal variance of $\mathrm{KCRV}_t(w)$ can be found via orthogonalization of the positive definite matrix implicit in (21). We do not go into the slightly involved details.

Equation (20) gives the uncertainty of the KCRV for a given $t$. For any given set of weights, the minimum value of the standard uncertainty of $\mathrm{KCRV}_t(w)$ will occur when the second term in (21) or $\sum_{i=1}^{I} w_i(t_i - t)^2$ equals zero. This occurs when $t$ takes the value of

$$t_{\mathrm{w}} = \sum_{i=1}^{I} w_i t_i. \qquad (23)$$

In this case, from (19), the KCRV has the same form as in the no-trend case and is given by

$$\mathrm{KCRV}_{t_{\mathrm{w}}}(w) = \sum_{i=1}^{I} w_i X_i. \qquad (24)$$

The corresponding uncertainty is given by

$$u^2_{\mathrm{KCRV}_{t_{\mathrm{w}}}(w)} = \sum_{i=1}^{I} w_i^2 u_i^2. \qquad (25)$$

As in the no-trend case, $u^2_{\mathrm{KCRV}_{t_{\mathrm{w}}}(w)}$ is minimized when the weights are given by

$$w_i(1) = \frac{1/u_i^2}{\sum_{k=1}^{I}(1/u_k^2)}. \qquad (26)$$

See Graybill and Deal [6]. The corresponding KCRV is given by

$$\mathrm{KCRV}_{t^*}(w(1)) = \sum_{i=1}^{I} w_i(1) X_i, \qquad (27)$$

where

$$t^* = \sum_{i=1}^{I} w_i(1) t_i \qquad (28)$$

and $w(1) = (w_i(1), \ldots, w_I(1))$. From (26) and (27), the uncertainty of this KCRV is the same as in the no-trend case and is given by

$$u^2_{\mathrm{KCRV}_{t^*}(w(1))} = \frac{1}{\sum_{i=1}^{I}(1/u_i^2)}, \qquad (29)$$

assuming that $u_i$ ($i = 1, \ldots, I$) in the weights $\{w_i(1)\}$ are the standard deviations for all laboratories. Otherwise, for the example when $u_i$ equals the experimental standard deviation or sample standard deviation $S_i$ for $i = 1, \ldots, I$, which are estimates of the respective standard deviations, the weights, as the estimates of $\{w_i(1),\ i = 1, \ldots, I\}$, equal

$$\frac{1/S_i^2}{\sum_{k=1}^{I}(1/S_k^2)}$$

for $i = 1, \ldots, I$. If these estimated weights are used, then (29) does not hold. The same comment applies to all the derivations in the rest of the paper when $\{w_i(1), i = 1, \ldots, I\}$ are used.

As a practical matter, we recommend the $\mathrm{KCRV}_{t^*}(w(1))$ with $t^*$, and the weights given by (26) and (28) to be the reference value of this key comparison.

## 5. Degrees of equivalence

### 5.1. Degrees of equivalence of the national measurement standards with respect to the KCRV

The degree of equivalence of the national measurement standard from the $i$th laboratory with respect to the $\mathrm{KCRV}_t$ is defined as the difference

$$D_{i,\mathrm{KCRV}(w)} = \hat{\alpha}_i + \hat{\beta}t - \mathrm{KCRV}_t(w)$$
$$= (1 - w_i)\hat{\alpha}_i + \sum_{j\neq i, j=1}^{I} w_j\hat{\alpha}_j. \qquad (30)$$

