

Do we really need to account for run bias when producing analytical results with stated uncertainty? - Reply

Author/Contributor:

O'Donnell, Gregory; Hibbert, D. Brynn

Publication details:

Analyst

v. 132

Chapter No. 12

pp. 1275-1277

0003-2654 (ISSN)

Publication Date:

2007

Publisher DOI:

<http://dx.doi.org/10.1039/b711403f>

License:

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/39126> in <https://unsworks.unsw.edu.au> on 2023-09-27

Reply to “Comment”: Do we really need to account for run bias when producing analytical results with stated uncertainty?

GREGORY E. O’DONNELL ^{1,2} AND D. BRYNN HIBBERT ²

1. TestSafe Australia, Chemical Analysis Branch, 5a Pioneer Ave, Thornleigh, NSW, 2120 AUSTRALIA
Fax: +61 2 9980 6849; Tel: +61 2 9473 4005;
E-mail: greg.odonnell@workcover.nsw.gov.au
2. School of Chemistry, University of New South Wales, Sydney, NSW, 2052 AUSTRALIA
Fax: +61 2 9385 6141; Tel: +61 2 9385 4713;
E-mail: b.hibbert@unsw.edu.au

Do we really need to account for run bias when producing analytical results with stated uncertainty? The answer is definitely in the affirmative. This is a reply to the “Comment” by Kadis¹ on the paper “Treatment of bias in estimating measurement uncertainty” by G. E. O’Donnell and D. B. Hibbert².

To quote from the Guide to Expression of Uncertainty in Measurement³ :

It is assumed that the result of a measurement has been corrected for all recognized significant systematic effects and every effort has been made to identify such effects (GUM 3.2.4)

An analyst can proclaim that a measurement result has no bias only if the bias has been measured and shown to be insignificant compared with the measurement uncertainty. How rigorously the check for bias has been performed contributes to the uncertainty of the test result. The more precisely this has been performed, the smaller the contribution to the uncertainty. This type of check is usually stipulated by accreditation bodies⁴ and is typically normal procedure in standard reference methods in the form of a recovery check within each analytical batch run. The need for a check for bias comes from the obligation to demonstrate that the analysis has no bias in that particular batch run. Verification that a current batch run is fit for purpose can only be achieved by such a measurement. Demonstration of the absence of bias in previous batch runs does not vindicate the current run.

The use of a matrix Certified Reference Material (CRM) for calibration is not generally encouraged (see 7.4.3 of the EURACHEM/CITAC Guide on Traceability in Chemical Measurement⁵). However, if a matrix CRM were used for calibration then it could be argued that a check for bias is no longer needed. However, the only way to proclaim an unbiased test result is to demonstrate it, and it would still be necessary to obtain an independent reference material from a secondary source to be used as the bias check.

The use of a CRM to measure bias maintains the metrological traceability of the corrected measurement result, and the uncertainty of the quantity value carried by the CRM is usually small. We acknowledge that CRMs are not always available, and when they are available very few are perfectly matrix matched. We also acknowledge the high cost of some CRMs that prohibits their regular use. If other standards from reputable suppliers are used for the bias check, the uncertainty interval is enlarged with lower degrees of freedom, and some work must be done to establish metrological traceability.

The argument that including the check for bias in the uncertainty budget is double counting the random component of uncertainty is nonsense. Measurement uncertainty comprises of the precision of the analysis of the sample and the uncertainty of the assessment of trueness (i.e., measurement of bias). Just as the precision of the analysis depends on the number of replicates of the sample analysed, then the precision of the assessment of trueness is dependent on the number of replicates of the trueness standard analysed. Therefore, the precision of the analysis is included in the uncertainty budget and the precision of the assessment of trueness needs to be included also. Consider a measurement equation for a measurand X , with measured quantity value x .

$$x = x(\text{initial}) + \delta \quad (1)$$

where $x(\text{initial})$ is the initially estimated quantity (the term coined by Dybkaer⁶) and δ is the bias. The measurement uncertainty of the value x is the combination of the uncertainties of $x(\text{initial})$ and δ . The fact that both quantity values are estimated by a measurement does not necessarily mean there is double counting. It might be that some common aspects of the measurements need to be considered as discussed by Hibbert⁷.

