Uncertainty Analysis of Interlaboratory Studies with Linear Trends

Nien F. Zhang, Hung-kung Liu, Nell Sedransk, and William E. Strawderman
National Institute of Standards and Technology, Gaithersburg, MD

Interlaboratory studies have been used to ensure accurate measurement capability among different laboratories. In this paper, we discuss a special interlaboratory comparison called key comparison. 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 is responsible to organizing the circulation and transport of standards or artifacts and ensuring that the participant NMIs make proper arrangements. See the Appendix F of the "Mutual Recognition of National Measurements Standards and of Measurement Certificates issued by National Metrology Institutes" [1]. Usually each non-pilot NMI makes measurements in one period while the pilot NMI makes measurements in multiple periods based on the circulation scheme. 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, 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, Sedransk, and Jarrett [2], an approach was proposed and applied to the CCEM-K2 Key Comparison of Resistance Standards at 10 $M\Omega$ and 1 $G\Omega$ (see Dziuba and Jarrett [3]). 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 KCRV in the existing methods dealing with trend, the results cannot be applied to the trivial case when the trend reduces to zero. In this article, we define the time-dependent KCRV in an appropriate way and propose a new approach that overcomes the shortcomings in [2] and [3]. The calculation of KCRV is consistent with the case in which there is no trend. The pair-wise comparison between any two national measurement standards and its corresponding uncertainty, which is independent of the time, are also given.

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 article we consider the case of a single artifact circulating through all laboratories. 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} , $k=1,...,K$, which are taken to have negligible associated uncertainty, and $X_{1k}, k=1,...,K$ with $K>2$, respectively. In practice, t_{1k} can be an average value of the time when the measurements were made in the kth period and then X_{1k} is the average value of the corresponding measurements in that period.

We assume that a simple linear regression model holds for the measurements made by the pilot laboratory,

$$X_{1k} = a_1 + b\,t_{1k} + e_{1k} \qquad (1)$$

for $k=1,...,K$. We assume that the random error e_{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

$$X_i = a_i + b\,t_i + e_i, \qquad (2)$$

where the random error e_i has a zero mean and standard uncertainty of u_i for $i = 2,...,I$. In a key comparison study, NMI's will provide x_i's, the values of X_i's, and the corresponding uncertainties including the Type A and Type B evaluations of uncertainty.

For the pilot laboratory, the least squares estimators of the regression parameters are given by

$$\hat{b} = \frac{\sum_{k=1}^{K}(t_{1k}-t_1)(X_{1k}-X_1)}{\sum_{k=1}^{K}(t_{1k}-t_1)^2} \qquad (3)$$

and

$$\hat{a}_1 = X_1 - \hat{b}\,t_1, \qquad (4)$$

where X_1 is the average of the measurements made by the pilot laboratory and t_1 is the average of $\{t_{11},...,t_{1K}\}$. For other laboratories, i.e., $i = 2,...,I$, the corresponding intercepts can be estimated by

$$\hat{a}_i = X_i - \hat{b}\,t_i, \qquad (5)$$

where t_i is the time when the i^{th} $(i \neq 1)$ laboratory made its measurement.

Key Comparison Reference Value

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

$$KCRV_t(w) = \sum_{i=1}^{I} w_i(\hat{a}_i + \hat{b}t) = \sum_{i=1}^{I} w_i X_i - \hat{b}\sum_{i=1}^{I} w_i(t_i - t) \qquad (6)$$

with weights $w = (w_1,..., w_I)$ satisfying $\sum_{i=1}^{I} w_i = 1$. $KCRV_t(w)$ is a function of t as well as the weights. The uncertainty of KCRV depends on t as well as the weights w. For any given t, a reasonable criterion to choose optimal weights is the minimum variance or minimum standard uncertainty of $KCRV_t(w)$. For any given set of weights, the minimum value of the standard uncertainty of $KCRV_t(w)$ will occur when $\sum_{i=1}^{I} w_i(t_i - t)^2$ equals zero. This occurs when t takes the value of $t_w = \sum_{i=1}^{I} w_i t_i$. In this case, from (6), the KCRV has the same form as in the no trend case and is given by

$$KCRV_{t_w}(w) = \sum_{i=1}^{I} w_i X_i . \qquad (7)$$

The corresponding uncertainty is given by

$$u^2_{KCRV_{t_w}(w)} = \sum_{i=1}^{I} w_i^2 u_i^2 . \qquad (8)$$

As in the no trend case, $u^2_{KCRV_{t_w}(w)}$ is minimized when the weights are given by

$$w_i(1) = \frac{\frac{1}{u_i^2}}{\sum_{k=1}^{I} \frac{1}{u_k^2}} . \qquad (9)$$

The corresponding KCRV is given by

$$KCRV_{t^*}(w(1)) = \sum_{i=1}^{I} w_i(1) X_i , \qquad (10)$$

where

$$t^* = \sum_{i=1}^{I} w_i(1) t_i , \qquad (11)$$

and $w(1) = (w_i(1),..., w_I(1))$. From (9) and (10), the uncertainty of this KCRV is the same as in the no-trend case and is given by

$$u^2_{KCRV_{t^*}(w(1))} = \frac{1}{\sum_{i=1}^{I} \frac{1}{u_i^2}} \qquad (12)$$

assuming that u_i ($i = 1,..., I$) in the weights $\{w_i(1)\}$ are the standard deviations for all laboratories. As a practical matter, we recommend the $KCRV_{t^*}(w(1))$ with t^* and the weights given by (9) and (11) to be the reference value of this key comparison.

