

## Solution to the data averaging challenge

Juris Meija<sup>1</sup>

Accepted: 4 May 2022 / Published online: 5 July 2022  
© Crown copyright in right of Canada 2022

*The winner of the data averaging challenge (published in volume 414 issue 1) is:*

*Cristhian Paredes, Instituto Nacional de Metrología de Colombia, Bogotá, D.C., Colombia.*

*The award entitles the winner to select a Springer book of his choice up to a value of €100,-.*

*Our Congratulations!*

Summarizing the data, by averaging them somehow, is one of the fundamental tasks of data analysis. The task of this Challenge [1] seems trivial at first, but consider this: the solutions received from the readers *all* had a different answer for the average atomic weight of tellurium, ranging from 126.3 to 126.6. Moreover, the uncertainty associated with this average differed by more than an order of magnitude, ranging from 0.01 to 0.20!

Clearly, this topic is not trivial and requires making numerous modelling assumptions. Are the data consistent with each other? Are all data reliable? Are the provided measurement uncertainties reliable? The answers to these questions will determine the appropriate statistical model to reduce these data. For example, in the case of consistent data with reliable uncertainties, one could adopt the uncertainty weighted mean,  $\mu$ , given by the following statistical model:

$$A_i(\text{Te}) = \mu + e_i \quad (i = 1 \dots 6),$$

where  $e_i$  are the measurement errors modelled as random draws from Gaussian distribution with zero mean and standard deviations  $u(A_i)$ . This simple model was indeed used by Clarke himself who reported  $\mu = 126.523$  with the associated standard uncertainty  $u(\mu) = 0.014$  [2]. But the tellurium data are not consistent! The results from methods #1 and #3 are more than 20 standard uncertainties apart from one other

(Fig. 1) and such a discrepancy cannot be explained by the provided measurement uncertainties alone.

Given the large discrepancies observed between these results, three general approaches can be taken. First, one can declare the provided uncertainties as unreliable and proceed without them by calculating the simple arithmetic average. Second, one can declare some of the data as unreliable and proceed by finding the largest consistent subset using a chi-squared test to assess consistency [3]. Third, one can still proceed with all the data but using a statistical model that provides allowance for errors in the individual results [4]. This approach is formally known as the random effects model, given in its most simple form as

$$A_i(\text{Te}) = \mu + \lambda_i + e_i \quad (i = 1 \dots 6),$$

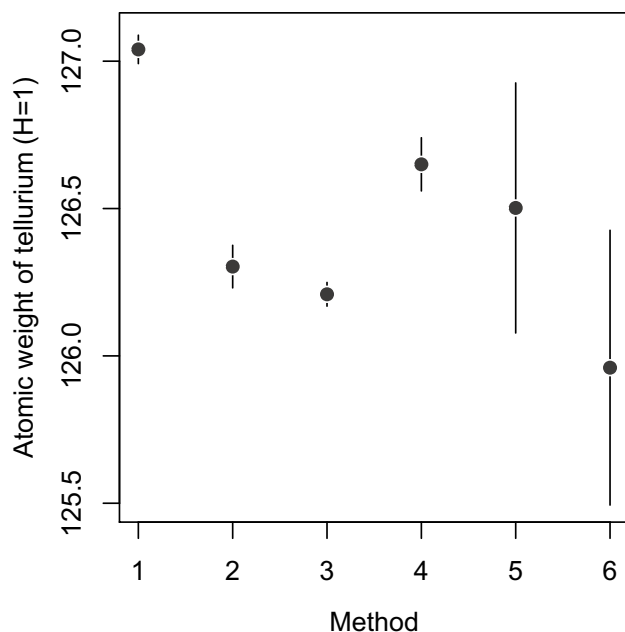
where the additional variable  $\lambda_i$  describes the laboratory effects which can be modelled as random draws from Gaussian distribution with zero mean and unknown standard deviation  $\tau$ . One of the most popular solutions of this statistical model gives  $\mu = 126.46$  with the associated standard uncertainty  $u(\mu) = 0.19$  [5].

These are, of course, by no means the only choices. Even with the same statistical model at hand, one can adopt a variety of classical or Bayesian methods to fit the model to the data [6]. And there are many other statistical models developed for this simple task [7–10], with one of the most recent methods inspired from the Darwinian theory of evolution [11]. Some of these methods are daunting in their complexity but in this particular example, dismissing the provided measurement uncertainties might very well be the best way to reduce this dataset: the simplest of all averages, the arithmetic mean, is 126.44 with the associated standard uncertainty 0.15.

This article is the solution to the Analytical Challenge to be found at <http://dx.doi.org/10.1007/s00216-021-03748-z>

✉ Juris Meija  
juris.meija@nrc-cnrc.gc.ca

<sup>1</sup> National Research Council Canada, Ottawa, ON K1A 0R6, Canada



**Fig. 1** Summary of atomic weight determinations of tellurium from Clarke's 1897 edition of *Recalculation of the Atomic Weights* with  $\pm 2$  standard uncertainty error bars [2]

## Declarations

**Conflict of interest** The author declares no competing interests.

## References

1. Meija J. Data averaging challenge. *Anal Bioanal Chem.* 2022;414:29–30.
2. Clarke FW. A Recalculation of the atomic weights. Washington DC: Smithsonian Institution; 1897. <https://archive.org/details/recalculationofa00clarrich>. Accessed 10 May 2022.
3. Cox MG. The evaluation of key comparison data: determining the largest consistent subset. *Metrologia.* 2007;44:187–200.
4. Toman B, Possolo A. Laboratory effects models for interlaboratory comparisons. *Accred Qual Assur.* 2009;14:553–63.
5. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7:177–88.
6. Koepke A, Lafarge T, Possolo A, Toman B. Consensus building for interlaboratory studies, key comparisons, and meta-analysis. *Metrologia.* 2017;54:S34–62.
7. Merktas C, Toman B, Possolo A, Schlamminger S. Shades of dark uncertainty and consensus value for the Newtonian constant of gravitation. *Metrologia.* 2019;56: 054001.
8. Huang H. A new method for estimating consensus values in interlaboratory comparisons. *Metrologia.* 2018;55:106–13.
9. Mana G. Interlaboratory consensus building. *Metrologia.* 2021;58: 055002.
10. Pomme S, Keightley J. Determination of a reference value and its uncertainty through a power-moderated mean. *Metrologia.* 2015;52:S200–12.
11. Coulon R, Judge S. An evolutionary algorithm for consensus building in interlaboratory comparisons. *Metrologia.* 2021;58: 065007.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.