MEASUREMENT SCIENCE REVIEW, **19**, (2019), No. 5, 204–208

sciendo

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: https://content.sciendo.com

About not Correcting for Systematic Effects

Katy Klauenberg, Gerd Wübbeler, and Clemens Elster

Physikalisch-Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin, Germany, Katy.Klauenberg@Ptb.de

In practice, measurement results are sometimes described by an estimate, which is not the best one as defined in the *GUM*. Such alternative estimates arise when the result of a measurement is not corrected for all systematic effects. No recommendation exists in the *GUM* for associating an uncertainty with an uncorrected estimate.
A common choice in guidelines and in the literature is the uncertainty $u(y') = \sqrt{u^2(y) + (y - y')^2}$ for an alternative estimate y'. It arises from the expected quadratic loss, on which, also in the *GUM*, the standard uncertainty $u(y)$, and the best estimate y are based. However, such an uncertainty is not a standard uncertainty and we establish, it may not be used for uncertainty propagation.
One consequence is, for example, that pairs $(y', u(y'))$ are not to be used in calibration certificates.

Keywords: Systematic effect, correction, uncorrected estimate, GUM, standard uncertainty, consistency, transferability.

1. Introduction – uncorrected effects

1.1. A metrological account

In metrology, a measurand is often related to its influencing quantities. Among all estimates for a measurand, the expectation of the associated state-of-knowledge distribution is usually considered the best estimate. In practice, however, alternative estimates are also employed for technical or economic reasons. For example, an input quantity whose effect on the measurand is estimated to be non-zero, is called a systematic effect (in clause 3.2 of the [1], hereafter called the *GUM*); and not correcting for such a systematic effect leads to an alternative (i.e., uncorrected) estimate.

The *GUM* requires that all significant systematic effects are corrected when evaluating the measurement uncertainty of a measurand (see *GUM* 3.2.4). However, it also concedes exceptions (Note 2 in 3.2.4, 6.3.1, and F.2.4.5) and international [2]–[7] as well as national [8]–[13] guidelines mention uncorrected systematic effects[1]. Also, publications like [14]–[20] and references therein are concerned with uncorrected systematic effects.

For the general case of uncorrected systematic effects, the *GUM* does not provide any guidance; only a special case is treated in F.2.4.5. The above documents either

(a) state vaguely that unperformed corrections may be incorporated in the uncertainty (e.g., [2], [3], [7]), or

(b) recommend separately reporting the uncorrected estimate (i.e., the estimate derived from not correcting for all systematic effects), together with the correction and the standard uncertainty associated with the corrected estimate (cf. [18]), or

(c) recommend quadratically adding the correction and the standard uncertainty associated with the corrected estimate, and then associating the root of this as the standard uncertainty with the uncorrected estimate (which has its seed in [15] and is the focus of this work), or

(d) suggest various ways to arrive at an expanded uncertainty or coverage interval which may be associated with the uncorrected estimate (e.g., [4]–[6], [14], [16]–[19], and guidelines in footnote 1),

or combine any of these options for treating uncorrected systematic effects.

This article examines uncertainties which can be associated with uncorrected estimates and possibly treated as standard uncertainty. To this end, we generalize option (c) and recall reasons for specifying such an uncertainty. Then we expose and discuss crucial consequences of applying this uncertainty. Notably, the uncertainty of an uncorrected estimate is inconsistent and not transferable and thus does not fulfil essential requirements of metrology.

1.2. A statistical account

Throughout this work, let Y and $g_Y(\eta)$ denote the random variable and its probability density function, representing the available information about a measurand. That is, g_Y shall be

[1] International guidelines mention uncorrected systematic effects, e.g., in clauses 2.26 in [2], 2.5.2 in [3], 4.4, 5.2 step 4, step 5 and 7.1 in [4], 7.5.2 in [5], 5.2.1.4.3 and B.2 in [6], and 5.3 in [7]. German guidelines mention uncorrected systematic effects, e.g., in equations (7.8), (7.16), and (7.17) in [8], (B.1) in [9], (B.3), and (B.4) in [10], (B.2) and (B.3) in [11], (8.7) in [12] as well as (12), (14), (19), (22), (29), and (30) in [13].

DOI: 10.2478/msr-2019-0026

the state-of-knowledge or degree-of-belief distribution mentioned in *GUM* 4.1.6 and in the introduction of [21]. We assume the expectation and the variance of the random variable Y exist.