The second equality is from the first equality of (19). From the second equality of (30), $D_{i,\mathrm{KCRV}(w)}$ is a function of the weights in forming the $\mathrm{KCRV}_t(w)$, and it is independent of $t$. From (13), (16), (18), and (30), the corresponding variance is

$$\mathrm{Var}[D_{i,\mathrm{KCRV}(w)}] = (1-w_i)^2\mathrm{Var}[\hat{\alpha}_i] + \mathrm{Var}\left[\sum_{j\neq i, j=1}^{I} w_j\hat{\alpha}_j\right]$$
$$+2(1-w_i)\,\mathrm{Cov}\left[\hat{\alpha}_i, \sum_{j\neq i,j=1}^{I} w_j\hat{\alpha}_j\right]$$
$$= (1-w_i)^2\left[u_i^2 + \frac{t_i^2\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}\right]$$
$$+\sum_{j\neq i,j=1}^{I} w_j^2\left[u_j^2 + \frac{t_j^2\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}\right]$$
$$+2\sum_{j,k\neq i}\sum_{j\neq k} w_j w_k \frac{t_j t_k\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}$$
$$+2(1-w_i)\sum_{j\neq i}\frac{w_j t_j t_k \sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}. \quad (31)$$

In particular, when $w = w(1)$ as given in (26), in the first equality of (30), let $t = t^* = \sum_{i=1}^{I} w_i(1)t_i$. Then

$$D_{i,\mathrm{KCRV}(w(1))} = \hat{\alpha}_i + \hat{\beta}t^* - \mathrm{KCRV}_{t^*}(w(1))$$
$$= \hat{\alpha}_i + \hat{\beta}t^* - \sum_{i=1}^{I} w_i(1)X_i. \quad (32)$$

By (17) and (29), the corresponding uncertainty is given by

$$u^2_{D_{i,\mathrm{KCRV}(w(1))}} = [1-2w_i(1)]u_i^2 + \frac{1}{\sum_{k=1}^{I}(1/u_k^2)}$$
$$+\frac{(t_i-t^*)^2\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}. \quad (33)$$

Again we recommend using $D_{i,\mathrm{KCRV}(w(1))}$ as the degree of equivalence of the national measurement standard from the $i$th laboratory with respect to the KCRV.

### 5.2. *Degrees of equivalence of pairs of national measurement standards*

The degree of equivalence between the national measurement standards at time $t$ is defined as

$$D_{i,k} = (\hat{\alpha}_i + \hat{\beta}t) - (\hat{\alpha}_k + \hat{\beta}t) = \hat{\alpha}_i - \hat{\alpha}_k, \quad (34)$$

when $i \neq k$. Thus, the quantity is independent of $t$. The corresponding uncertainty is given by

$$u^2_{D_{i,k}} = u^2_{\hat{\alpha}_i} + u^2_{\hat{\alpha}_k} - 2\,\mathrm{Cov}[\hat{\alpha}_i, \hat{\alpha}_k]$$
$$= u^2_{\hat{\alpha}_i} + u^2_{\hat{\alpha}_k} - 2\frac{t_i t_k\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}, \quad (35)$$

where $u_{\hat{\alpha}_i}$ can be found in (13) and (16). Therefore,

$$u^2_{D_{i,k}} = u_i^2 + u_k^2 + \frac{(t_i-t_k)^2\sigma_{1,\mathrm{A}}^2}{\sum_{k=1}^{K}(t_{1k}-t_1)^2}. \quad (36)$$

## 6. An example

To illustrate the approach, we applied it to the key comparison CCEM-K2. An international comparison of dc resistance at 10 MΩ and 1 GΩ was organized under the auspices of the Consultative Committee for Electricity and Magnetism (CCEM) and piloted by the National Institute of Standards and Technology (NIST), with 14 other NMIs participating [1]. In this key comparison, three 10 MΩ wire-wound resistors and three 1 GΩ film-type resistors were used as the transport standards. During the comparison, the transport standards were measured at the pilot NMI, NIST, for seven separate periods with two different measurement systems that are used on a regular basis to calibrate customer high resistance standards [1]. For each period, an average value of the dates when the measurements were made is called a mean date of the period. Each non-pilot NMI reported a mean value and a mean date of measurement for each of the six artefacts. An uncertainty budget that includes the Type A and Type B evaluations of uncertainties for each NMI's measurement process was also reported. In this example, we only consider the 10 MΩ case and use the measurement data only for the artefact S/N HR7551 listed in table 1. The uncertainties for all NMIs are also listed in table 1.