The definition of bias used in the paper² was taken from ISO 3534-1⁸ and was clearly defined in the paper as

Bias: *The difference between the expectation of the test results and an accepted reference value.*

The definition of laboratory bias was also taken from ISO 3534-1 defined as

Laboratory bias: *The difference between the expectation of the test results from a particular laboratory and an accepted reference value.*

Method bias was taken from ISO-5725-1⁹ and was defined as

Method bias: *The difference between the expectation of test results obtained from all laboratories using that method and an accepted reference value.*

Then it follows that run bias is the difference between the expectation of the test results from a particular run and an accepted reference value. These definitions were all given in the paper² and clearly illustrated in Figure 1 of the paper². However, there seems to be further confusion of the difference between bias and the components of bias. Figure 1 illustrates the difference between run bias and the run component of bias and laboratory bias and the laboratory component of bias. The Analytical Methods Committee¹⁰ and Thompson¹¹ introduced the idea of the “ladder of errors” where the uncertainty of a single analysis can be obtained from the following components of the repeatability, run bias, laboratory bias and the method bias as shown in the following equation.

$$\text{Result} = \text{true value} + \text{method bias} + \text{laboratory bias} + \text{run bias} + \text{repeatability error} \quad (2)$$

They further state that the corresponding combined uncertainty of the measurement is obtained from the addition of the above components in quadrature

$$u_c = \sqrt{\sigma_{\epsilon}^2 + \sigma_{\text{Run}}^2 + \sigma_{\text{Lab}}^2 + u_b^2} \quad (3)$$

where σ_{ϵ}^2 is the variance of the repeatability, σ_{Run}^2 is the between run within laboratory variance of bias, σ_{Lab}^2 is the between laboratory within method variance of the bias, and u_b^2 is the variance of the standard uncertainty of the method bias which includes the between method variance and the uncertainty of the reference material. However, the bias terms used in equation 2 and the corresponding variances in equation 3 do not comply to the definitions set out in ISO 3534-1 or ISO 5725-1. For example, the definition of laboratory bias states that it relates to a “particular laboratory” clearly indicating that laboratory bias is the bias of a single laboratory and not the average of number of laboratories. Therefore, the variance of laboratory bias should be the variance of the bias within a laboratory and not the variance between laboratories. What is represented in equation 2 and 3 are the components of the overall bias that is attributed to the run, the laboratory and the method effects. The following diagram illustrates the differences between the components of bias and the actual bias.

