DRAFT B, Final Report on CIPM key comparison of 1 kg standards in stainless steel (CCM.M-K1)

C. Aupetit[1], L.O. Becerra[2], N. Bignell[3], W. Bich[4], G.D. Chapman[5], J.W. Chung[6], J. Coarasa[7], S. Davidson[8], R. Davis[7], N.G. Domostroeva[9], K.M.K. Fen[3], M. Gläser[10], W.G. Lee[6], M. Lecollinet[1], Q. Li[11], A. Ooiwa[12], R. Spurny[13], A. Torino[4], J.C.G.A. Verbeek[14], Z.J. Jabbour[15]

1. Introduction

This report describes a Key Comparison of 1 kg standards in stainless steel undertaken by the Consultative Committee on Mass and Related Quantities (CCM). The report has been prepared following the *Guidelines for key comparisons carried out by Consultative Committees (draft of 1 March 1999)*. It has not been possible to follow these guidelines in all particulars because the measurements were begun in 1995.

The Bureau International des Poids et Mesures (BIPM) was the pilot laboratory for these comparisons. All other participants are member countries of the CCM except Mexico which is an official observer:

Laboratory		Country
Commonwealth Scientific and Industrial Research Organisation	CSIRO	Australia
National Research Council of Canada	NRC	Canada
National Institute of Metrology	NIM	China
Bureau National de Métrologie-INM/CNAM	BNM-INM/CNAM	France
Physikalisch-Technische Bundesanstalt	PTB	Germany
Istituto di Metrologia “G. Colonnetti”	IMGC	Italy
National Metrology Institute of Japan, National Institute of Advanced Science and Technology	NMIJ/AIST[16] (ex NRLM)	Japan
Korea Research Institute of Standards and Science	KRISS	Republic of Korea

[1] Bureau National de Métrologie (BNM-INM/CNAM), 75003 Paris, France
[2] Centro Nacional de Metrologia (CENAM), CP 76900 Querétaro, Mexico
[3] Commonwealth Scientific and Industrial Research Organisation (NML-CSIRO), Lindfield NSW 2070, Australia
[4] Istituto di Metrologia G. Colonnetti (IMGC), 10135 Torino, Italy
[5] National Research Council of Canada (NRC-CNRC), K1A OR6 Ottawa, Canada
[6] Korea Research Institute of Standards and Science (KRISS), 305-600 Daejeon, Republic of Korea
[7] Bureau International des Poids et Mesures (BIPM), 92312 Sèvres Cedex, France
[8] National Physical Laboratory (NPL), Teddington, Middlesex TW11 OLW, United Kingdom
[9] D.I. Mendeleyev Institute for Metrology, Gosstandart of Russia (VNIIM), 198005 St. Petersburg, Russia
[10] Physikalisch-Technische Bundesanstalt (PTB), 38023 Braunschweig, Germany
[11] National Institute of Metrology (NIM), 100013 Beijing, China
[12] National Institute of Advanced Industrial Science and Technology (NMIJ/AIST), Ibaraki 305-8563, Japan
[13] Slovak Metrological Institute (SMU), 84255 Bratislava, Slovakia
[14] Van Swinden Laboratorium (NMi/VSL), 2600 Delft, Netherlands
[15] National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, United States of America

[16] Throughout the body of this report, we refer to the “NRLM” although this laboratory has now become the NMIJ/AIST.

Centro Nacional de Metrologia	CENAM	Mexico
Van Swinden Laboratorium	VSL	Netherlands
D.I. Mendeleyev Institute for Metrology	VNIIM	Russia
Slovak Metrological Institute	SMU	Slovakia
National Physical Laboratory	NPL	United Kingdom
National Institute of Standards and Technology	NIST	USA

We begin with the results as obtained in each laboratory. We then explain the steps we have taken to account for: imperfect stability of the travelling standards[17], possible changes in the BIPM working standards and the fact that half the laboratories in the comparison used one set of travelling standards and half used a second set.

The comparisons were carried out over a period of about 1000 days between 1995-02 and 1997-11. All participants had reported their essential results to the pilot laboratory by 1998-02. The period of the comparisons was much longer than originally foreseen. Only a few months of the prolongation can be attributed to the addition of two more participants (KRISS and CENAM) after the comparisons had begun, since the added laboratories were among the most efficient. Rather, the explanation lies primarily in various scheduling changes required by participants and several difficulties in arranging for hand transport of the travelling standards. To this list must be added a short period during which the balance used by the pilot laboratory underwent repair.

Most reports were received from participants within a reasonable time but a few were greatly delayed. We should note that the present *Guidelines* (which, of course, did not exist when we began the comparisons) ask that the report to the pilot laboratory should be written within six weeks after completion of the measurements.

2. Summary of results received from participants

Seven participating laboratories determined the mass of travelling standards VSL-1 and J2 (Package 1) while the remaining seven determined the mass of travelling standards VSL-2 and J3 (Package 2). As pilot laboratory, the BIPM determined the mass of each package at the beginning and the end of the comparisons and four times during the course of the comparisons. These results are summarized in Table 1. The basic model describing the measurements of each participant is given in Appendix 4.

In the following, it has seemed preferable to examine the average mass for each package and the average difference for each package. The first quantity depends sensitively on traceability to the international prototype (involving an important correction for air buoyancy) but averages over other influences. The second quantity is virtually independent of the international prototype and involves a much less important correction for air buoyancy. Consequently, the second quantity more clearly indicates the relative stability of the travelling standards and the reproducibility of the balance used in the measurement. The difference in mass as obtained by each participant and the pilot laboratory thus serves as a valuable diagnostic tool.

[17] The protocol for the comparisons refers to these as “transfer standards”. However, it seems that this term may be ambiguous or confusing. For this reason we use “travelling standards” in this report to refer to standards VSL-1, J2, VSL-2 and J3.

Approx. Date	**Lab.**	**(m_{VSL1}-1kg)/ mg**	**(m_{J2}-1kg)/ mg**	**(m_{VSL2}-1kg)/mg**	**(m_{J3}-1kg)/mg**	**u_{c} / µg**	**drift / µg**	**u(drift)/ µg**
95/02	**BIPM**	**0.478**	**3.378**	**0.042**	**3.564**	**12**		
95/05	NMi,c	0.457	3.353			18	-5	3
95/07	NIST			0.010	3.540	19	-7	4
95/07	NPL	0.478	3.367			16	-5	3
95/09	NRC,c			0.012	3.540	16	-7	4
95/09	**BIPM**	**0.471**	**3.365**			**12**		
96/02	NRLM,c			0.007	3.538	13	-7	4
96/02	VNIIM	0.529	3.396			24	-2	1
96/05	**BIPM**	**0.473**	**3.357**	**0.024**	**3.554**	**12**		
96/07	CSIRO			0.025	3.557	14	1	0
96/08	PTB	0.467	3.354			12	0	0
96/09	**BIPM**			**0.024**	**3.556**	**12**		
96/10	NIM			0.087	3.573	21	0	9
96/12	SMU	0.532	3.412			22	0	0
97/02	**BIPM**	**0.474**	**3.383**	**0.083**	**3.586**	**12**		
97/02	**BIPM**	**0.473**	**3.360**	**0.041**	**3.564**	**12**		
97/03	KRISS	0.468	3.357			14	0	0
97/04	**BIPM**	**0.471**	**3.364**			**12**		
97/05	IMGC			0.034	3.561	13	-2	1
97/06	**BIPM**			**0.033**	**3.564**	**12**		
97/07	BNM	0.470	3.367			10	-2	1
97/09	**BIPM**	**0.468**	**3.359**			**12**		
97/09	CENAM			0.031	3.556	13	-4	2
97/10	**BIPM**			**0.023**	**3.559**	**12**		

Table 1. Column 1 gives the approximate date at which the laboratory identified in column 2 made its measurements. The reported results found for the travelling standards are given in columns 3 through 6. The combined standard uncertainty (k = 1) reported by each participant for the average mass of both travelling standards is given in column 7. Columns 8 and 9 are discussed in the text. These last two columns show the drift in the average mass of the travelling standards and its estimated standard uncertainty.

The results of Table 1 are used directly in Figures 1 through 5. The following information is necessary to understand these figures :

- The uncertainty bars in Figures 1 and 2 have been provided by the participants. These are the same as the values given in column 7. Only in the case of perfect correlation will the uncertainty assigned by a participant to the mass of an individual travelling standard be the same as the uncertainty assigned to the average mass of both travelling standards. Most participants (including the pilot laboratory) claimed almost perfect correlation. The remaining participants assigned the following combined standard uncertainties to each

1 kg travelling standard: NPL (17 µg), CSIRO (15 µg), NIM (25 µg), BNM (12 µg), CENAM (16 µg).

- As explained below, we have assigned the same standard uncertainty to all BIPM measurements. For clarity, the uncertainty bars are only shown for the first BIPM measurements.
- The NMi was the first to report its measurements. Since this was (at the time) an unexpected result, and since the *Guidelines* had not yet been developed, we asked the NMi to re-examine its data to see if there could be some error that had been overlooked. We at the pilot laboratory also did additional checks of our own results. As part of its research, the NMi sent its prototype (No. 53) to the PTB to seek confirmation of the certified value. (Note: The NMi had cleaned and washed its prototype prior to calibrating its own working standards in stainless steel; the PTB did not attempt to clean the NMi prototype.) The PTB results indicated a problem and so the NMi then sent their prototype to the BIPM for recalibration. On arrival, the NMi prototype was 44 µg heavier than the value reported in the third verification of national prototypes of the kilogram [1]. (Successive cleaning of the NMi prototype removed about 30 µg of the excess.) Discussions between the NMi and the BIPM failed to locate the cause of this unexpected behaviour [2].

  At any rate, a correction of 44 µg was added to the original NMi result for Package 1. The corrected result is referred to as NMi,c in this report. The standard uncertainty originally reported by the NMi has not been changed.

- The average difference in the mass of the travelling standards was used to control their stability (see Figures 3 and 4 and Appendix 1). Thus we had reason to suspect that the Package 1 might have changed after its measurement by SMU and that Package 2 might have changed at some time after it left the BIPM for delivery to the NIM. In fact, the NIM reported that, due to nearby construction, they feared that there might have been a problem with dust in their laboratory. We were concerned that an abnormal accumulation of surface contamination might make the travelling standards less stable. On the other hand, we did not believe it advisable to clean the standards using liquids. We therefore brushed the standards more vigorously than normal to see if they would recover their previous values. The results, as indicated by the arrows, were reasonably successful. Needless to say, the results of the brushing were not considered when evaluating prior measurements.