For the trend, the assumption is that the resistor drifts in a linear fashion. Any non-linear effects are caused probably by several physical or mechanical changes during the transportation process. Although other models were considered for the drift, the consensus of the review subgroup for CCEM-K2 was that the resistor had experienced some type of non-linear behaviour during transport, that it would be difficult to model the behaviour without more information, and that a linear regression line was an acceptable drift model for this resistor. A detailed discussion can be found in section 5.1 of [4].

For the pilot NMI, within each of the seven periods (say the $k$th period, $k = 1, \ldots, 7$) when the measurements were made, the mean dates and the corresponding mean value of the measurements were taken as $t_{1k}$ and $X_{1k}$ as in section 2 and listed in table 1.

A linear regression line was fitted to the NIST measurements. Figure 1 shows the measurement results (mean values) of all NMIs corrected by the nominal value 10 MΩ in the units of μΩ/Ω and the regression line based on the NIST measurements. The estimates of the intercept for NIST, the slope, and the residual standard deviation are $\hat{\alpha}_1 = 6.03$, $\hat{\beta} = 1.05$, and $\hat{\sigma}_{1,\mathrm{A}} = 1.09$ from (8), (9), and (14), respectively. We used $\hat{\sigma}_{1,\mathrm{A}}$ as $\sigma_{1,\mathrm{A}}$ in all the formulae. The date for the minimum standard deviation of the KCRV is $t^* = 1998.23$, given by (28). The KCRV at $t^*$ with weights $w(1)$ given by (26) and (27) is 8.03, and the corresponding uncertainty is 0.28, given by (29). The $D_{i,\mathrm{KCRV}_{t^*}(w(1))}$ for all NMIs and their corresponding uncertainties, $u_{D_{i,\mathrm{KCRV}_{t^*}(w(1))}}$, are calculated from (30) and (33) and listed in table 2.

The degrees of equivalence of pairs of NMIs and their corresponding uncertainties are calculated using (34) and (36), respectively, and are listed in table 3.

**Table 1.** Information for the standard S/N HR7551.

| Laboratory | Mean date of measurement | Measurement ($\times 10^{-6}$) | Type A uncertainty ($\times 10^{-6}$) | Type B uncertainty ($\times 10^{-6}$) |
|---|---|---|---|---|
| NIST | 1996.65 | 4.6 | 0.2[a] | 1.51[a] |
| NRC | 1996.80 | 5 | 1.88 | 2.29 |
| NIST | 1996.94 | 6.7 | 0.2[a] | 1.51[a] |
| BNM-LCIE | 1997.17 | 6.97 | 0.50 | 0.35 |
| NPL | 1997.35 | 7.1 | 0.52 | 0.61 |
| PTB | 1997.50 | 7.5 | 1.0 | 2.19 |
| NIST | 1997.62 | 8.1 | 0.2[a] | 1.51[a] |
| CSIRO-NML | 1997.82 | 7.3 | 0.07 | 2.56 |
| MSL | 1998.03 | 7.3 | 0.04 | 0.59 |
| CSIR-NML | 1998.13 | −20 | 50 | 13.52 |
| NIST | 1998.33 | 8.9 | 0.2[a] | 1.51[a] |
| SP | 1998.49 | 8.7 | 0.17 | 1.79 |
| OFMET | 1998.62 | 8.9 | 0.39 | 0.58 |
| IEN | 1998.74 | 9.4 | 0.79 | 2.53 |
| NMI-VSL | 1998.98 | 9.1 | 0.80 | 3.04 |
| NIST | 1999.15 | 7.8 | 0.2[a] | 1.51[a] |
| KRIST | 1999.39 | 7.1 | 0.30 | 3 |
| NIST | 1999.60 | 8.5 | 0.2[a] | 1.51[a] |
| NIM | 1999.87 | 10.2 | 0.10 | 0.83 |
| VNIIM | 2000.03 | 10 | 0.25 | 1.03 |
| NIST | 2000.20 | 10 | 0.2[a] | 1.51[a] |