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 $KCRV_t$ is defined as the difference:

$$\begin{aligned} D_{i,KCRV(w)} &= \hat{a}_i + \hat{b}t - KCRV_t(w) \\ &= (1-w_i)\hat{a}_i + \sum_{j\neq i, j=1}^{I} w_j \hat{a}_j. \end{aligned} \qquad (13)$$

when $w = w(1)$ as given in (9), in the first equality of (13), let $t = t^* = \sum_{i=1}^{I} w_i(1)t_i$. Then

$$\begin{aligned} D_{i,KCRV(w(1))} &= \hat{a}_i + \hat{b}t^* - KCRV_{t^*}(w(1)) \\ &= \hat{a}_i + \hat{b}t^* - \sum_{i=1}^{I} w_i(1)X_i. \end{aligned} \qquad (14)$$

The corresponding uncertainty is given by

$$u^2_{D_{i,KCRV(w(1))}} = [1-2w_i(1)]u_i^2 + \frac{1}{\sum_{k=1}^{I} \frac{1}{u_k^2}} + \frac{(t_i - t^*)^2 s^2_{1,A}}{\sum_{k=1}^{K}(t_{1k} - t_1)^2}, \qquad (15)$$

where $s_{1,A}$ is the Type A evaluation of uncertainty of the measurements made by the pilot laboratory. An unbiased estimator of $s^2_{1,A}$ in (15) is given by the "mean-square residual" based on the model in (1), i.e.,

$$\hat{s}^2_{1,A} = \frac{\sum_{k=1}^{K}(X_{1k} - \hat{a}_1 - \hat{b}t_{1k})^2}{K-2}. \qquad (16)$$

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

Degrees of equivalence of pairs of national measurement standards

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

$$D_{i,k} = (\hat{a}_i + \hat{b}t) - (\hat{a}_k + \hat{b}t) = \hat{a}_i - \hat{a}_k \tag{17}$$

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

$$u_{D_{i,k}}^2 = u_i^2 + u_k^2 + \frac{(t_i - t_k)^2 s_{1,A}^2}{\sum_{k=1}^{K}(t_{1k} - t_1)^2}. \tag{18}$$

To illustrate the approach, we applied it to the key comparison CCEM-K2. In this key comparison, three 10 MΩ wirewound 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 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 artifacts. 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 artifact, 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. A linear regression line was fit to the NIST measurements. The estimates of the intercept for NIST, the slope, and the residual standard deviation are $\hat{a}_1 = 6.03$, $\hat{b} = 1.05$, and $\hat{s}_{1,A}$=1.09, respectively. We used $\hat{s}_{1,A}$, the residual standard deviation from (16) as an estimate of $s_{1,A}$ in (15) and (18). The date for minimum standard deviation of KCRV is $t^* = 1998.23$ given by (11). The KCRV at t^* with weights $w(1)$ given by (9) and (10) is 8.03 and the corresponding uncertainty is 0.28 given by (12). The $D_{iKCRV_{t^*}(w(1))}$ for all NMIs and their corresponding uncertainties $u_{D_{iKCRV_{t^*}(w(1))}}$ were calculated.

The degrees of equivalence of pairs of NMIs and their corresponding uncertainties are calculated by (17) and (18), respectively.

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

1. 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, 1999.

2. Zhang, N. F., Sedransk, N., and Jarrett, D. G., Statistical Uncertainty Analysis of Key Comparisons CCEM-K2, *IEEE Transaction Instrumentation & Measurement*, 2003, 491-494.

3. Dziuba, R. F. and Jarrett, D. G. Final Report CCEM-K2 Key Comparison of Resistance Standards at 10 MΩ and 1GΩ. Appendix B of *Mutual Recognition Arrangement*, 2002, Bureau International des poids et Measures (BIPM), Paris France.

Table 1. Information for the Standard S/N HR7551

Lab	Mean date of Measurement	Measurement (x 10^{-6})	Type A uncertainty (x 10^{-6})	Type B uncertainty (x 10^{-6})
NIST	1996.65	4.6	0.2 *	1.51*
NRC	1996.80	5	1.88	2.29
NIST	1996.94	6.7	0.2 *	1.51*
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 *	1.51*
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 *	1.51*
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 *	1.51*
KRIST	1999.39	7.1	0.30	3
NIST	1999.60	8.5	0.2 *	1.51*
NIM	1999.87	10.2	0.10	0.83
VNIIM	2000.03	10	0.25	1.03
NIST	2000.20	10	0.2 *	1.51*