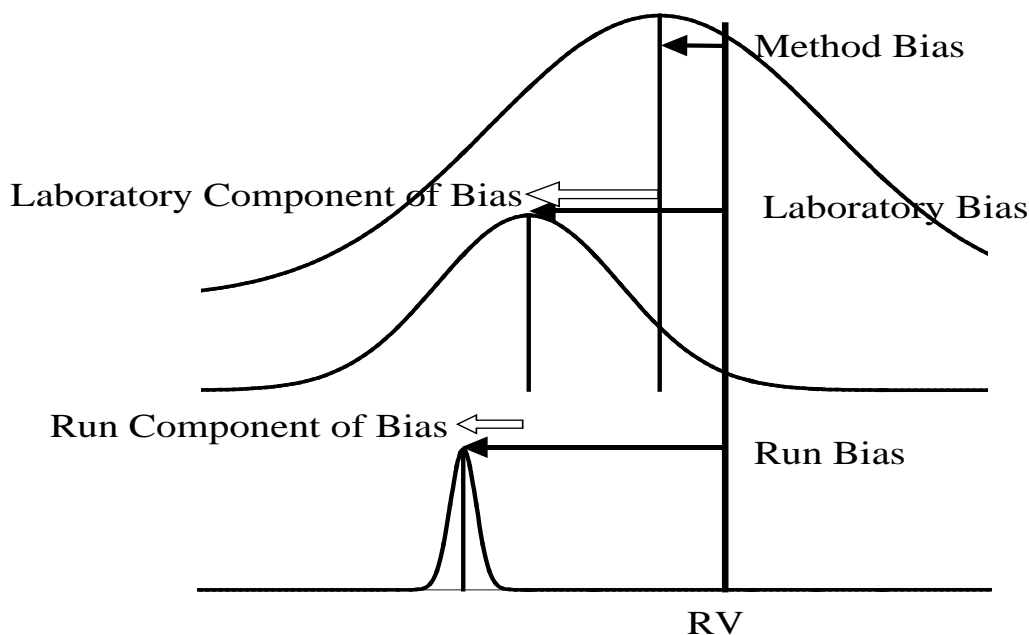


Figure 1. The distributions represent the spread of test results from within a run, a laboratory and a method in ascending order. The definitions of bias according to ISO 3534-1⁸. Run bias, laboratory bias and method bias are shown as black arrows. The run component of bias and the laboratory component of bias are shown as white block arrows. RV is the reference value.

The variances of these components of bias can be extracted from an interlaboratory trial followed by ANOVA analysis, for the use of combining back together again to get an estimate of the combined uncertainty under reproducibility conditions. One should ask if this procedure is really necessary? The laboratory component of bias is defined in ISO 5725-1 as the difference between the laboratory bias and the bias of the measurement method. For one to combine the variances of the components of bias in this manner one needs to be sure that the components are truly independent variables.

When the matrix of a sample interferes with the analysis it is obviously very important to matrix match both the calibration standards and the standards used for the bias check. This was stated in the paper² and is acknowledged by the authors again.

The questions that are really being asked by Kadis¹ are “Is it correct to estimate measurement uncertainty based on the repeatability of test results?” and “Shouldn’t it be estimated by intermediate precision or reproducibility?”

Estimating the random component of measurement uncertainty by repeatability standard deviation is consistent with ideas outlined in the “Guide to Expression of Uncertainty in Measurement” (GUM)³ described in Annexe D. More recently the uncertainty approach that underpins the GUM, and the forthcoming edition of the International Vocabulary of Metrology (VIM), has been explained by Ehrlich et al.¹² Briefly, the uncertainty of a measurement comprises the dispersion of the observations and the uncertainty of the applied correction for all recognised systematic effects. All other effects are unknowable.

The GUM states in its introduction that the ideal method for evaluating measurement uncertainty should be universal to all kinds of measurements. It says that the actual uncertainty quantity should be internally consistent. That is to say that the quantity value

... should be directly derivable from the components that contribute to it, as well as independent of how these components are grouped and of the decomposition of the components into subcomponents. (GUM 0.4)

and the value should be transferable

...to use directly the uncertainty evaluated for one result as a component in evaluating the uncertainty of another...(GUM 0.4)

The GUM explains in Annex E that what is presented in the GUM is

... a realistic rather than a safe value of uncertainty based on the concept that there is no inherent difference between an uncertainty component arising from a random effect and one arising from a correction for a systematic effect. (GUM E.1.1)

It goes on to point out that in the past there was a commonly held belief that uncertainty should be “safe” or “conservative” and it was often made deliberately large. Furthermore, that the random and systematic components were often reported separately and if they were combined were often reported in such a manner to satisfy this safety requirement. Often the result of a measurement is used as an input quantity value for a subsequent measurement. When this is the case, for the uncertainty estimate to be transferable it should be the best estimate and not a safe one. We would hold that conservative, or ‘worst-case scenario’ approaches to uncertainty of measurement results should be applied to the interpretation and use of the results, not to the expression of the measurement uncertainty.