[a] Single values of Type A and Type B uncertainties were listed for the pilot NMI—NIST—for all seven measurement periods. The Type A uncertainty for the pilot NMI is computed from (14).

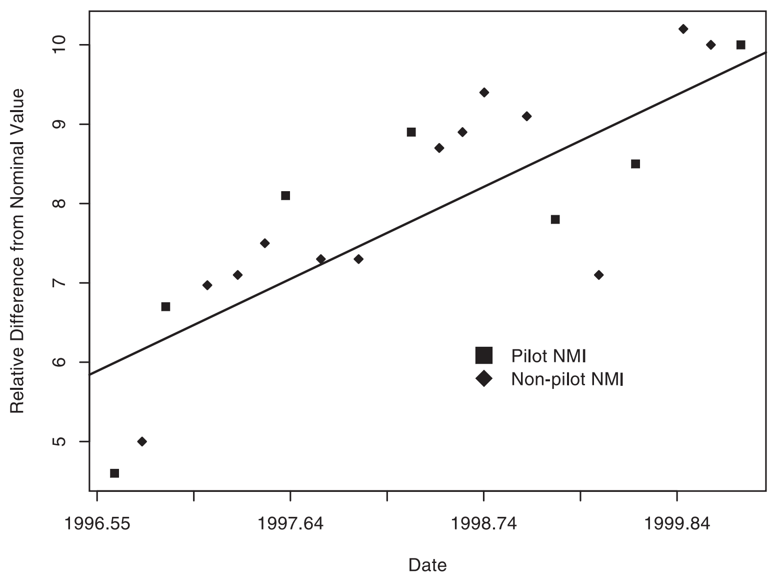


**Figure 1.** Measurements of 10 MΩ standard S/N HR7551 by all participants.

**Table 2.** The degrees of equivalence of national measurement standards with respect to KCRV and their uncertainties.

| | NIST | NRC | BNM-LCIE | NPL | PTB | CSIRO-NML | MSL | CSIR-NML | SP | OFMET | IEN | NMI-VSL | KRISS | NIM | VNIIM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $D_{i,\mathrm{KCRV}_{t^*}(w(1))}$ | −0.35 | −1.54 | 0.04 | −0.01 | 0.24 | −0.29 | −0.52 | −27.92 | 0.40 | 0.46 | 0.83 | 0.28 | −2.15 | 0.46 | 0.09 |
| $u_{D_{i,\mathrm{KCRV}_{t^*}(w(1))}}$ | 1.54 | 2.99 | 0.64 | 0.80 | 2.40 | 2.55 | 0.53 | 51.79 | 1.78 | 0.65 | 2.64 | 3.14 | 3.03 | 0.95 | 1.18 |

## 7. A note on the residual standard deviation

In the example above, we used the residual standard deviation of the regression, i.e. $\hat{\sigma}_{1,\mathrm{A}}$, from (14) to estimate the standard deviation of $X_{1k,\mathrm{A}}$ as in CCEM-K5 Comparison of 50/60 Hz Power [7]. In fact, for this CCEM-K2 key comparison study, the pilot NMI provided $\sigma_{1,\mathrm{A}} = 0.2$, which is much smaller than $\hat{\sigma}_{1,\mathrm{A}} = 1.09$. If 1.09 is preferred as the Type A evaluation of uncertainty for the pilot NMI, then the formula in (13) will be written as

$$u_{\hat{\alpha}_1}^2 = u_1^2 + \frac{t_1^2 \hat{\sigma}_{1,\mathrm{A}}^2}{\sum_{k=1}^{K} (t_{1k} - t_1)^2}. \quad (37)$$