One estimate for the measurand is the expectation of Y, which is

$$y := \mathbb{E}(Y) = \int \eta g_Y(\eta) d\eta.$$

It is considered the best estimate under quadratic loss

$$\mathbb{E}\left((Y-y)^2\right) \leq \mathbb{E}\left((Y-y')^2\right) := \mathrm{QL}(y') \tag{1}$$

in the *GUM* as well as in Bayesian statistics (see, e.g., [22]). That is, y minimizes the squared error risk compared to any other estimate y' for the measurand.

The standard uncertainty associated with the best estimate y is defined by

$$u^2(y) := \mathbb{E}\left((Y-y)^2\right) = \mathbb{V}(Y) \tag{2}$$

(cf. *GUM* 2.3.1, where the symbol $\mathbb{V}$ denotes the variance), and minimizes the expected quadratic loss (1). That is, the standard uncertainty and the best estimate are the pair representing the minimum of the loss function QL and where it is realized.

This work considers alternative estimates $y' \neq y$. In metrology, these alternative estimates may arise, for example, when systematic effects are not corrected for. Despite not being the best estimate, an uncertainty is to be associated with such an uncorrected estimate – one that ideally can be used as a standard uncertainty. Section 2 generalizes the suggestion from [15] and defines the associated uncertainty through the same loss function QL. Section 3 establishes that important metrological properties of the standard uncertainty are missing there. Finally, section 4 discusses consequences of this finding.

2. Quadratic-loss uncertainty

For any estimate y' of the measurand, let us define the associated uncertainty by

$$u^2(y') := \mathbb{V}(Y) + \left(\mathbb{E}(Y) - y'\right)^2. \tag{3}$$

We call $u(y')$ quadratic-loss (QL) uncertainty because, with $u^2(y') = \mathrm{QL}(y')$, it adopts the loss criterion (1), which is identical to the one the standard uncertainty (2), and thus the *GUM*, is based on. However, the QL uncertainty does not minimize this loss function as $\mathrm{QL}(y') = \mathrm{QL}(y) + (y'-y)^2$ and is therefore larger than the standard uncertainty $u(y)$ for any alternative estimate $y' \neq y$.

The QL uncertainty can be thought to extend the standard uncertainty (2), and thus the *GUM*, to uncorrected estimates. It provides a possibility to report uncertainties when not correcting for all systematic effects. However, not all metrological properties are preserved, as will be shown in section 3.2.

2.1. Special case in metrology

A special case of the QL uncertainty in metrology was suggested in [15]. Let us assume the measurand is related to input quantities $X_1, \ldots, X_N$ through a linear measurement model and to systematic effects B via

$$Y = f(X_1, \ldots, X_N) + B. \tag{4}$$

If the effects subsumed in B are independent of the other effects and are expected to cause the correction b, the authors in [15] define the uncertainty associated with the uncorrected estimate $y' := y - b$ for the measurand Y by

$$u^2(y') = u^2(y) + b^2, \tag{5}$$

with uncertainty $u(b)$ subsumed in $u(y)$.

2.2. Usage and terminology

The QL uncertainty is frequently used in metrology for uncorrected systematic effects, e.g., in [6], [16], [19], [20], [23], [24] and indirectly for expanded uncertainties in [5], [17], [18]. The publications [6], [15], [19] even call the QL uncertainty a standard uncertainty, whereas [20], [23] avoid this on purpose.

Clearly, uncertainty (2) – and not the QL uncertainty (3) – is the standard uncertainty of the measurand under the current knowledge. However, the (squared) standard uncertainty is not equal to the expected quadratic loss for the uncorrected estimate y'; i.e., reporting the pair $(y', u(y))$ is thus not an option.

More importantly, can the QL uncertainty be used as if it were a standard uncertainty? So far, no restrictions have been made in the literature; for example, in [6], [15], [16], [19], [20], [24]. Let us judge this practice with respect to the following metrological properties: internal consistency, transferability, and validity.

3. Metrological properties

Quite fundamentally, the *GUM* states in clause 0.4:

> The actual quantity used to express uncertainty should be
>
> - *internally consistent*: it 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;
> - *transferable*: it should be possible to use directly the uncertainty evaluated for one result as a component in evaluating the uncertainty of another measurement in which the first result is used.