  To analyze the results, we have assumed that the travelling standards changed after they left the SMU. Therefore results obtained at the PTB and the SMU are compared only to the BIPM values of 96/5.

  Subsequent to the initial analysis of the data (Draft A), the NIM has provided additional data supporting the hypothesis that the travelling standards changed before the measurements at the NIM. A standard uncertainty component of 8.5 µg (1 degree of freedom) has been assigned to this hypothesis based on the supporting data.

- The initial data submitted by the NRLM has been recalculated by that laboratory to take account of the increase in mass of their national prototype since the third verification of national prototypes and to eliminate a correction for the difference in the mass of surface moisture between platinum-iridium and stainless steel standards between ambient and vacuum conditions. The previous values reported by the NRLM were:

m_{VSL2}-1kg = -0.014 mg; m_{J3} –1kg = 3.517 mg. Each value had been assigned a standard uncertainty of 8 µg.

- The values originally reported by the NRC had failed to take account of the buoyancy correction that must be applied to small masses (about 94 mg) added to their stainless steel working standards when they were calibrated with respect to the national prototype of Canada. The correction had the effect of increasing the mass values originally reported for VSL2 and J3 by about 0.014 mg each.

- In order to see the approximate behaviour of individual standards, their masses are shown in Figure 5. Uncertainty bars have been omitted.

3. Interpretation of comparisons

The BIPM has two functions as pilot laboratory. First, it carries out what is essentially a bilateral comparison with each participant. This allows each participant to compare its assigned mass values to those of the BIPM. The second function is to allow each participant to compare its assigned mass values to those of the other 13 participating laboratories. It will be useful to examine schematically how each function is accomplished. This will point to the most efficient way to present the results.

Consider two laboratories, A and B that have participated in these comparisons. Laboratory A has assigned a value to the average mass m_{A} of the travelling standards and has reported the combined standard uncertainty $u_{\mathrm{c,A}}$. The same travelling standards were measured at the BIPM before and after the measurements of laboratory A. The measurements at the BIPM that most closely bracket those of laboratory A are $m_{\mathrm{BI,1}}$ and $m_{\mathrm{BI,2}}$. Each of these has essentially the same combined standard uncertainty, $u_{\mathrm{c,BI}}$.

We estimate *diff*(A-BIPM), the average difference in mass of the travelling standards assigned by laboratory A and the BIPM at the same moment in time, as

$$\mathit{diff}(\mathrm{A}-\mathrm{BIPM}) = m_{\mathrm{A}} - \frac{m_{\mathrm{BI,1}} + m_{\mathrm{BI,2}}}{2} \quad . \qquad (1)$$

The combined standard uncertainty of this quantity is estimated as

$$u_{\mathrm{c}}(\mathit{diff}(\mathrm{A}-\mathrm{BIPM})) = \left({u_{\mathrm{c,A}}}^2 + {u_{\mathrm{c,BI}}}^2 + \frac{(m_{\mathrm{BI,1}} - m_{\mathrm{BI,2}})^2}{12} \right)^{½} . \qquad (2)$$

Equations (1) and (2) result from several important assumptions :

- The best estimate of the BIPM mass value for comparing with laboratory A is the average of $m_{\mathrm{BI,1}}$ and $m_{\mathrm{BI,2}}$. (note: This simple view has been amended in the cases of PTB, SMU and NIM as mentioned above).
- For lack of a better model, the probability distribution for the "BIPM value" between measurements BI,1 and BI,2 is taken to be a rectangular distribution. This added component is required to account for changes in the travelling standards (see also Appendix 3). Usually, it is not when the change took place and under what circumstances.

- $m_{\mathrm{BI},1} - m_{\mathrm{BI},2}$ is significant compared with $u(m_{\mathrm{BI},1} - m_{\mathrm{BI},2})$. That is, the average mass of the standards has changed significantly compared to the standard uncertainty of the change. If this is not the case, then the third component of (2) is negligible.
- The measurements of the BIPM and laboratory A are uncorrelated[18].

Measurements reported by a second laboratory, B, are handled in the same way:

$$diff(\mathrm{B}-\mathrm{BIPM}) = m_{\mathrm{B}} - \frac{m_{\mathrm{BI},3} + m_{\mathrm{BI},4}}{2} \tag{3}$$

and

$$u_{\mathrm{c}}(diff(\mathrm{B}-\mathrm{BIPM})) = \left(u_{\mathrm{c,B}}{}^2 + u_{\mathrm{c,BI}}{}^2 + \frac{(m_{\mathrm{BI},3} - m_{\mathrm{BI},4})^2}{12} \right)^{1/2} . \tag{4}$$

In general, the measurements of laboratory B are bracketed by different BIPM measurements which we label here as BI,3 and BI,4. Equations (3) and (4) may pertain either to the same travelling standards as used by laboratory A or to different travelling standards.

It is now evident that the comparison between laboratory A and laboratory B can be determined from the quantity *diff*(A-BIPM)-*diff*(B-BIPM) = *diff*(A-B) :

As M. Gläser has pointed out, the standard uncertainty of *diff*(A-B) depends in general on whether laboratories A and B are in the same loop or in different loops. Appendix 3 provides a full explanation. For A and B in the same loop:

$$u_{\mathrm{c}}(diff(\mathrm{A}-\mathrm{B})) = \left(u_{\mathrm{c,A}}{}^2 + u_{\mathrm{c,B}}{}^2 + u'_{\mathrm{BI}}{}^2 + \frac{(m_{\mathrm{BI},1} - m_{\mathrm{BI},2})^2}{12} \right)^{1/2} \tag{5}$$

whereas for A and B in different loops

$$u_{\mathrm{c}}(diff(\mathrm{A}-\mathrm{B})) = \left(u_{\mathrm{c,A}}{}^2 + u_{\mathrm{c,B}}{}^2 + 2u'_{\mathrm{BI}}{}^2 + \frac{(m_{\mathrm{BI},1} - m_{\mathrm{BI},2})^2}{12} + \frac{(m_{\mathrm{BI},3} - m_{\mathrm{BI},4})^2}{12} \right)^{1/2} . \tag{6}$$

It is assumed that the measurements of laboratory A and B are uncorrelated.

Equations (5) and (6) contain a component due to the reproducibility of BIPM comparisons with respect to BIPM working standards in stainless steel. Additional components are due to the stability of the travelling standards. It will be shown below that u'_{BI} is estimated as 2 µg and, therefore, produces a negligible increase in the combined uncertainty.

[18] All mass measurements are traceable to the international prototype of the kilogram, the mass of which is taken to be 1 kg, without uncertainty. Other sources of correlation, such as common correction terms or the weighing design used to produce the results of the 3rd verification of national prototypes of the kilogram [1] are sufficiently small to be neglected.

4. Stability of BIPM working standards in stainless steel

Throughout the period of the comparisons, the BIPM measurements were made with respect to two 1 kg working standards in stainless steel. These are referred to as N2 and N3 and have essentially the same physical characteristics as travelling standards VSL-1 and VSL-2. Standards N2 and N3 have not been cleaned in many years. The BIPM also used as check standards two additional 1kg standards in stainless steel. These are referred to as SMUE1 and SMUE2 and were supplied by the SMU, which fabricated them. Their densities were determined by the IMGC in anticipation of their use in these international comparisons.

4.1. Internal consistency of N2 and N3

The average sum of the mass of N2 and N3, $0.5\cdot(m_{\mathrm{N2}} + m_{\mathrm{N3}})$, was used as the constraint in each BIPM calibration of the travelling standards. In order to use (5), we need to know that $m_{\mathrm{N2}} + m_{\mathrm{N3}}$ does not change between successive measurements of the travelling standards at the BIPM. The longest of these periods was 15 months. It is necessary (but not sufficient) that the measured values of $0.5\cdot(m_{\mathrm{N2}} - m_{\mathrm{N3}})$ be constant in time if neither m_{N2} nor m_{N3} changes in time. The quantity $0.5\cdot(m_{\mathrm{N2}} - m_{\mathrm{N3}})$ was determined during eight periods throughout the duration of the international comparisons (about 1000 days). These periods correspond to measurements of the mass of the travelling standards at the BIPM. There is no sign of any drift in the measured difference over this period. The standard deviation of a single measurement is $\sigma(0.5\cdot(m_{\mathrm{N2}} - m_{\mathrm{N3}})) = 0.5$ µg. It may be of interest that, although σ is small, it is nevertheless significantly greater than the standard deviation of the mean found during each of the eight periods (about 0.2 µg).

4.2. Measurements of SMUE1 and SMUE2

The standards SMUE1 and SMUE2 were measured during seven of the eight periods mentioned in the previous paragraph, also spanning the 1000 days of the international comparisons. In this case there is evidence that $0.5\cdot(m_{\mathrm{SMUE1}} - m_{\mathrm{SMUE2}})$ has changed slowly, and linearly, over time. This conclusion is independent of any assumptions about the stability of N2 and N3. In addition, the value of $0.5\cdot(m_{\mathrm{SMUE1}} + m_{\mathrm{SMUE2}})$ also has changed linearly with time with respect to the mass of N2 and N3. These results are summarized in Table 2.

Quantity	coefficient / (µg per day)	std. dev. / (µg per day)
$0.5\cdot(m_{\mathrm{SMUE1}} - m_{\mathrm{SMUE2}})$	0.0055	0.0007
$0.5\cdot(m_{\mathrm{SMUE1}} + m_{\mathrm{SMUE2}})$	–0.0128	0.0023

Table 2.

From the first row of Table 2, it is clear that either one or the other (or both) standards are drifting with time, but at different rates. Nothing similar was observed for $0.5\cdot(m_{\mathrm{N2}} - m_{\mathrm{N3}})$. We therefore interpret the second row of the table as evidence of a change in the masses of SMUE1 and SMUE2 rather than a change in N2 and N3.

It is also interesting to examine the standard deviation of a predicted point for the two linear fits:

$$\sigma(0.5\cdot(m_{\mathrm{SMUE1}} - m_{\mathrm{SMUE2}})) = 0.6\ \mu\mathrm{g}$$

$$\sigma(0.5\cdot(m_{\mathrm{SMUE1}} + m_{\mathrm{SMUE2}})) = 2.0\ \mu\mathrm{g}$$

The first standard deviation is consistent with that found for $\sigma(0.5\cdot(m_{\mathrm{N2}} - m_{\mathrm{N3}}))$. The second standard deviation suggests that the reproducibility with which the BIPM could measure the mass of a 1 kg standard in stainless steel with respect to its working standards was about 2 µg throughout the course of the international comparisons.