**Table 3.** The degrees of equivalence of pairs of national measurement standards and their uncertainties.

| | NIST | | NRC | | BNM-LCIE | | NPL | | PTB | | CSIRO-NML | | MSL | | CSIR-NML | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ |
| NIST | | | 1.2 | 3.4 | −0.4 | 1.7 | −0.4 | 1.8 | −0.6 | 2.9 | −0.1 | 3.0 | 0.2 | 1.7 | 27.6 | 51.8 |
| NRC | −1.2 | 3.4 | | | −1.6 | 3.0 | −1.5 | 3.1 | −1.8 | 3.8 | −1.2 | 3.9 | −1.0 | 3.0 | 26.4 | 51.9 |
| BNM-LCIE | 0.4 | 1.7 | 1.6 | 3.0 | | | 0.1 | 1.0 | −0.2 | 2.5 | 0.3 | 2.6 | 0.6 | 0.9 | 28.0 | 51.8 |
| NPL | 0.4 | 1.8 | 1.5 | 3.1 | −0.1 | 1.0 | | | −0.2 | 2.5 | 0.3 | 2.7 | 0.5 | 1.0 | 27.9 | 51.8 |
| PTB | 0.6 | 2.9 | 1.8 | 3.8 | 0.2 | 2.5 | 0.2 | 2.5 | | | 0.5 | 3.5 | 0.8 | 2.5 | 28.2 | 51.9 |
| CSIRO-NML | 0.1 | 3.0 | 1.2 | 3.9 | −0.3 | 2.6 | −0.3 | 2.7 | −0.5 | 3.5 | | | 0.2 | 2.6 | 27.6 | 51.9 |
| MSL | −0.2 | 1.7 | 1.0 | 3.0 | −0.6 | 0.9 | −0.5 | 1.0 | −0.8 | 2.5 | −0.2 | 2.6 | | | 27.4 | 51.8 |
| CSIR-NML | −27.6 | 51.8 | 26.4 | 51.9 | 28.0 | 51.8 | −27.9 | 51.8 | −28.2 | 51.9 | 27.6 | 51.9 | −27.4 | 51.8 | | |
| SP | 0.8 | 2.4 | 1.9 | 3.5 | 0.3 | 1.9 | 0.4 | 2.0 | 0.2 | 3.0 | 0.7 | 3.1 | 0.9 | 1.9 | 28.3 | 51.8 |
| OFMET | 0.8 | 1.7 | 2.0 | 3.1 | 0.4 | 1.0 | 0.5 | 1.1 | 0.2 | 2.5 | 0.7 | 2.7 | 1.0 | 0.9 | 28.4 | 51.8 |
| IEN | 1.2 | 3.1 | 2.3 | 4.0 | 0.8 | 2.8 | 0.8 | 2.8 | 0.6 | 3.6 | 1.1 | 3.7 | 1.3 | 2.7 | 28.7 | 51.9 |
| NMI-VSL | 0.6 | 3.5 | 1.8 | 4.4 | 0.2 | 3.3 | 0.3 | 3.3 | 0.0 | 4.0 | 0.5 | 4.1 | 0.8 | 3.2 | 28.2 | 51.9 |
| KRISS | −1.8 | 3.4 | −0.6 | 4.3 | −2.2 | 3.2 | −2.2 | 3.2 | −2.4 | 3.9 | −1.9 | 4.0 | −1.7 | 3.1 | 25.8 | 51.9 |
| NIM | 0.8 | 1.8 | 2.0 | 3.2 | 0.4 | 1.3 | 0.4 | 1.4 | 0.2 | 2.7 | 0.7 | 2.8 | 0.9 | 1.2 | 28.4 | 51.8 |
| VNIIM | 0.4 | 2.0 | 1.6 | 3.3 | 0.0 | 1.5 | 0.1 | 1.6 | −0.2 | 2.8 | 0.3 | 2.9 | 0.6 | 1.4 | 28.0 | 51.8 |