In addition, the method for evaluating and expressing uncertainty should be *universal* and *valid* according to *GUM* 0.4. That is, it ‘should be applicable to all kinds of measurements and to all types of input data used in measurements,’ and it ‘should be capable of readily providing [...] an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject to measurement.' (The term *validity* was coined by [25].)

These properties are important in metrology to quantitatively indicate the quality of a measurement result. Universality permits uncertainties to be evaluated wherever measurements play their part – in science, engineering, commerce, industry, regulation, and so on. Consistency is crucial because it is a necessity to a consensus on the evaluation of uncertainty worldwide. Transferability is the foundation of the traceability of measurements to the SI. The international comparability of measurements cannot be established without consistency or without transferability – neither for national metrology institutes, for accreditation laboratories, in industry, nor for end users.

3.1. Properties of the standard uncertainty

The standard uncertainty (2) fulfils the properties of internal consistency, transferability, and validity.

Consistency and transferability require a notion of 'uncertainty components' which relate to the measurement result. Within the scope of the *GUM,* let us assume input quantities $X_1,\dots,X_N$, a linear measurement model $Y=\sum_{i=1}^N a_iX_i+a$, and a decomposition into, say, two arbitrary groups $X_1,\dots,X_j$ and $X_{j+1},\dots,X_N$. Then the standard uncertainty is *consistent*, because it may be calculated directly from the sum

$$\mathbb{V}(Y)=\mathbb{V}\left(\sum_{i=1}^N a_iX_i\right)=\sum_{i=1}^N a_i^2\mathbb{V}(X_i)+\sum_{i=1}^N\sum_{k\neq i}a_ia_k\mathbb{C}ov(X_i,X_k),\tag{6}$$

called the law of propagation of uncertainty (LPU, see *GUM* 5.1.1, 5.1.2, where $\mathbb{C}ov$ denotes the covariance); or alternatively from the groups of components

$$\begin{aligned}
&\mathbb{V}\left(\sum_{i\le j}a_iX_i+\sum_{i>j}a_iX_i\right)\\
&=\mathbb{V}\left(\sum_{i\le j}a_iX_i\right)+\mathbb{V}\left(\sum_{i>j}^N a_iX_i\right)+2\mathbb{C}ov\left(\sum_{i\le j}a_iX_i,\sum_{i>j}a_iX_i\right)\\
&=\sum_{i\le j}a_i^2\mathbb{V}(X_i)+2\sum_{i\le j}\sum_{k<i}a_ia_k\mathbb{C}ov(X_i,X_k)+\sum_{i>j}a_i^2\mathbb{V}(X_i)\\
&\quad+2\sum_{i>j}\sum_{k>i}a_ia_k\mathbb{C}ov(X_i,X_k)+2\sum_{i\le j}\sum_{k>j}a_ia_k\mathbb{C}ov(X_i,X_k)\\
&=\sum_{i=1}^N a_i^2\mathbb{V}(X_i)+\sum_{i=1}^N\sum_{k\neq i}a_ia_k\mathbb{C}ov(X_i,X_k)=\mathbb{V}(Y).
\end{aligned}$$

The result is identical.

The standard uncertainty is also *transferable* for linear models, because it is the output as well as the only input of the LPU in (6), when the covariance is thought of as the multivariate extension of the variance (cf. uncertainty matrix in 3.11 of [21]).

In addition, the standard uncertainty may be *valid* in part. That is, when the central limit theorem is known to hold (cf. *GUM* G.2), expanded uncertainties and coverage intervals can be derived as detailed in *GUM* 6.2 and 6.3.

As a side note, probability distributions and their propagation are also internally consistent, transferable and valid – even for nonlinear measurement models and beyond the central limit theorem. In contrast, coverage intervals are generally not consistent and not transferable on their own.

3.2. Properties of the quadratic-loss uncertainty

As noted in section 2, the QL uncertainty is per definition not a standard uncertainty for the measurand. Nevertheless, can it fulfil the needs in metrology defined above?

Firstly, the QL uncertainty is *valid* in part. That is, when the central limit theorem is known to hold, expanded uncertainties and coverage intervals with exact coverage can be derived either from the QL uncertainty or otherwise (see, e.g., [18]).