5. Comparison of results among participants

In order to compare results among participants, we need to consider two cases, as described above. The first case is to compare results between two participants A and B where neither A nor B is the BIPM. The second case is to compare the results between the BIPM and each of the other participants.

5.1. Comparison of results between two participants, neither being the BIPM

In this case, we use (5) and (6). Since the term $u'_{\mathrm{BI}}{}^2$ already accounts for some apparent instability of the travelling standards as determined by the BIPM, the additional terms are only significant if $|m_{\mathrm{BI,n+1}} - m_{\mathrm{BI,n}}| > 5$ µg. For each participant, the value of

$$0.5\cdot(\ m_{\mathrm{BI,n+1}} - m_{\mathrm{BI,n}})$$

and its standard uncertainty are given in the last two columns of Table 1.

The results are shown in graphical form in Figure 6. The origin of the vertical axis has been fixed at the BIPM value for no other reason than convenience. The uncertainty bars have been enlarged to account for instability in the travelling standards and for the reproducibility of BIPM measurements with respect to N2 and N3. Thus, the uncertainty bar for laboratory A is calculated as

$$u_{\mathrm{c}}(\mathit{diff}(\mathrm{A})) = \left(u_{\mathrm{c,A}}{}^2 + \frac{u'_{\mathrm{BI}}{}^2}{2} + \frac{(m_{\mathrm{BI,1}} - m_{\mathrm{BI,2}})^2}{12} \right)^{½} \tag{7}$$

where the term for instability of the standards is only significant in some cases. The last component of (7) is given in the last column of Table 1. The differences among laboratories are shown in Table 3.

Note that (6) adds the uncertainties in quadrature so that the uncertainty bars of Figure 6 may be slightly misleading. The result of addition in quadrature can be found in Table 4 (see Appendix 2).

5.2. Comparison of results between the BIPM and other participants

This is based on (1) and (2). These results are also shown in Tables 3 and 4 (see also Appendix 2). To calculate (2), we have used the following relation :

$$\left(u_\mathrm{c}(\mathit{diff}(\mathrm{A}-\mathrm{BIPM}))\right)^2 = \left(u_\mathrm{c}(\mathit{diff}(\mathrm{A}))\right)^2 + {u_\mathrm{c,BI}}^2 - \frac{\left(u'_\mathrm{BI}\right)^2}{2} \qquad (8)$$

The square root of the final two terms in (8) is 12 µg. The dashed lines in Figure 6 represent ± 12 µg. Once again, we emphasize that Figure 6 is meant to help visualize the *differences* found between laboratories as inferred from these comparisons. A single point and a single standard uncertainty have limited significance.

Perhaps a word should now be said about how the BIPM assigned values and uncertainties to its working standards N2 and N3.

First, it has been assumed that the mass of N2 and N3 has been stable throughout the comparisons. These standards are compared periodically with platinum-iridium prototypes No. 9 and No. 31. Seven such comparisons have been made between 1994-03 and 1998-03. The two prototypes were not cleaned at the time of the third verification, although they were calibrated during the third verification. They are also calibrated at five-year intervals with respect to prototype No. 25, which is washed and cleaned by the BIPM method on these occasions. Based on history prior to 1994-03, we had extrapolated a small increase in the average mass of prototypes No. 9 and No. 31 [3]. This correction was applied each time N2 and N3 were recalibrated. Finally, in 1998-03, prototypes No. 9 and No. 31 were again compared with prototype No. 25. These last measurements are consistent with the assumed extrapolation in mass.

As with the participating laboratories, our greatest uncertainty component in the calibration of N2 and N3 is in the correction for air buoyancy. As all participants except the NRLM, we have used the CIPM-81/91 formula to determine the density of air. The seven calibrations of $0.5\cdot(m_\mathrm{N2} + m_\mathrm{N3})$ with respect to No. 9 and No. 31 do not show any obvious drift. The standard deviation of a single determination is 4.3 µg and the standard deviation of the mean of the seven measurements is 1.6 µg. While the statistical scatter is relatively low, estimated uncertainties that are common to all seven measurements lead us to assign a combined standard uncertainty of 12 µg to the average. When the international comparisons began in 1995-02, the value that the BIPM had assigned to $0.5\cdot(m_\mathrm{N2} + m_\mathrm{N3})$ was within 2.5 µg of the mean of the seven measurements described above.

5.3 Comparison of results of participants with respect to the Key Comparison Reference Value

The calculation of the Key Comparison Reference Value (KCRV) is described in Appendix 2. The degree of equivalence of each participant with respect to the KCRV is given in Table 5.

6. Effective degrees of freedom

The protocol asked each laboratory to report the effective degrees of freedom ν_eff using the Welch-Satterhwaite formula [4] for the measurements shown in Figures 1 and 2. The effective degrees of freedom were greater than 10 for all laboratories. Ten laboratories (including the BIPM) reported the number of degrees of freedom to be greater than 20.

However, two laboratories expressed some concern that the Welch-Satterhwaite formula does not provide a good estimate for ν_{eff} if data are correlated.

Recall that the uncertainties given in Table 4 are based either on (2) or (6). The effective degrees of freedom of the calculated uncertainties will, in general, be greater than 10. How much greater depends on the individual case and on one's confidence in the Welch-Satterhwaite formula. In this report, we have assumed that a coverage factor of two (k=2) represents the 95% confidence limit. This approximation may underestimate the 95% confidence limit by a factor of approximately 1.15 in some cases.

References

1. Girard, G., “International Report: the third periodic verification of national prototypes of the kilogram (1988-1992)”, *Metrologia* **31** (1994) 317-336.

2. *Procès-verbaux du CIPM*, **65** (1997) p. 298.

3. *Procès-verbaux du CIPM*, **64** (1996), p. 171.

4. *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization (Geneva, Switzerland), 1993, 101 p.

5. Müller, J. “Possible advantages of a robust evaluation of comparisons”, *Rapport BIPM-95/2*, April 1995, 7 p. Reprinted with minor changes in *J. Res. NIST*, 105 (2000) 551-555.

6. Vecchia, D.F. and Splett, J.D., “Outlier-resistant methods for estimation and model fitting”, *ISA Trans.* **33** (1994) 411-420.

Figure captions

Figure 1. The value of 0.5·($m_{\mathrm{VSL}\text{-}1}$+$m_{\mathrm{J2}}$) as reported by each participant. The uncertainty bars are taken from column 7 of Table 1. The open diamonds represent measurements at the BIPM. For clarity, the uncertainty bars for the BIPM measurements are only shown once. The arrows are explained in the text. The spacing of horizontal lines of the grid is 20 µg.

Figure 2. The value of 0.5·($m_{\mathrm{VSL}\text{-}2}$+$m_{\mathrm{J3}}$) as reported by each participant. The uncertainty bars are taken from column 7 of Table 1. The open squares represent measurements at the BIPM. For clarity, the uncertainty bars for the BIPM are only shown once. The arrow is explained in the text. The spacing of horizontal lines of the grid is 20 µg.

Figure 3. The value of 0.5·($m_{\mathrm{VSL}\text{-}1}$ - m_{J2}) as reported by each participant. The open diamonds represent measurements at the BIPM. The arrow is explained in the text. The spacing of horizontal lines of the grid is 5 µg.

Figure 4. The value of 0.5·($m_{\mathrm{VSL}\text{-}2}$ - m_{J3}) as reported by each participant. The open squares represent measurements at the BIPM. The arrow is explained in the text. The spacing of horizontal lines of the grid is 5 µg.

Figure 5. The mass values of the individual standards as reported by the pilot and other participants. Constants offsets have been added to the data in order to display them on the same graph.

Figure 6. The results of all participants relative to the pilot laboratory (BIPM). The participants have been listed in chronological order. Measurements of Package 1 are represented by diamonds while measurements of Package 2 are represented by squares. Uncertainty bars are calculated from (7). The dashed lines represent the uncertainty of the BIPM. Standard uncertainties (k=1) have been used throughout. Note that the uncertainty bars are generally larger than those of Figures 1 and 2 for reasons explained in the text.

Figure 7 The results of all participants, including the BIPM, relative to the median of all measurements. The bars represent expanded uncertainties (k=2).

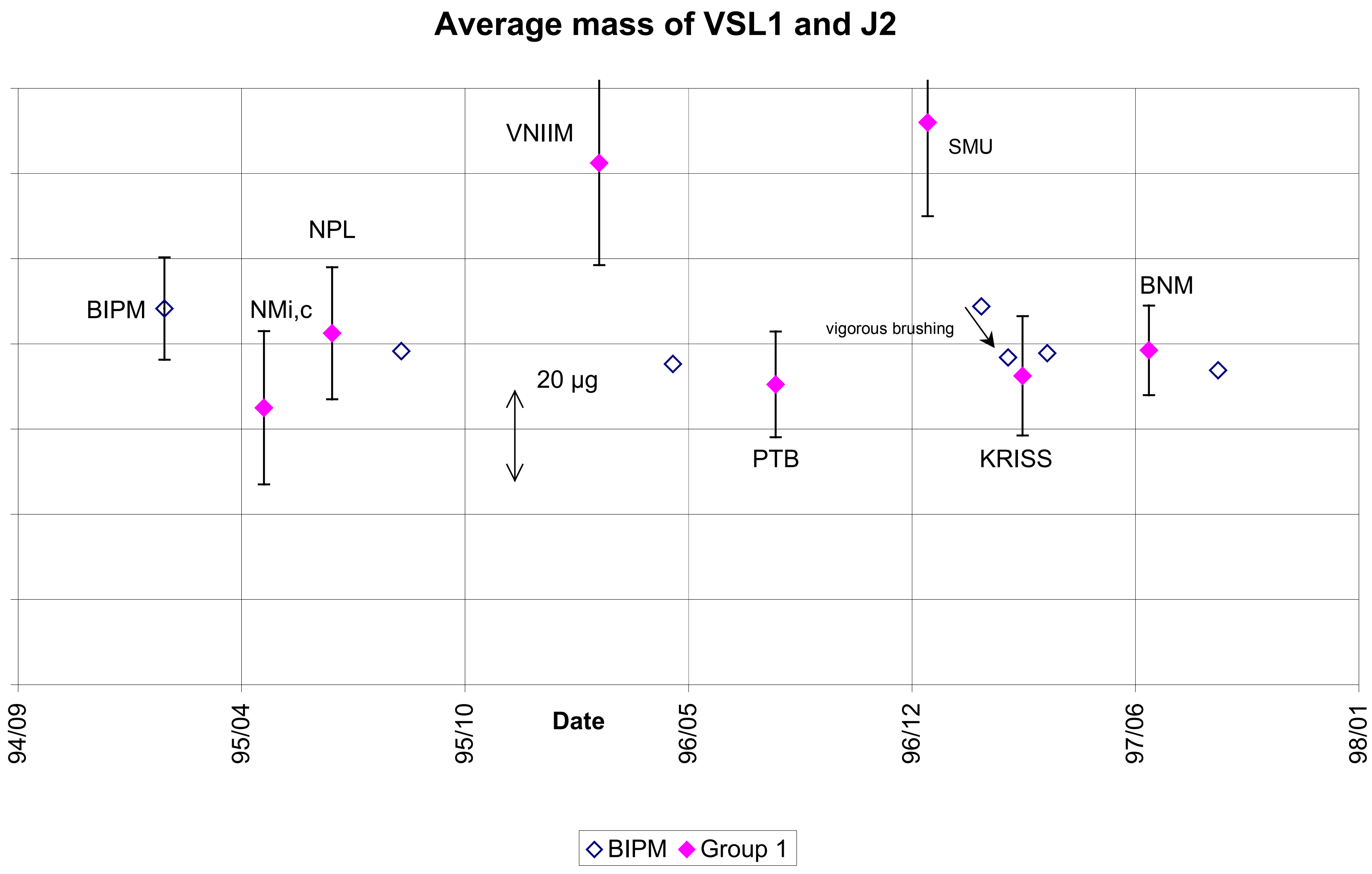

Average mass of VSL1 and J2
BIPM
NMi,c
NPL
VNIIM
20 µg
PTB
vigorous brushing
SMU
KRISS
BNM
94/09
95/04
95/10
96/05
96/12
97/06
98/01
Date
◇ BIPM ◆ Group 1

Figure 1.

