Response to Mark A. Mackay and Tony C. Badrick: on the holy grail of clinical chemistry

Wytze P. Oosterhuis

Department of Clinical Chemistry and Haematology, Zuyderland Medical Center, Heerlen, The Netherlands
Correspondence to: Wytze P. Oosterhuis. Department of Clinical Chemistry and Haematology, Zuyderland Medical Center, Henri Dunantstraat 5, 6419 PC Heerlen, The Netherlands. Email: w.oosterhuis@zuyderland.nl.
Response to: Mackay MA, Badrick TC. Response to Wytze P. Oosterhuis—analytical performance specifications in clinical chemistry: the holy grail? J Lab Precis Med 2019;4:24.

After the Milan Strategic conference in 2014 a consensus statement was published on analytical performance specifications (APS), revising the Stockholm consensus of 1999. One of the task and finish groups (T&F) groups that were started was one on the total error (TE) concept. Several issues did need to be resolved concerning TE and the measurement uncertainty (MU) models. Most importantly were flaws in the model for calculating the permissible (or allowable) TE. These models are based upon biological variation to derive APS, the second model of the Milan consensus.

The concept of TE is closely connected with the work of Westgard. It represents the combined effect of the random and systematic errors (bias) of the method, and the TE is compared to a permissible TE.

The concept of bias is complicated, both in the estimation of the value, and in the integration of bias in a mathematical model. Bias is commonly excluded from the MU model, where only the uncertainty of the (estimation of) bias is taken into account. There is no general consensus on these issues, and challenges remain to integrate the different concepts, both for the definition of performance, of APS as for quality control (QC) procedures.

In my paper, the "holy grail" referred to the APS and the fact that there is no general consensus how to define or derive these specifications based on biological variation (1). The T&F group did not solve all the issues, but I hope it brought some problems to light. However, consensus has been reached (2). The two important results were,

that it was agreed that the current calculation leads to an overestimation of the permissible TE based on biological variation. Two maximum permissible errors (bias and imprecision) are added, derived under the mutually exclusive assumptions of zero bias and zero imprecision, respectively. Secondly, the TE model needs a standard to define the error, as is the case in quality QC were a reference standard (or method mean) are available. In patients, however, the MU model applies, as a direct reference standard is lacking here (this is to be derived from the traceability chain) and only the uncertainty of the test result—but not the error—can be estimated.

Concerning the bias concept, it can be a valuable distinction between systematic method bias that can be corrected (e.g., between methods), and drift as a form of "bias" that, due to its variable nature, is included as component of analytical variation. However, without the definition of the period of estimating a QC mean and analytical variation (SDa) this distinction will be in part arbitrary.

There is no disagreement that APS and QC limits have different levels. However, APS is not a "half-way point" between clinical purpose and inadequate assays. They are meant to be the actual QC limit for clinical purpose. Quality limits in QC procedures should be stricter that the APS limit: with QC limits equal to the APS, and with a bias equal to this APS, 50% of QC results will still be within this limit due to random error. This level of error detection is obviously too low.

We agree that the within-patient variation CV_i is better to derive APS than the combined within-patient and group variation, as most tests are used for monitoring, where only CV_i is in play. The authors assume an imprecision of $CV_i/4$, a quite strict quality goal not achievable for many measurands. For sodium with $CV_i =0.5\%$ (3), the optimum imprecision is 0.125%. How will that be achieved? The authors illustrated the low assay quality in their study, with about 50% of the routine tests in the range below $0.25CV_i$ (4). We derived APS based the Number of Distinct Categories (NDC), a concept used in industry (5). This resulted in the APS of $CV_i/3$.

From their paper, it is not understood which QC rules were applied with sufficient error detection. It is still open for debate whether strict QC methods will lead to test improvement. More frequent QC with the concomitant higher frequency of calibration will introduce another source of uncertainty.

The issue of the Six Sigma based QC might not be understood correctly. This has been subject to debate in the T&F group. The Six Sigma model includes quality specifications derived from clinical specifications (or the surrogate: biological variation). In cases where biological variation is not small with a high sigma metric (e.g., 6), this results in relaxed quality limits. However, from a purely technical perspective, the measurement procedure could be outside the $\pm3\ SD_a$ QC limits and be out of control, while still within the clinical specifications. The question here is: do we want to know or care? Some argue that as long as the test is within the clinical specifications, we don't need to know. Others argue that a test that is out of its technical limits always needs attention.

An important difference between the TE and MU concepts seems to be overlooked by the authors. In the TE-model, as applied in QC, the error is estimated relative to the standard. In most cases an external QC result is used to estimate the bias. In the MU model, "true value" concept has been abandoned. The value of a test result is traceable through a chain of standards and reference methods to the highest standard, with each step adding an uncertainty component to the patient test result. The two different situations, QC and patient testing, are reflected in different models, TE and MU.

The holy grail is something that we keep searching, without ever reaching our goal. The TE model in the form that it is commonly applied has several flaws. There is no consensus what alternative model could provide performance specifications based on biological variation, QC procedures and an estimate of the MU, while correctly dealing with the concepts of bias and imprecision. The authors have presented interesting methods. However, without a general consensus our quest has not ended.

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Laboratory and Precision Medicine*. The article did not undergo external peer review.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jlpm.2019.06.02). The author has no conflicts of interest to declare.

Ethical Statement: The author is accountable for all aspects of the work in ensuring that questions to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Oosterhuis WP. Analytical performance specifications in clinical chemistry: the holy grail? J Lab Precis Med 2017;2:78.
2. Oosterhuis WP, Bayat H, Armbruster D, et al. The use of error and uncertainty methods in the medical laboratory. Clin Chem Lab Med 2018;56:209-19.
3. EFLM Biological variation database. Available online: http://www.biologicalvariation.eu/
4. Mackay M, Hegedus G, Badrick T. A simple matrix of

analytical performance to identify assays that risk patients using External Quality Assurance Program data. Clin Biochem 2016;49:596-600.

5. Oosterhuis WP, Severens MJ. Performance specifications and Six Sigma theory: Clinical chemistry and industry compared. Clin Biochem 2018;57:12-7.