However, the QL uncertainty is *not consistent*. Evaluating the QL uncertainty in (3) for an estimate y' and a linear measurement model $Y=\sum_{i=1}^N a_iX_i+a$, gives

$$\begin{aligned}
u^2(y')&=\mathbb{V}(Y)+\left(\mathbb{E}(Y)-y'\right)^2\\
&=\mathbb{V}\left(\sum_{i=1}^N a_iX_i\right)+\left(\mathbb{E}\left(\sum_{i=1}^N a_iX_i+a\right)-y'\right)^2.
\end{aligned}$$

In contrast, let us evaluate the QL uncertainty from two arbitrary groups $X_1,\dots,X_j$ and $X_{j+1},\dots,X_N$, which result in the best estimates y_1 and y_2, and alternative estimates y_1' and y_2', respectively, such that $y=y_1+y_2$ and $y'=y_1'+y_2'$. For simplicity, let the groups of input quantities be independent of each other. While $\mathbb{V}(Y)$ itself is consistent (see section 3.1), we have

$$\begin{aligned}
u^2(y_1')+u^2(y_2')&=\mathbb{V}\left(\sum_{i\le j}a_iX_i\right)+\left(\mathbb{E}\left(\sum_{i\le j}a_iX_i+a_1\right)-y_1'\right)^2\\
&\quad+\mathbb{V}\left(\sum_{i>j}a_iX_i\right)+\left(\mathbb{E}\left(\sum_{i>j}a_iX_i+a_2\right)-y_2'\right)^2\\
&=\mathbb{V}(Y)+\left(y_1-y_1'\right)^2+\left(y_2-y_2'\right)^2\\
&=\mathbb{V}(Y)+\left(\mathbb{E}(Y)-y'\right)^2+2\left(y_1-y_1'\right)\left(y_2-y_2'\right)\\
&\neq u^2(y'),\quad\text{when } y_1'\neq y_1 \text{ and } y_2'\neq y_2.
\end{aligned}$$

The reason for this inconsistency lies in the general inequality $(d+e)^2\neq d^2+e^2$, when the estimate of the systematic effect $(y'-y)$ is decomposed and squared.

In addition, the QL uncertainty is *not transferable*. Per definition (3), the QL uncertainty is dependent on the specific estimate y'. Only having knowledge of the QL uncertainty of each input quantity therefore cannot be sufficient.

For the sake of completeness, similar arguments hold for the special case of alternative estimates in F.2.4.5 of the *GUM.* That is, also the uncertainty $u_c(y')$ associated with an estimate y' when corrections from a calibration curve are not applied, is not a standard uncertainty and should not be used as such.

4. Discussion, recommendation

This article focused on measurement results that are described by an estimate which is not the expectation of the current state-of-knowledge distribution. In metrology, such an estimate arises when not all systematic effects are corrected for. The central question for such a measurement result is then: how does one assign a measure for its quality which meets the requirements of metrology?

A natural choice for associating an uncertainty with these alternative estimates adopts the expected quadratic loss which the *GUM* also does for the standard uncertainty. The uncertainty for an estimate y', which is not equal to the best estimate y, is then $u(y') = \sqrt{u^2(y) + (y - y')^2}$ and is here called quadratic-loss uncertainty. It generalizes a previous suggestion [15] and existing practice, to alternative estimates which do not necessarily arise from uncorrected systematic effects.

This quadratic-loss uncertainty is not a standard uncertainty for alternative estimates. Nevertheless, it is applied in practice and provides a possibility to report uncertainties for estimates other than the best ones. Specifically, it uses the same foundation as the *GUM* for the evaluation of the standard uncertainty.

However, we demonstrated that and why the quadratic-loss uncertainty is not sufficient for the general needs in metrology. We proved that it is not consistent and is thus not to be used for uncertainty propagation.

As a consequence, the pair of an uncorrected estimate and its quadratic-loss uncertainty is not suitable for calibration certificates. The same pair may be suitable for characterizing measurement results which are not used for uncertainty propagation, e.g., for end devices at the factory floor level.

The authors recommend reviewing standards and guidelines that mention uncorrected effects in the light of these findings. That is, in the first place a re-examination is recommended whether and how a correction can be made. If uncorrected estimates cannot be avoided, it could be contemplated whether a restricted use of uncorrected effects is possible.

Acknowledgement

The authors would like to thank Dr. Röske (PTB) for providing a first list of guidelines which treat uncorrected systematic effects.

References

[1] Joint Committee for Guides in Metrology (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*. JCGM 100:2008.

[2] Joint Committee for Guides in Metrology (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. JCGM 200:2012, 3 edition.