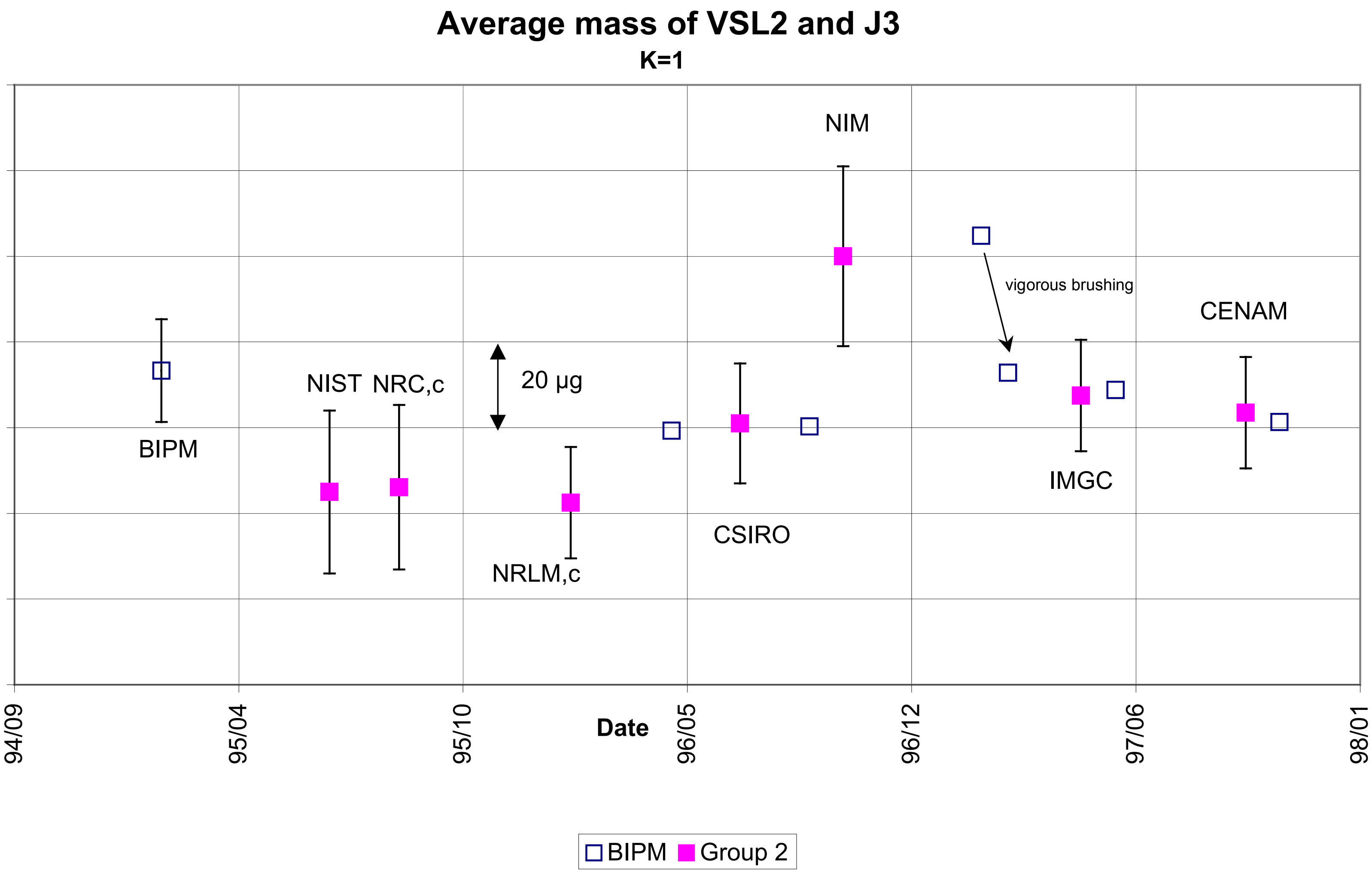

Average mass of VSL2 and J3
K=1
NIM
vigorous brushing
CENAM
BIPM
NIST
NRC,c
20 µg
NRLM,c
CSIRO
IMGC
94/09
95/04
95/10
96/05
96/12
97/06
98/01
Date
□ BIPM ■ Group 2

Figure 2.

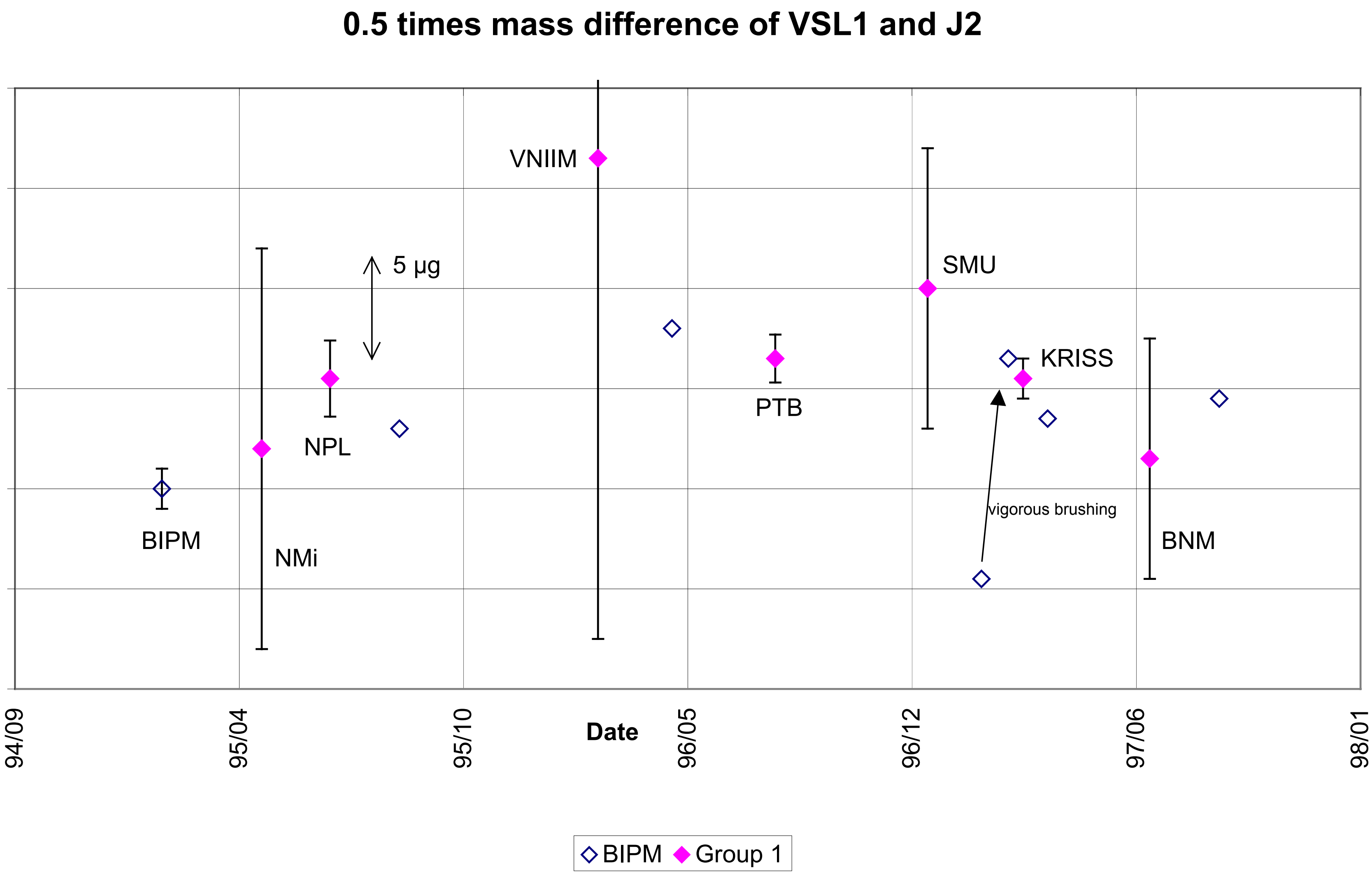

0.5 times mass difference of VSL1 and J2
VNIIM
5 µg
SMU
KRISS
PTB
NPL
BIPM
NMi
vigorous brushing
BNM
94/09
95/04
95/10
Date
96/05
96/12
97/06
98/01
◇ BIPM ◆ Group 1

Figure 3.