For measurement uncertainty based on intermediate precision the uncertainty estimate is calculated based on the between batch run variance within a laboratory. Often this is calculated by the repeated analysis of a quality control or standard material in a number of batch runs over an extended period of time to obtain a reasonable estimate of the intermediate precision. Hence, intermediate precision is an estimate of the combination of the within-run (repeatability) and between-batch variance. The significance of bias is often then determined by comparison to this variance. Usually the within-run (repeatability) variance is usually much smaller than the between-batch

variance. The between-batch run variance contains the uncertainty of the laboratory bias, so to test the bias of a particular run for significance against the between-batch variance is comparing the bias to its own distribution. Therefore, in the long run the bias would only ever be deemed significant 5% of the time. What is being compared here is not the significance of bias in comparison to the precision of the analysis, but the measured bias against the expected value of zero for that particular analysis. When estimates of intermediate precision are used the bias could vary substantially from run to run but still be considered non-significant. In practice, industry groups establish control limits for the significance of bias and are typically in the range of 70 – 120% for a level of analyte 0.01 to 0.1 mg/kg, as in the food sector¹³. Many industry groups often stipulate that recovery data be reported on a test report separately from the test result. Clearly, this approach is not internally consistent or transferable, but is ostensibly designed to provide a safe conservative limit.

An estimation of measurement uncertainty based on reproducibility is obtained from the variances of all laboratories that participate in an interlaboratory study or method validation program. This type of measurement uncertainty estimation is essentially what GUM refers to as the “maximum error bound”. GUM makes the statement concerning this type of approach as follows:

In particular, if the “maximum error bound” (the largest conceivable deviation from the putative best estimate) is used in equation (E3), the resulting uncertainty will have an ill-defined meaning and will be unusable by anyone wishing to incorporate it into subsequent calculations of the uncertainties of other quantities. (GUM E.4.1)

The distribution that is generated by this method arises from the between-laboratory biases, the use of which is contrary to the GUM approach. Any bias estimated in a particular run would rarely be considered significant in comparison to this distribution. Indeed, the argument for the use of reproducibility from an interlaboratory study as measurement uncertainty is that it includes every possible systematic effect randomized in a single Type A quantity. By offering a single measurement uncertainty for all measurements made with a particular method, this approach equates the best laboratory with the worst irrespective of the different QC controls in place or even the different number of replicates that could be performed in each laboratory.

Comparability of test results is surely better achieved by having unbiased test results rather than by merely increasing the uncertainty interval to take account of putative, but unknown bias. This can be achieved by correcting an initial test result with the measured run bias. Realistic uncertainty intervals are what laboratory clients actually require to be able to make truly informed decisions.

REFERENCES

- 1 R. Kadis, *Analyst*, 2007, DOI: 10.1039/b703615a.
- 2 G. E. O'Donnell, D. B. Hibbert, *Analyst* **2005**, 130, 721.
- 3 ISO, BIPM, IEC, IFCC, IUPAC, IUPAP, OIML, *Guide to the Expression of Uncertainty in Measurement*, 1st ed., ISO, Geneva, Switzerland, **1993**.
- 4 National Association of Testing Authorities (NATA), *Technical note 23*, Sydney, Australia, **2006**.

- 5 EURACHEM, CITAC, *Traceability in Chemical Measurement*, 1st ed., EURACHEM / CITAC, Teddington, UK, **2003**.
- 6 R. Dybkaer, *Accred. Qual. Assur.* **2005**, *10*, 302.
- 7 D. B. Hibbert, *Accred. Qual. Assur.* **2003**, *8*, 195.
- 8 ISO 3534-1, *Statistics - Vocabulary and symbols - Part 1:*, 1st ed., ISO, Geneva, Switzerland, **1993**.
- 9 ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions*, 1st ed., ISO, Geneva, Switzerland, **1998**.
- 10 Analytical Methods Committee, *Analyst*, **1995**, *120*, 2303.
- 11 M. Thompson, *Analyst*, **2000**, *125*, 2020.
- 12 C. Ehrlich, R. Dybkaer, W. Woger, *Accred. Qual. Assur.* **2007**, *12*, 201.
- 13 A. Ambrus, *Accred. Qual. Assur.* **2004**, *9*, 288.