

Is “consensus value” a correct term for the product of pooling measurement results?

Paul De Bièvre

Accreditation and Quality Assurance
Journal for Quality, Comparability and
Reliability in Chemical Measurement

ISSN 0949-1775
Volume 17
Number 6

Accred Qual Assur (2012) 17:639-640
DOI 10.1007/s00769-012-0938-7



Your article is protected by copyright and all rights are held exclusively by Springer-Verlag Berlin Heidelberg. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Is “consensus value” a correct term for the product of pooling measurement results?

Paul De Bièvre

Published online: 7 November 2012
© Springer-Verlag Berlin Heidelberg 2012

We often pool measurement results (entry 2.9 in [1]) and expect somehow a more adequate value from such an operation, especially in two cases.

In the process leading to a certified reference material, CRM (entry 5.14 in [1]), we frequently pool various measurement results obtained by various measurement procedures (sometimes in various laboratories) expecting to come closer to a more adequate value for embodiment in the CRM than just the one value obtained by one measurement laboratory only. By a more adequate value, we mean a value that makes the CRM into a more reliable tool to assist in calibration or in the search of compensation for systematic effects (entry 2.17 in [1]) in a measurement.

In a Proficiency Testing Scheme (one of the various types of Interlaboratory Programme—see description of ILC types in Table 7.2–1 in [2]), we frequently pool various measurement results obtained by possibly various measurement procedures, expecting to establish a more adequate “reference value” (entry 5.18 in [1]) for the Scheme than just the one value obtained by one measurement laboratory only. Again, by more adequate, we mean a value that makes a reference value in a PTS into a more reliable tool to assist in the evaluation of the performance of the PTS participants, that is, of their “measurement capability” (concept 7.3–1 in [2]).

Thus, in both cases, we expect that pooling measurement results enables us to come closer to a more adequate value for its intended use.

But what is the thinking underlying the pooling of multiple measurement results?

There is this overwhelming belief that the greater the number of measurement results we pool, the better the resulting mean value. But, in an ILC with the purpose of arriving at a reliable value for certification, the goal is different from that in a PTS: detecting unknown systematic effects. Let us analyse the application of that thinking a little deeper.

First, one would think that the individual measurement results of a set of results obtained by different analysts by definition includes a measurement uncertainty (entry 2.26 in [1]). All of them are intended to assist somehow in the certification of a CRM and therefore ought to be reliable, in principle, within their respective stated measurement uncertainty. One would expect such results to be metrologically compatible (entry 2.47 in [1]) even if obtained through possibly different measurement procedures, reference measurement procedures, or even primary reference measurement procedures (entries 2.6–2.7–2.8 in [1]). One would even expect that these results are metrologically equivalent, that is, that they are “acceptable for the same specified intended use” (concept 5–4 in [2]). The question then arises whether we indeed need multiple results for the same measurand (entry 2.3 in [1]). As they should all be equivalent in the sense of the definition, we could logically conclude that one of these results—with its measurement uncertainty—is sufficient to attribute a trustworthy value to the measurand, thus making the other values superfluous.

So, what could be the point of having multiple results? The answer seems to be that these additional—if superfluous—values do offer confirmation of each other, thus making all

Disclaimer: The author is a member of the Joint Committee on Guides for Metrology (JCGM), Working Group 2 (VIM). The opinions expressed in this Column do not necessarily represent the view of the Working Group or of ACQUAL.

P. De Bièvre (✉)
Kasterlee, Belgium
e-mail: paul.de.bievre@skynet.be

participating analysts and their laboratories feel more satisfied, individually and as a group. Attaining unanimity on a value would be ideal, but probably difficult to reach. In such an important context as the certification of a value embodied in a CRM or the establishment of a reference value for a PTS, “consensus” would already be highly satisfactory. What then is “consensus”? It is defined as “absence of sustained opposition” [3] in ISO contexts where conclusions in the Technical Committees (TCs) must somehow be drawn, no matter what problem is discussed. Similarly, a discussion about a value to be embodied in a CRM or to be made into a reference value for a PTS is important. Arriving at it by such a “consensus” is very useful and the term “consensus value” looks adequate to name that value.

But the question may be asked whether a consensus value is, metrologically speaking, a measured value (entry in 2.10 [1]) since it is the product of a choice, that is, it is the consequence of a *decision* supposedly taken after “absence of sustained opposition”. Think for example about decisions during the pooling process about eliminating some values, usually after discussion, or about removing outliers from the set, both because metrological concepts were not respected during the measurements. The fact is that such a consensus increases trust in the certified value and that is an important psychological consideration, sometimes even an important political or commercial consideration. The increase in trust

seems to be the overriding and understandable drive towards attempting to achieve a “consensus value”. This raises another interesting and—as yet unanswered—question: if we want to pool measurement results, what are the criteria for doing so? When, that is under which conditions, can they be pooled? Another topic for continuing discussion.

We conclude that the process of pooling measurement results and concluding to one value is adequately described by the term “consensus value”, but that that is not a measured value.

As usual, any comment, question, or amendment is welcome, preferably as a contribution to the Discussion Forum of this Journal.

References

1. BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, OIML (2012) The international vocabulary of metrology—Basic and general concepts and associated terms (VIM), 3rd edn. JCGM 200. <http://www.bipm.org/vim>
2. De Bièvre P, Dybkaer R, Fajgelj A, Hibbert B (2011) Metrological traceability of measurement results in chemistry—concepts and implementation. *Pure Appl Chem* 83:1873–1935. <http://iupac.org/publications/pac/83/10/1873>
3. ISO-IEC Directives (2012) Part 1, Procedures for the technical Work, 9th edn. Section 2.5.6. <http://iso.org/> (accessed on 2012-09-30 and followed up from there)