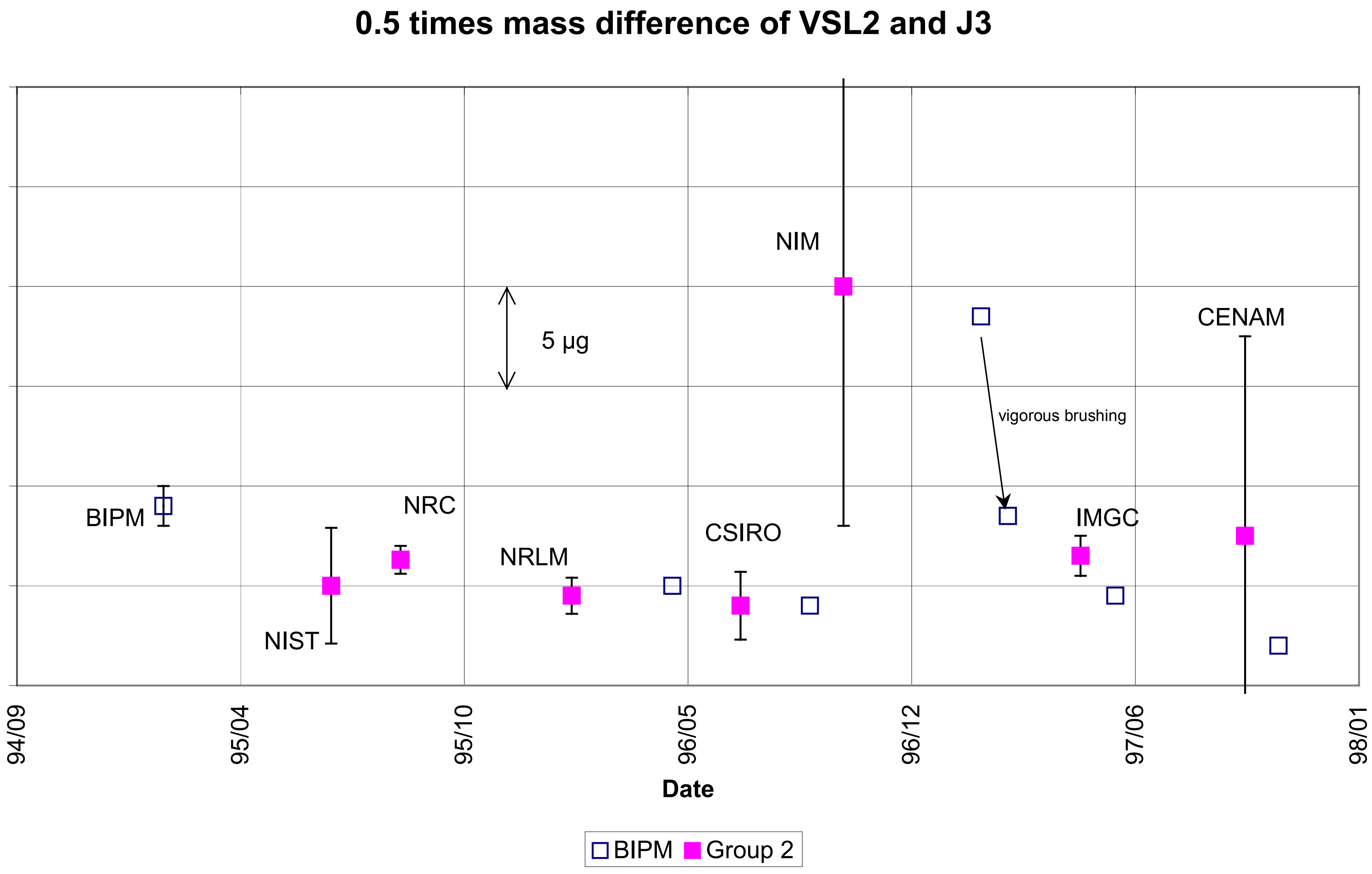

0.5 times mass difference of VSL2 and J3
BIPM
NIST
NRC
NRLM
CSIRO
NIM
vigorous brushing
IMGC
CENAM
5 µg
94/09
95/04
95/10
96/05
96/12
97/06
98/01
Date
□ BIPM ■ Group 2

Figure 4.

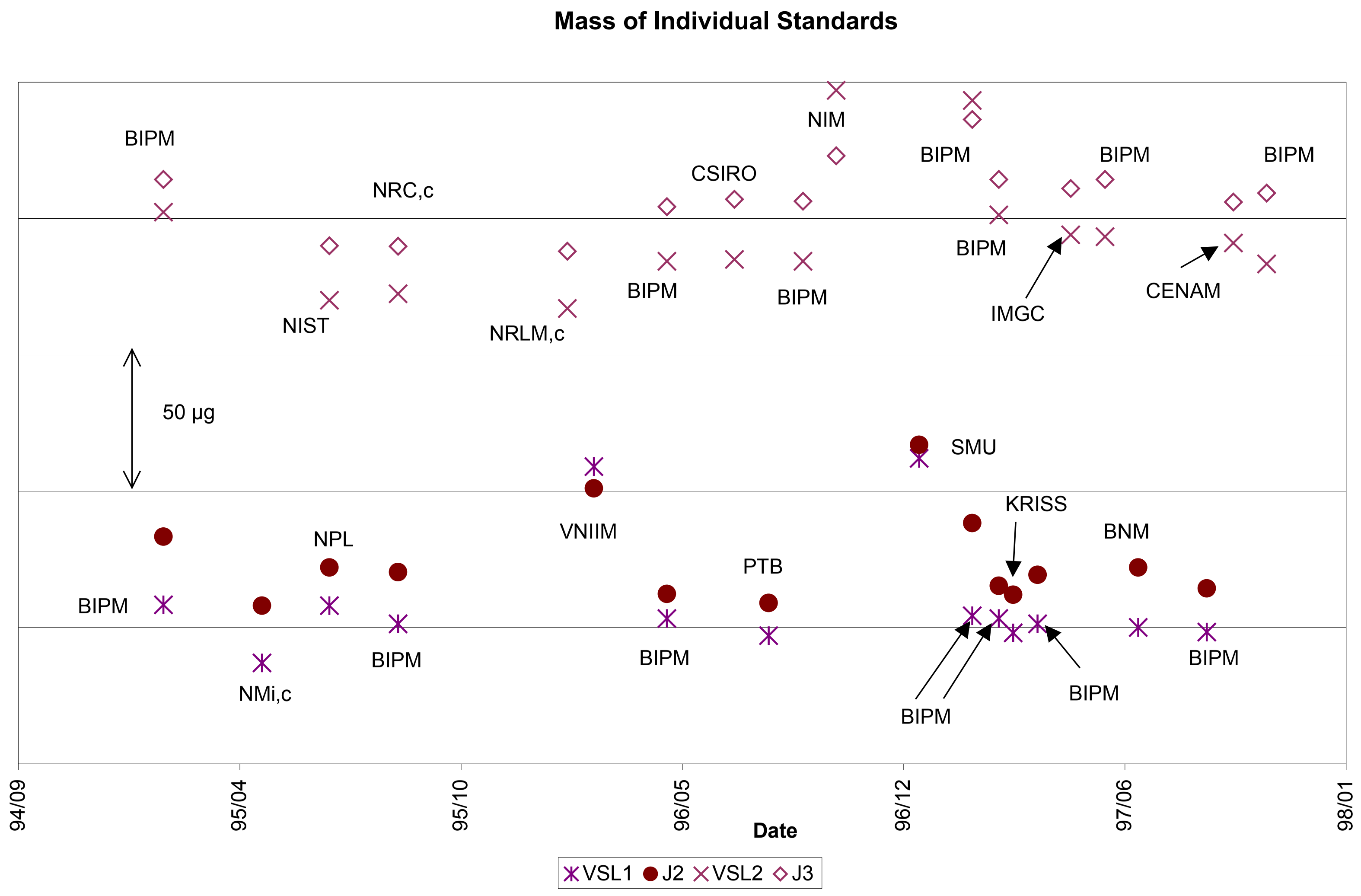

Mass of Individual Standards
BIPM
NRC,c
NIST
NRLM,c
CSIRO
NIM
BIPM
BIPM
BIPM
BIPM
BIPM
BIPM
IMGC
CENAM
50 µg
SMU
VNIIM
KRISS
NPL
BNM
PTB
BIPM
BIPM
BIPM
BIPM
BIPM
BIPM
NMi,c
BIPM
94/09
95/04
95/10
96/05
96/12
97/06
98/01
Date
VSL1
J2
VSL2
J3

Figure 5.

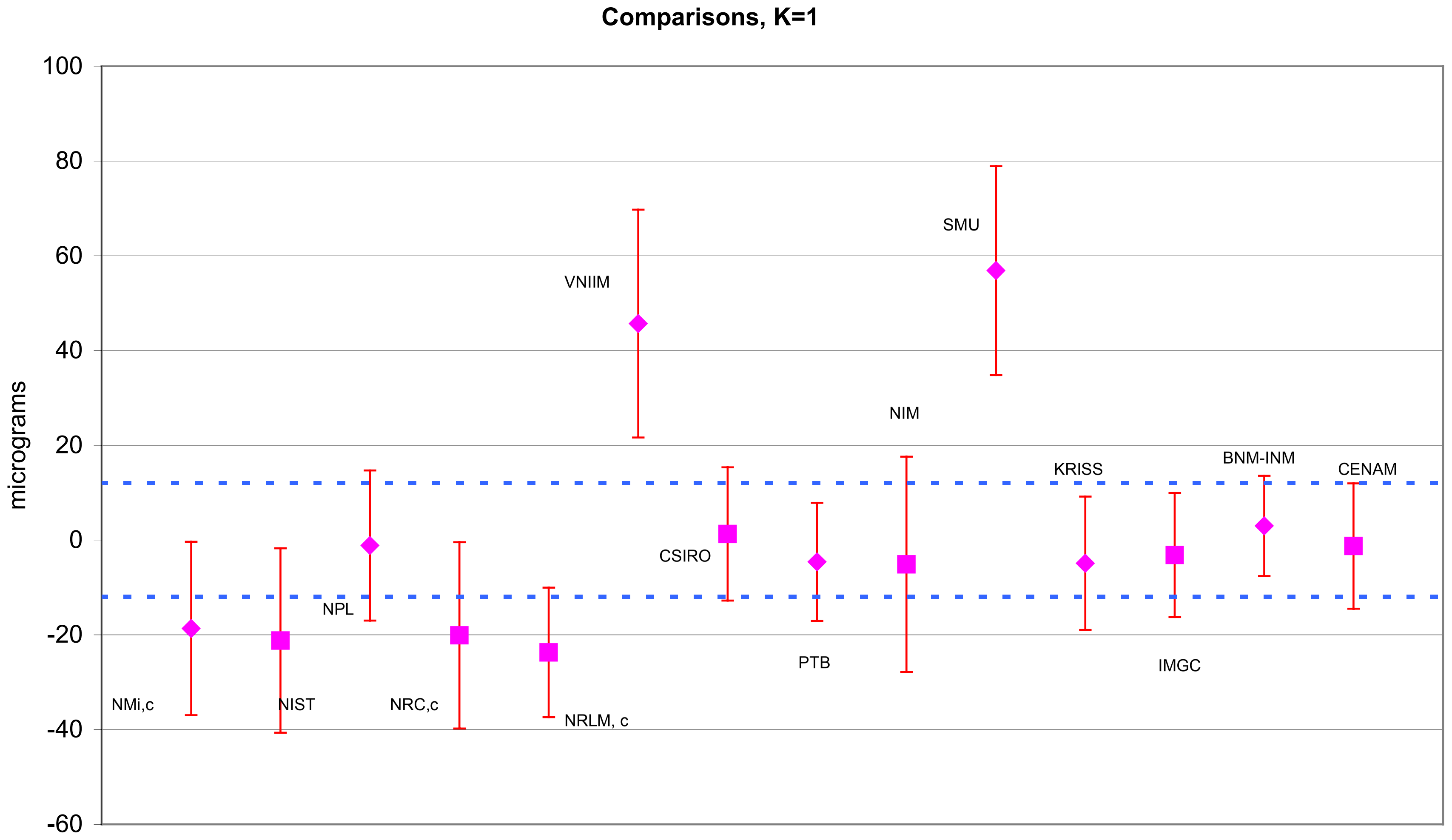

Comparisons, K=1
micrograms
100
80
60
40
20
0
-20
-40
-60
NMi,c
NIST
NPL
NRC,c
NRLM, c
VNIIM
CSIRO
PTB
NIM
SMU
KRISS
IMGC
BNM-INM
CENAM

Figure 6.