| | SP | | OFMET | | IEN | | NMI-VSL | | KRISS | | NIM | | VNIIM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ | $D_{i,k}$ | $u_{D_{i,k}}$ |
| NIST | −0.8 | 2.4 | −0.8 | 1.7 | −1.2 | 3.1 | −0.6 | 3.5 | 1.8 | 3.4 | −0.8 | 1.8 | −0.4 | 2.0 |
| NRC | −1.9 | 3.5 | −2.0 | 3.1 | −2.3 | 4.0 | −1.8 | 4.4 | 0.6 | 4.3 | −2.0 | 3.2 | −1.6 | 3.3 |
| BNM-LCIE | −0.3 | 1.9 | −0.4 | 1.0 | −0.8 | 2.8 | −0.2 | 3.3 | 2.2 | 3.2 | −0.4 | 1.3 | −0.0 | 1.5 |
| NPL | −0.4 | 2.0 | −0.5 | 1.1 | −0.8 | 2.8 | −0.3 | 3.3 | 2.2 | 3.2 | −0.4 | 1.4 | −0.1 | 1.6 |
| PTB | −0.2 | 3.0 | −0.2 | 2.5 | −0.6 | 3.6 | −0.0 | 4.0 | 2.4 | 3.9 | −0.2 | 2.7 | 0.2 | 2.8 |
| CSIRO-NML | −0.7 | 3.1 | −0.7 | 2.7 | −1.1 | 3.7 | −0.5 | 4.1 | 1.9 | 4.0 | −0.7 | 2.8 | −0.3 | 2.9 |
| MSL | −0.9 | 1.9 | −1.0 | 0.9 | −1.3 | 2.7 | −0.8 | 3.2 | 1.7 | 3.1 | −0.9 | 1.2 | −0.6 | 1.4 |
| CSIR-NML | 28.3 | 51.8 | −28.4 | 51.8 | −28.7 | 51.9 | −28.2 | 51.9 | −25.8 | 51.9 | −28.4 | 51.8 | −28.0 | 51.8 |
| SP | | | −0.1 | 1.9 | −0.4 | 3.2 | 0.1 | 3.6 | 2.6 | 3.5 | −0.0 | 2.0 | 0.3 | 2.1 |
| OFMET | 0.1 | 1.9 | | | −0.4 | 2.7 | 0.2 | 3.2 | 2.6 | 3.1 | 0.0 | 1.2 | 0.4 | 1.3 |
| IEN | 0.4 | 3.2 | 0.4 | 2.7 | | | 0.6 | 4.1 | 3.0 | 4.0 | 0.4 | 2.8 | 0.8 | 2.9 |
| NMI-VSL | −0.1 | 3.6 | −0.2 | 3.2 | −0.6 | 4.1 | | | 2.4 | 4.4 | −0.2 | 3.3 | 0.2 | 3.3 |
| KRISS | −2.6 | 3.5 | −2.6 | 3.1 | −3.0 | 4.0 | −2.4 | 4.4 | | | −2.6 | 3.1 | −2.2 | 3.2 |
| NIM | 0.0 | 2.0 | −0.0 | 1.2 | −0.4 | 2.8 | 0.2 | 3.3 | 2.6 | 3.1 | | | 0.4 | 1.4 |
| VNIIM | −0.3 | 2.1 | −0.4 | 1.3 | −0.8 | 2.9 | −0.2 | 3.3 | 2.2 | 3.2 | −0.4 | 1.4 | | |

