EDITORIAL

Chemists' views on measurement results are influenced too much by statistical considerations and not enough by the application of simple metrological principles

Paul De Bièvre

Statistical considerations are often used to describe the quality of measurement results (entry 2.9 in [1]). Measured quantity values (entry 2.10 in [1]) are thereby treated as "samples" from homogeneous populations ("samples" is here used in the meaning the term has in statistics, not in the meaning it has in chemistry). Thus, the entire array of statistical evaluation procedures is available to perform evaluations of the "quality" of these results (see [2] for more on "quality of a measurement result"). However, in chemical measurement, many of these populations are not homogeneous because a "best" measurement result is believed to be attainable in the form of an average, mean, median, or otherwise, when different measurement procedures (entry 2.6 in [1]) and different analysts (entry 2.24 in [1]) are involved to measure the same measurand (entry 2.3 in [1]), thus making said population inherently inhomogeneous by definition. As can be learned from experience, the spread of measured quantity values (entry 2.10 in [1]) constituting these populations, frequently is not "normal". For a number of illustrations showing such "non-normaliy", see [3]. Yet, for the purpose of constructing a "best" result, they are "assumed" to be normal nevertheless.

Two observations jump to the mind:

First observation: sets of measured quantity values not compatible with the quoted assumption of "normal distribution" generate an almost irresistible trimming process to eliminate outliers, making the assumed distribution model prevail. Also, construction of criteria is attempted to eliminate "abnormal" or "outlying" measured values on the basis of statistical assumptions. Experimental data not compatible with a normal distribution are considered "outliers" and are put aside because they make evaluation of the measurement results in terms of assumed statistical models too difficult or outright impossible. Thus, the desire to come close(st) to a supposedly best value conflicts with the assumption of normal distribution.

Second observation: chemical measurements cannot be repeated frequently in practice because of cost or time restrictions (in chemistry, measurements can require a complex or lengthy chemical sample preparation); hence, a large number of results is not available nor even pursued. On the other side, a small number of measurement results may invalidate any statistical conclusion based by definition on large numbers. Hence, in practice, a problem arises in the form of a conflict between the requirement for a large number of measurement results with an understandable and defendable limitation of that number.

There is more. Chemistry is, by its very nature, the science of conversion of chemical compounds with given molecular structures and their associated chemical properties into other molecules with other molecular structures and their associated other chemical properties [4]. Such interactions do occur in almost all chemical measurements, and they must be described on the basis of good knowledge of the measurement procedure of them in the first place. That can be checked by asking the questions: which stable references have been used as reference for such transient phenomena? And, also: how do we then assess the "quality" of such a measurement result?

In any measurement or experiment, we want the measurement result to be of "high quality". So, we must also define "quality of a measurement result". Pending a widely discussed and accepted definition, we here define quality as "property of a measurement result consisting of exceeding the end user's expectation". To answer that question, we verify whether essential metrological principles have been identified and respected when setting up and carrying out

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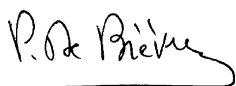e-mail: paul.de.bievre@skynet.be

the measurement. Has the achieved measurement result "metrological traceability" (entry 2.42 in [1])? Has it been evaluated against a preset "target measurement uncertainty" (entry 2.34 in [1]), the ultimate criterion for the fitness for intended use of the result? "Metrological traceability" provides us with the "trace" along which the result comes to us. Thereby, the "reference" for the metrological traceability is identified, either the measurement unit chosen, or (the value of) a "measurement standard" (entry 5.1 in [1]), or (the value obtained by) a "measurement procedure" (entry 2.6 in [1]), and these can be identified in the metrological traceability chain (entry 2.42 in [1]). The comparison of a full measurement uncertainty according to GUM with a preset target measurement uncertainty enables a simple evaluation of the quality of the measurement result in terms of the intended quality expressed as a fitness for the intended use of this result.

We submit that the above is sufficient to describe the quality of a measurement result.

We also note that we did not need statistics to arrive at a judgement of the quality of a measurement result. It could be done by applying metrological principles.

Are chemists indeed too much influenced by statistics and not enough by the use of simple metrological principles?

As usual, comments are invited.

Paul De Bièvre
Editor-in-Chief

By the way, looking for justification of metrology in chemistry?

The establishment of a consistent reference framework of concepts and associated terms needed to communicate reliably our measurement results in chemistry across continents (in order to believe "tested once, accepted everywhere").

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, The international vocabulary of metrology—basic and general concepts and associated terms (VIM), JCGM 200:2008, www.bipm.org/vim
2. De Bièvre P (2006) On quality of a measurement result. Accred Qual Assur 11:597–598
3. De Bièvre P (2009) Comparing results of chemical measurements: time to raise some basic questions from practice, Chap 8. In: Pavese F, Forbes AB (eds) Data modelling for metrology and testing in measurement science. Birkhauser, Boston, pp 255–273. ISBN 978-0-8176-4804-6. doi:10.1007/978-0-8176-4804-6
4. De Bièvre P (2011) Integer numbers and their ratios are key concepts in describing the interactions of atoms and molecules. Accred Qual Assur 16:117–120