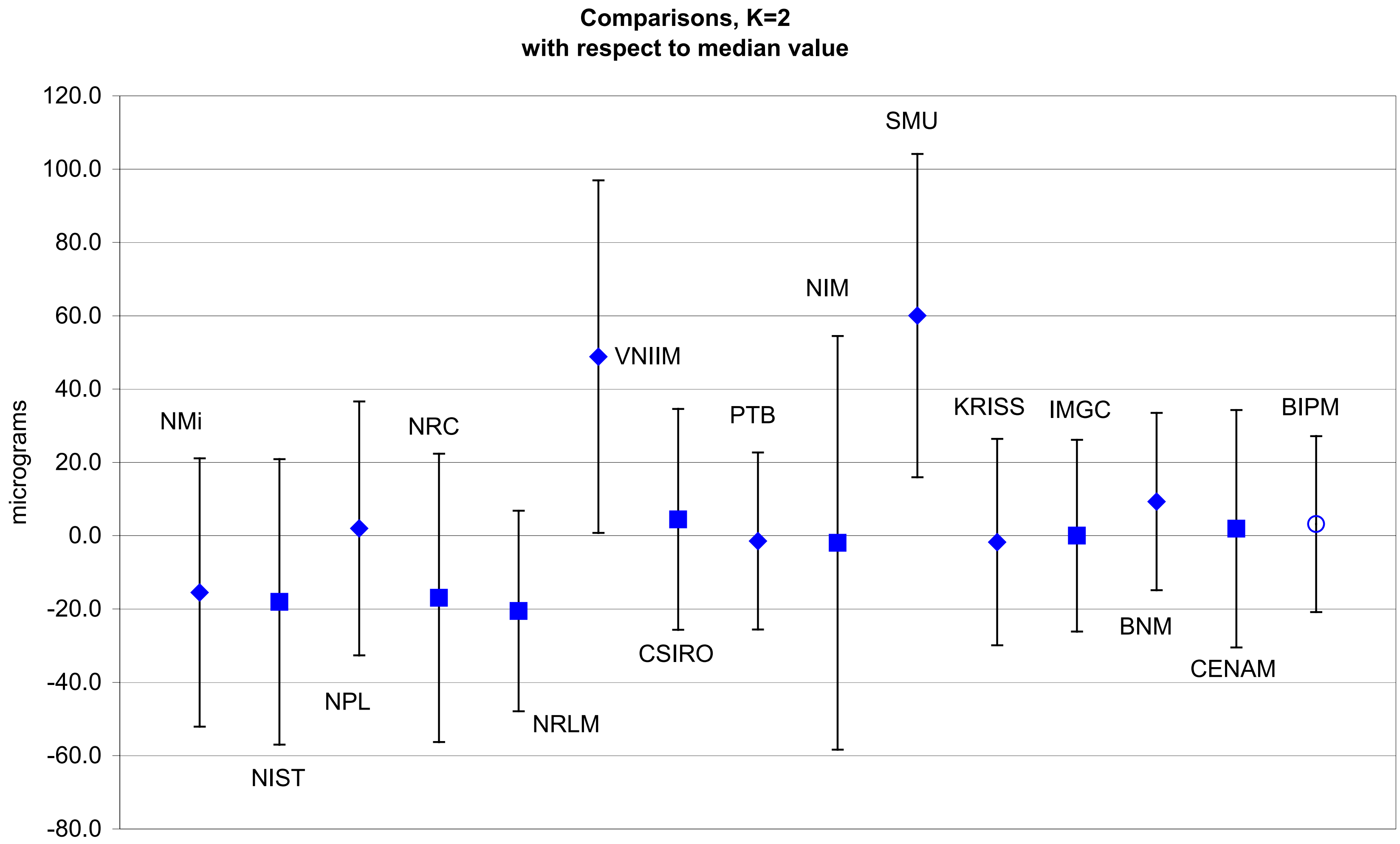

Comparisons, K=2
with respect to median value
micrograms
120.0
100.0
80.0
60.0
40.0
20.0
0.0
-20.0
-40.0
-60.0
-80.0
NMi
NIST
NPL
NRC
NRLM
VNIIM
CSIRO
PTB
NIM
SMU
KRISS
IMGC
BNM
CENAM
BIPM

Figure 7.

APPENDIX 1. Stability of the travelling standards

Taking Package 1 and Package 2 together, the travelling standards returned on ten different occasions to the BIPM. The median value of the ten measures of

$$0.5\cdot(m_{\mathrm{BI,n+1}} - m_{\mathrm{BI,n}})$$

is –1.8 µg. The estimated standard deviation of a single value about the median is 3.4 µg. We have already used Figures 3 and 4 to look for unusual changes in the travelling standards. We now confirm our conclusions by carrying out a test for "outliers", as described in reference [6]. The test is whether an individual datum lies more than 2.5 standard deviations from the median. (Of course, the appropriate standard deviation is that of a single measurement). From Figures 3 and 4, we had already noted a change in the travelling standards upon their return from SMU and from NIM. The outlier test just described gives a value of +2.6 standard deviations (just significant) for the return from SMU and +7.4 standard deviations (very significant) for the return from NIM. The remaining eight points lie within the range ±1.7 standard deviations (not significant).

From Table 1, the mass of VSL-1 (as measured by the BIPM) did not change significantly between measurements that most closely bracket the period in which Package 1 was sent to the PTB and the SMU. Thus the average change that was seen in the mass of Package 1 during this period may be attributed entirely to a change in J2. This view is consistent with the fact that vigorous brushing of VSL-1 had no effect on its mass. As mentioned above, it seems clear that the change to J2 occurred after it left the SMU and before it returned to the BIPM. We have therefore omitted the second BIPM value (the "return" value) in computing *diff*(PTB-BIPM) and *diff*(SMU-BIPM). The results obtained in this way are nearly identical to those obtained by using (1).

From Table 1, the mass of both VSL-2 and J3 increased during the period that Package 2 was at the NIM. The average mass of Package 2, as measured by the NIM, is close to the succeeding BIPM value. One might, therefore, be tempted to omit the preceding BIPM value for this reason alone. But this reasoning assumes *a priori* that agreement between the BIPM and the NIM is good. As mentioned above, the NIM has subsequently reported additional measurements of the difference in mass between VSL-2 and J3. These do tend to support the hypothesis that the change in the standard occurred before measurements at the NIM were begun. This hypothesis reduces the NIM result by almost 20 µg compared with a calculation based on (1). This difference, though large, is consistent with the combined estimated uncertainties of the two hypotheses.

Although the stability of the travelling standards was not perfect, we can conclude that only in the case of the NIM did this lead to a significant ambiguity in analyzing the results of these comparisons. It would, therefore, be useful to repeat a comparison with the NIM in the near future.

APPENDIX 2. Reference value and degrees of equivalence

The *Guidelines* of 1 March 1999 ask that discussion of the reference value of a key comparison and the degrees of equivalence among the participants be discussed in an appendix to Draft B.

Reference value of the comparisons

The *Guidelines* ask that a "reference value" be established for the comparisons. In the body of this report, the value of the BIPM (as pilot laboratory) has been taken as the provisional reference. However, since the major uncertainty budgeted by all laboratories taking part in the comparisons is due to the correction for air buoyancy determinations between Pt-Ir prototypes and stainless steel working standards, there is no reason to accept the BIPM result *a priori*. Therefore we have established a reference value based on the results of all participants. It will, however, be sufficient to calculate how much the BIPM differs from this reference. This is because previous calculations have taken the BIPM results as the provisional reference.

Note that the choice of a reference value will have no effect on Tables 3 and 4.

The overwhelming majority of participants have suggested that the median be used as the reference.

The median has been proposed as a "robust" estimator for reference values [5]. It has been computed from the 15 values under the column of Table 3 that is labelled "BIPM". The result of this calculation is as follows :

Median(Labs – BIPM) :	-3.2 μg (using all corrected results)
std. deviation of median [5] :	2.2 μg

The median is the line of abscissas of Figure 7.

Degree of Equivalence of laboratories with respect to the reference value

For each laboratory, this degree of equivalence is defined by the following two quantities: (1) the difference between the value obtained by the laboratory and the reference value and (2) the expanded uncertainty of the value given in (1) at a level of 95% confidence. These values are given in Table 5, where a coverage factor of k=2 has been used to estimate the confidence limit. Possible correlations between the measurements of each participant and the uncertainty of the median have not been taken into account. It is noted, however, that the variance of the median is small compared to the variances inferred from the last column of Table 5.

Degree of Equivalence of laboratories between two laboratories

This is defined by the following two quantities: (1) the difference in results obtained by the two laboratories and (2) the expanded uncertainty of the value given in (1) at a level of 95% confidence. These results have been given above in Tables 3 and 4.

The 95% confidence level has been approximated with a coverage factor of two (k=2) as mentioned in the body of the report.

Reference value appropriate to Appendix B of the MRA

The reference value as defined above was required in order to compute the degrees of equivalence. Once this has been accomplished, the actual mass of the travelling standards has no lasting metrological importance. The degrees of equivalence would, in principle have been the same if any other high-quality, 1-kg travelling standards in stainless steel had been used. For this reason, it is proposed to give the nominal value of 1 kg as the reference value for the purposes of Appendix B of the Key Comparison Database (KCDB).