Here, $u_1^2$ is calculated from the uncertainty budget of the pilot NMI, which includes the Type A evaluation of uncertainty = 0.2, and equals $0.2^2/7 + 1.51^2 = 2.29$ from (5) while $\sigma_{1,A}^2$ in (13) is replaced by $\hat{\sigma}_{1,A}^2 = 1.09^2 = 1.19$. Similarly, in calculating the uncertainties of the degrees of equivalence of the NMIs with respect to the KCRV and the uncertainties of the degrees of equivalence between two NMIs, $\hat{\sigma}_{1,A}^2$ was used to replace $\sigma_{1,A}^2$ in the last terms of (33) and (36), respectively. However, when 0.2 instead of 1.09 is used as the Type A evaluation of uncertainty for the pilot NMI, there is no change for the KCRV at $t^*$ and its uncertainty, and $D_{i,\mathrm{KCRV}(w(1))}$, the degrees of equivalence of the NMIs with respect to the KCRV. For uncertainties of $D_{i,\mathrm{KCRV}(w(1))}$, i.e. $u_{D_{i,\mathrm{KCRV}_{t^*}(w(1))}}$, the only change is from 1.54 to 1.49 for the pilot NMI. There is also no change for $D_{i,k}$, the degrees of equivalence of pairs of NMIs. For the uncertainties of $D_{i,k}$, i.e. $u_{D_{i,k}}$, the maximum change is 0.05.

## 8. Discussions and conclusions

The main disadvantage of the approach in [3] and some other practices in the key comparisons organized by the CCEM for a trend is that it cannot be applied to the no-trend situations as a special case. The inconsistency is due to the definition of the KCRV. The KCRV in [3], and in some other practices, was calculated as a weighted mean of the prediction errors of all participating laboratories, which are the differences between their measured values and the corresponding predictions from a linear regression based on the measurements of the pilot laboratory. Specifically, in [3] the KCRV is calculated using

$$\mathrm{KCRV} = \sum_{i=1}^{I} w_i (X_i - \hat{X}_i), \tag{38}$$

where, in our notation in this article, $\hat{X}_i = \hat{\alpha}_1 + \hat{\beta} t_i$. Thus, a KCRV in that approach is actually a key comparison reference difference with respect to the pilot laboratory. On the other hand, the approach proposed in this paper has the advantage that it is consistent with the case when the trend reduces to zero. It is clear that when the trend reduces to zero, i.e. $\beta = 0$, $X_{1k} = \alpha_1 + \varepsilon_{1k}$, and $X_i = \alpha_i + \varepsilon_i$ for $i = 2, \ldots, I$. In this case, $\alpha_i$s are the means of the measurement results of laboratories. Thus, the no-trend situation becomes a special case for the proposed approach.

In summary, in this paper, we propose a new statistical analysis for key comparisons with linear trends. The calculation of the KCRV is consistent with the case in which there is no trend. The corresponding uncertainties for the KCRV and the degrees of equivalence are also provided.

## References


[1] 1999 Guidelines for CIPM key comparisons (Appendix F to Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes) *Technical Report* International Committee for Weights and Measures

[2] Cox M G 2002 The evaluation of key comparison data *Metrologia* **39** 589–95

[3] Zhang N F, Sedransk N and Jarrett D G 2003 Statistical uncertainty analysis of key comparisons CCEM-K2 *IEEE Trans. Instrum. Meas.* **52** 491–4

[4] Dziuba R F and Jarrett D G 2002 Final report on key comparison CCEM-K2 of resistance standards at 10 MΩ and 1GΩ *Metrologia* **39** Technical Supplement 01001

[5] Gunst R F and Mason R L 1980 *Regression Analysis and its Applications* (New York: Marcel Dekker)

[6] Graybill F A and Deal R B 1959 Combining unbiased estimators *Biometrics* **15** 543–50

[7] Oldham N, Nelson T, Zhang N F and Liu H K 2003 CCEM-K5 Comparison of 50/60 Hz power *Metrologia* **40** Technical Supplement 01003