[3] Ellison, S. L. R., Williams, A. (eds.) (2012). *EURACHEM / CITAC Guide CG 4 – Quantifying Uncertainty in Analytical Measurement*. EURACHEM, 3 edition. ISBN 978-0-948926-30-3.

[4] EURAMET (2011). *Uncertainty of Force Measurements*, Vol. cg-4. EURAMET, 2 edition. (03/2011) ISBN 978-3-942992-03-9.

[5] EURAMET (2015). *Guidelines on the Calibration of Non-Automatic Weighing Instruments*, Vol. cg-18. EURAMET, 4 edition. (11/2015) ISBN 978-3-942992-40-4.

[6] International Organization for Standardization (2005). *Gas analysis – Investigation and treatment of analytical bias*. ISO 15796:2005.

[7] International Organization for Standardization (2007). *Air quality – Guidelines for estimating measurement uncertainty*. ISO 20988:2007.

[8] Deutscher Kalibrierdienst (DKD) (2018)a. *Richtlinie DKD-R 3-3, Kalibrierung von Kraftmessgeräten*. Braunschweig, Germany: Physikalisch-Technische Bundesanstalt.

[9] Deutscher Kalibrierdienst (DKD) (2018)b. *Richtlinie DKD-R 3-5, Kalibrierung von Drehmomentmessgeräten für statische Wechseldrehmomente*. Braunschweig, Germany: Physikalisch-Technische Bundesanstalt.

[10] Deutscher Kalibrierdienst (DKD) (2018)c. *Richtlinie DKD-R 3-7, Statische Kalibrierung von anzeigenden Drehmomentschlüsseln*. Braunschweig, Germany: Physikalisch-Technische Bundesanstalt.

[11] Deutscher Kalibrierdienst (DKD) (2018)d. *Richtlinie DKD-R 3-8, Statische Kalibrierung von Kalibriereinrichtungen für Drehmomentschraubwerkzeuge*. Braunschweig, Germany: Physikalisch-Technische Bundesanstalt.

[12] Deutscher Kalibrierdienst (DKD) (2018)e. *Richtlinie DKD-R 3-9, Kontinuierliche Kalibrierung von Kraftaufnehmern nach dem Vergleichsverfahren*. Braunschweig, Germany: Physikalisch-Technische Bundesanstalt.

[13] Deutscher Kalibrierdienst (DKD) (2014). *Richtlinie DKD-R 6-1, Kalibrierung von Druckmessgeräten*. Braunschweig, Germany: Physikalisch-Technische Bundesanstalt. Revision 2.

[14] Phillips, S. D., Eberhardt, K. R., Parry, B. (1997). Guidelines for expressing the uncertainty of measurement results containing uncorrected bias. *Journal of Research of the National Institute of Standards and Technology* 102(5), 577–585.

[15] Lira, I. H., Wöger, W. (1998). Evaluation of the uncertainty associated with a measurement result not corrected for systematic effects. *Measurement Science and Technology* 9(6), 1010.

[16] Haesselbarth, W. (2004). Accounting for bias in measurement uncertainty estimation. *Accreditation and Quality Assurance* 9(8), 509–514.

[17] Synek, V. (2005). Attempts to include uncorrected bias in the measurement uncertainty. *Talanta* 65(4), 829–837.

[18] Magnusson, B., Ellison, S. L. R. (2008). Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. *Analytical and Bioanalytical Chemistry* 390(1), 201–213.

[19] Hernla, M. (2008). Messunsicherheit bei nicht korrigierten systematischen Messabweichungen. *tm Technisches Messen* 75(11), 609–615.

[20] Physikalisch-Technische Bundesanstalt (2010). Erklärung der PTB zur Behandlung systematischer Abweichungen bei der Berechnung der Messunsicherheit. https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_8/8.4_mathematische_modellierung/8.40/GUM_-_systematische_Abweichungen.pdf.

[21] Joint Committee for Guides in Metrology (2008). *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method*. JCGM 101:2008.

[22] Robert, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer.

[23] Wöger, W. (2001). Zu den modernen Grundlagen der Datenauswertung in der Metrologie. *PTB-Mitteilungen* 3, 210–225.

[24] Lira, I. (2002). *Evaluating the Measurement Uncertainty: Fundamentals and Practical Guidance*. CRC Press.

[25] Wang, C. M., Iyer, H. K. (2010). On multiple-method studies. *Metrologia* 47(6), 642.

Received May 6, 2019.
Accepted August 30, 2019.