	BIPM	NMi,c	NIST	NPL	NRC,c	NRLM,c	VNIIM	CSIRO	PTB	NIM	SMU	KRISS	IMGC	BNM	CENAM
BIPM		18	21	1	20	23	-46	-2	4	4	-57	4	3	-3	1
NMi,c	-18		3	-17	2	5	-64	-20	-14	-14	-75	-14	-15	-21	-17
NIST	-21	-3		-20	-1	2	-67	-23	-17	-17	-78	-17	-18	-24	-20
NPL	-1	17	20		19	23	-47	-2	4	4	-58	4	2	-4	1
NRC,c	-20	-2	1	-19		4	-66	-21	-15	-15	-77	-16	-17	-23	-19
NRLM,c	-23	-5	-2	-23	-4		-69	-25	-19	-19	-80	-19	-20	-26	-22
VNIIM	46	64	67	47	66	69		44	50	50	-11	50	49	43	47
CSIRO	2	20	23	2	21	25	-44		6	6	-55	6	5	-1	3
PTB	-4	14	17	-4	15	19	-50	-6		0	-61	0	-1	-7	-3
NIM	-4	14	17	-4	15	19	-50	-6	0		-61	0	-1	-7	-3
SMU	57	75	78	58	77	80	11	55	61	61		61	60	54	58
KRISS	-4	14	17	-4	16	19	-50	-6	0	0	-61		-1	-7	-3
IMGC	-3	15	18	-2	17	20	-49	-5	1	1	-60	1		-6	-2
BNM	3	21	24	4	23	26	-43	1	7	7	-54	7	6		4
CENAM	-1	17	20	-1	19	22	-47	-3	3	3	-58	3	2	-4	

Table 3. Average difference in assigned values (in micrograms) between laboratory A (left column) and laboratory B (top row). For example, from the grid shown above, the average mass determined by the NPL was 3.8 µg above the average mass determined by the KRISS.

	BIPM	NMi,c	NIST	NPL	NRC,c	NRLM,c	VNIIM	CSIRO	PTB	NIM	SMU	KRISS	IMGC	BNM	CENAM
BIPM		44	46	40	41	36	54	37	35	51	50	37	35	32	36
NMi,c	44		54	48	50	46	61	46	44	58	57	46	45	42	45
NIST	46	54		50	50	47	62	48	46	60	59	48	47	45	47
NPL	40	48	50		46	42	58	43	41	55	54	43	41	38	41
NRC,c	41	50	50	46		42	59	44	42	56	55	44	42	40	43
NRLM,c	36	46	47	42	42		56	39	37	53	52	39	38	35	38
VNIIM	54	61	62	58	59	56		56	54	66	65	56	55	53	55
CSIRO	37	46	48	43	44	39	56		38	54	52	40	39	36	39
PTB	35	44	46	41	42	37	54	38		52	51	38	36	33	37
NIM	51	58	60	55	56	53	66	54	52		63	54	53	50	53
SMU	50	57	59	54	55	52	65	52	51	63		52	51	49	52
KRISS	37	46	48	43	44	39	56	40	38	54	52		39	36	39
IMGC	35	45	47	41	42	38	55	39	36	53	51	39		34	38
BNM	32	42	45	38	40	35	53	36	33	50	49	36	34		34
CENAM	36	45	47	41	43	38	55	39	37	53	52	39	38	34	

Table 4. The combined standard uncertainties (k = 2) in micrograms for the corresponding values of Table 3. Since the BIPM was the pilot laboratory, the uncertainty in comparing two laboratories other than the BIPM necessarily includes a component due to measurements at the BIPM. However, in this case only that part of the BIPM uncertainty that is not systematic between the two other laboratories has been included. In the case of comparison between the BIPM and another laboratory, the total BIPM uncertainty has been included.

	diff /μg	*U* /μg
BIPM	3	24
NMi,c	-15	37
NIST	-18	39
NPL	2	32
NRC,c	-17	34
NRLM,c	-20	28
VNIIM	49	48
CSIRO	5	29
PTB	-1	26
NIM	-1	46
SMU	60	44
KRISS	-1	29
IMGC	0	27
BNM	6	22
CENAM	2	27

Table 5. Column two gives the difference from the reference value for each laboratory listed in column one. The expanded uncertainty (k=2) of this difference is given in column 3.

APPENDIX 3.

Differences in mass determination of travelling standards and corresponding uncertainties in the frame of interlaboratory comparisons comprising different loops

Michael Gläser
Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

This is a consideration of the differences in mass determination and the associated uncertainties, when a travelling standard circulates between the pilot laboratory P and laboratories A and B within the same loop, or when two travelling standards circulate in two independent loops comprising P and A in the first and P and B in the second loop. The following four cases are considered.

Case 1. Mass comparison between laboratory A and the pilot laboratory P.
P1 indicates the first, P2 the second measurement at P.

It is assumed, that the mass difference between A and P is the average $\Delta m_{\mathrm{A,P}}$ of the differences m_{A} -m_{P1} and m_{A} - m_{P2}.

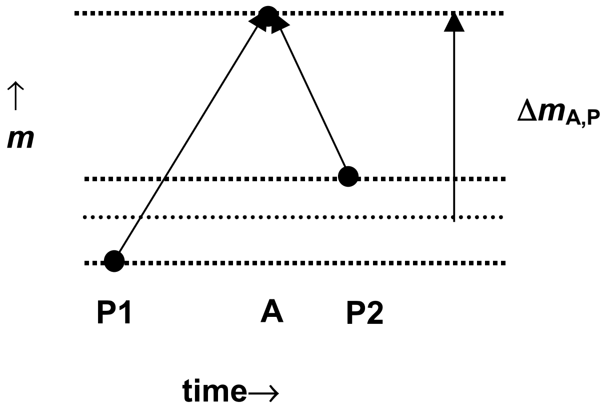


Figure A1. Mass determinations P1-A-P2.

Mass difference:

$$\Delta m_{\mathrm{A,P}} = m_{\mathrm{A}} - \left(\frac{m_{\mathrm{P1}} + m_{\mathrm{P2}}}{2} \right) \qquad \text{(A1)}$$

or (with the independent quantities m_{A}, m_{P1}, $\Delta m_{\mathrm{P1,P2}}$) :

$$\Delta m_{\mathrm{A,P}} = m_{\mathrm{A}} - m_{\mathrm{P1}} - \frac{\Delta m_{\mathrm{P2,P1}}}{2} \qquad \text{(A2)}$$

$$\text{with: } \frac{\Delta m_{\text{P2,P1}}}{2} = \frac{m_{\text{P2}} - m_{\text{P1}}}{2}$$

For the difference between the two pilot measurements, $\Delta m_{\text{P2,P1}}$, a rectangular probability distribution is assumed. The uncertainties of the two pilot measurements are assumed to be equal.

Variance:

$$u^2\left(\Delta m_{\text{A,P}}\right) = \frac{1}{3}\left(\frac{m_{\text{P2}} - m_{\text{P1}}}{2}\right)^2 + u^2\left(m_{\text{P}}\right) + u^2\left(m_{\text{A}}\right) \qquad \text{(A3)}$$

Case 2. Mass comparison between two laboratories A and B in two independent loops: P1-A-P2 and P3-B-P4

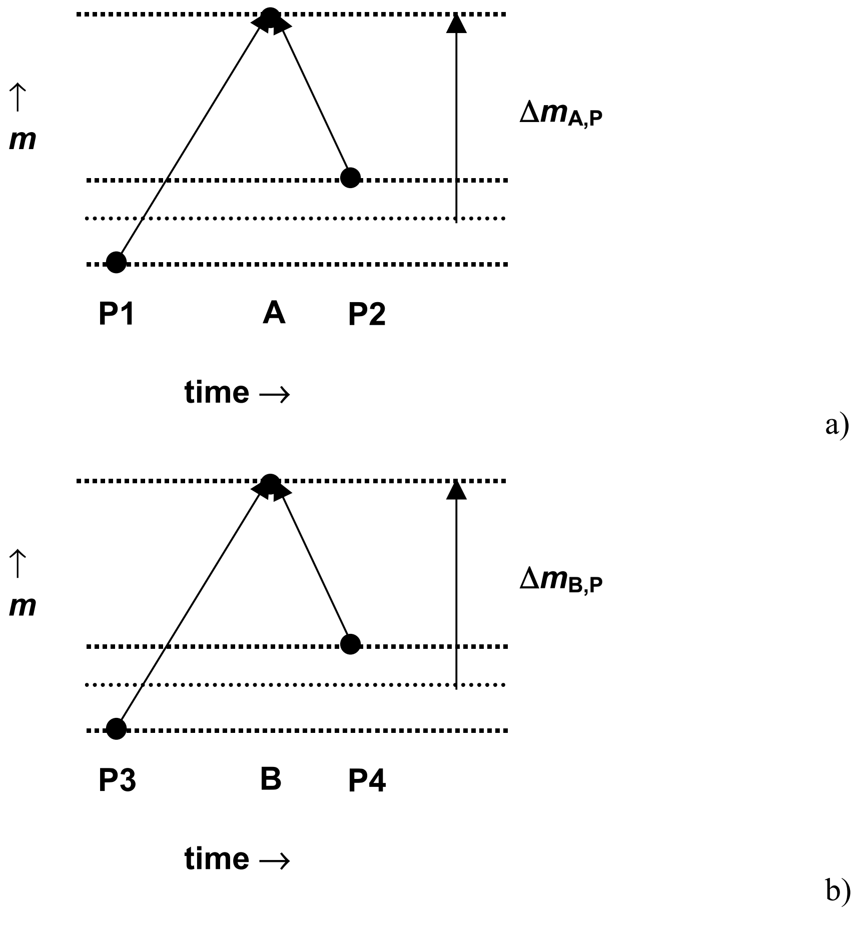


Figure A2. Mass determinations: a) P1-A-P2; b) P3-B-P4.

Mass difference:

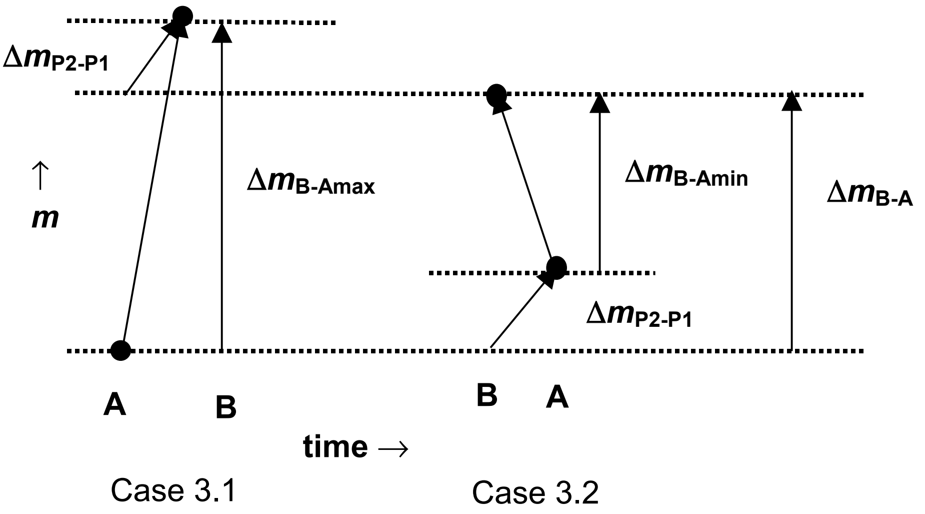
$$\Delta m_{\text{B-A}} = m_{\text{B}} - \left(\frac{m_{\text{P4}} + m_{\text{P3}}}{2}\right) - m_{\text{A}} + \left(\frac{m_{\text{P2}} + m_{\text{P1}}}{2}\right) \tag{A4}$$

Variance:

$$u^2\left(\Delta m_{\text{B-A}}\right) = \frac{1}{3}\left(\frac{m_{\text{P4}} - m_{\text{P3}}}{2}\right)^2 + \frac{1}{3}\left(\frac{m_{\text{P2}} - m_{\text{P1}}}{2}\right)^2 + 2u^2\left(m_{\text{P}}\right) + u^2\left(m_{\text{A}}\right) + u^2\left(m_{\text{B}}\right) \tag{A5}$$

Case 3. Mass comparison between two laboratories A and B within the same loop: P1-A-B-P2 (case 3.1) or P1-B-A-P2 (case 3.2). If A or B is the first laboratory, is not known.

It is assumed, that the difference between the measured values at A and B is the average between the maximum and minimum possible values due to the instability observed by P. In case 3.1, the maximum possible value is assumed if A measures the travelling standard at the same time as P1 and B at the same time as P2, in case 3.2 vice versa.
The absolute differences A-P and B-P have thus no influence on the difference between A and B.



$\Delta m_{\text{P2-P1}}$
↑
m
$\Delta m_{\text{B-Amax}}$
$\Delta m_{\text{B-Amin}}$
$\Delta m_{\text{B-A}}$
$\Delta m_{\text{P2-P1}}$
A
B
B
A
time →
Case 3.1
Case 3.2

Figure A3. Mass determinations P1-A-B-P2 (case 3.1) or P1-B-A-P2 (case 3.2).

Mass difference:

Case 3.1

$$\Delta m_{\text{B-A}} = m_{\text{B}} - m_{\text{A}} + \left(m_{\text{P2}} - m_{\text{P1}}\right) \tag{6}$$

Case 3.2

$$\Delta m_{\text{B-A}} = m_{\text{B}} - m_{\text{A}} - \left(m_{\text{P2}} - m_{\text{P1}}\right) \tag{7}$$

Average:

$$\Delta m_{\text{B-A}} = m_{\text{B}} - m_{\text{A}} \tag{8}$$

The uncertainty of the difference m_{P2} - m_{P1} is composed of the contribution from the unknown drift of the transfer standards (rectangular distribution) and of the contribution from the measurement uncertainty of P, $u(\Delta m_P)$. Because we don't know the order of A and B, the result has maximum drift deviations of ± (m_{P2} - m_{P1}).

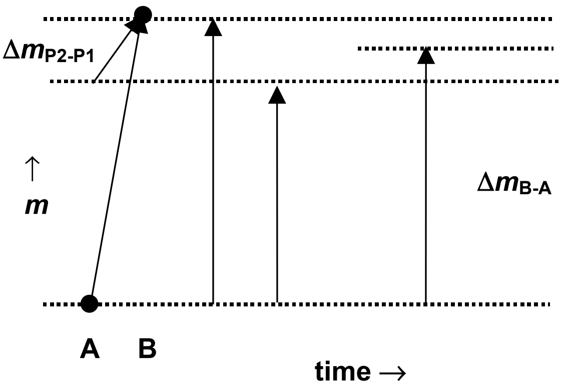
Variance:

$$u^2\left(\Delta m_{\text{B-A}}\right)=\frac{1}{3}\left(m_{\text{P2}}-m_{\text{P1}}\right)^2+u^2\left(m_{\text{A}}\right)+u^2\left(m_{\text{B}}\right)+u^2\left(\Delta m_{\text{P}}\right) \qquad (9)$$

Case 4.Mass comparison between two laboratories A and B within the same
loop: P1-A-B-P2. It is known, that A is the first and B the second laboratory.

It is assumed, that the difference between the measured values at A and B is the average between the maximum and minimum possible values due to the instability observed by P.
The absolute differences A-P and B-P have no influence on the difference between A and B.

Case 4.1



Case 4.2

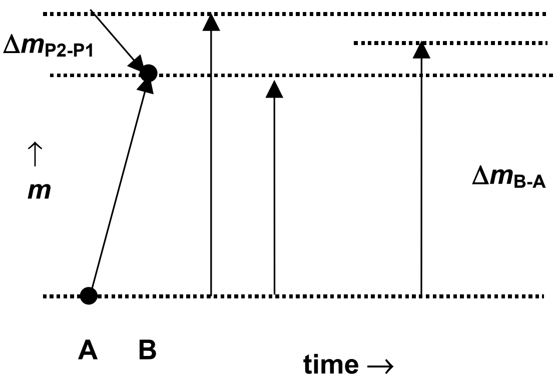


Figure A4. Mass determinations P1-A-B-P2; case 4.1: $m_{P2} > m_{P1}$; case 4.2: $m_{P1} > m_{P2}$.

Mass difference:

Case 4.1

$$\Delta m_{\text{B-A}} = m_{\text{B}} - m_{\text{A}} + \frac{m_{\text{P2}} - m_{\text{P1}}}{2} \qquad \text{(A10)}$$

Case 4.2

$$\Delta m_{\text{B-A}} = m_{\text{B}} - m_{\text{A}} - \frac{m_{\text{P2}} - m_{\text{P1}}}{2} \qquad \text{(A11)}$$

Average:

$$\Delta m_{\text{B-A}} = m_{\text{B}} - m_{\text{A}} \qquad \text{(A12)}$$

In contrast to case 3, here we know the order of A and B. Therefore the maximum drift deviation is only ± (m_{P2} - m_{P1})/2.

Variance:

$$u^2\left(\Delta m_{\text{B-A}}\right) = \frac{1}{3}\left(\frac{m_{\text{P2}} - m_{\text{P1}}}{2}\right)^2 + u^2\left(m_{\text{A}}\right) + u^2\left(m_{\text{B}}\right) + u^2\left(\Delta m_{\text{P}}\right) \qquad \text{(A13)}$$

APPENDIX 4. Basic model of the comparisons

In general, all laboratories make use of the following quantities, either directly or indirectly. All these quantities may be sources of uncertainty:

M_P : Mass of national prototype, P (in platinum-iridium)
M_S : Mass of stainless steel secondary standard, S
M_T : Mass of stainless steel transfer standard, T
ρ_a1 : Density of air during calibration of S
t_1 : Temperature in balance during calibration of S
ρ_a2 : Density of air during calibration of T
t_2 : Temperature in balance during calibration of T
V_P0 : Volume of P at 0 °C (given in BIPM certificate)
α_P0 : Volumetric coefficient of expansion of P, from 0 °C (given in BIPM certificate)
V_P20 : Volume of P at 20 °C
α_P : Volumetric coefficient of expansion of P, from 20 °C
V_S20 : Volume of S at 20 °C
α_S : Volumetric coefficient of expansion of S
V_T20 : Volumetric of T at 20 °C (provided to participants)
α_T : Volumetric coefficient of expansion of T (provided to participants)
h_P : Height of the centre of mass of P above its base (19.5 mm)
h_S : Height of the centre of mass of S above its base
h_T : Height of the centre of mass of T above its base (27.5 mm)
g' : Relative gravitational gradient in the balance: $(1/g)(\mathrm{d}g/\mathrm{d}h)$
Y_1 : Difference in balance scale readings between S and P, converted to mass
Y_2 : Difference in balance scale readings between T and S, converted to mass

In highly simplified form, the calibration of S in terms of P becomes

$$M_\mathrm{S} = M_\mathrm{P} + \rho_\mathrm{a1}(V_\mathrm{S20} - V_\mathrm{P20}) + \rho_\mathrm{a1}(t_1 - 20\ ^\circ\mathrm{C})(V_\mathrm{S20}\alpha_\mathrm{S} - V_\mathrm{P20}\alpha_\mathrm{P}) - g'(h_\mathrm{S} - h_\mathrm{P})(1\mathrm{kg}) + Y_1 \qquad (1)$$

In many laboratories, small mass standards are added to S and P in order that Y_1 be made small. In this case, eq. (1) is, of course, amended to take account of the added weights.

The calibration of T in terms of S is then given by eq. (2) in similar fashion:

$$M_\mathrm{T} = M_\mathrm{S} + \rho_\mathrm{a2}(V_\mathrm{T20} - V_\mathrm{S20}) + \rho_\mathrm{a2}(t_2 - 20\ ^\circ\mathrm{C})(V_\mathrm{T20}\alpha_\mathrm{T} - V_\mathrm{S20}\alpha_\mathrm{S}) - g'(h_\mathrm{T} - h_\mathrm{S})(1\mathrm{kg}) + Y_2 \qquad (2)$$

If M_S *has not changed* between the first and second calibrations, then the combined measurement model becomes :

$$\begin{aligned} M_\mathrm{T} = M_\mathrm{P} &+ \rho_\mathrm{a1}(V_\mathrm{S20} - V_\mathrm{P20}) + \rho_\mathrm{a2}(V_\mathrm{T20} - V_\mathrm{S20}) + \\ &\rho_\mathrm{a1}(t_1 - 20\ ^\circ\mathrm{C})(V_\mathrm{S20}\alpha_\mathrm{S} - V_\mathrm{P20}\alpha_\mathrm{P}) + \rho_\mathrm{a2}(t_2 - 20\ ^\circ\mathrm{C})(V_\mathrm{T20}\alpha_\mathrm{T} - V_\mathrm{S20}\alpha_\mathrm{S}) - \\ &g'(h_\mathrm{T} - h_\mathrm{P})(1\mathrm{kg}) + Y_1 + Y_2 \end{aligned} \qquad (3)$$