

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,100

Open access books available

116,000

International authors and editors

120M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Uncertainty of Measurement in Medical Laboratories

---

Paulo Pereira

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62437>

---

## Abstract

The “Guide to the Expression of Uncertainty in Measurement” (GUM) is not systematically used in medical laboratories, for what the laboratorian should understand the Uncertainty Approach and its importance to recognize the level of realism of results. This chapter presents, discusses, and recommends the models fulfilling GUM principles. An example is given to a single test for an easier understanding of the determination of measurement uncertainty. All the practice uses a freeware. Results with larger measurement uncertainty intervals have a significant probability of being unrealistic, arising a high risk of the uncorrected clinical decision. A flow chart to the selection of models for the determination of measurement uncertainty in a medical laboratory is recommended.

**Keywords:** bias, bottom-up, GUM, measurement uncertainty, medical metrology, precision, top-down, trueness

---

## 1. Introduction to the “Guide to the Expression of Uncertainty in Measurement”

“The lack of consensus in the international scientific community regarding the expression of measurement uncertainty occurred in 1977. Two years later, the Bureau International des Poids et Mesures (BIPM) and 21 laboratories agreed upon it was important to develop an international procedure for expressing measurement uncertainty and for combining individual uncertainty components into a single total uncertainty measurement in chemistry and physics. However, there was no consensus regarding the calculus for expression of measurement uncertainty. In 1980, the BIPM Working Group on the Statement of Uncertainties developed the Recommendation INC-1 “Expression of Experimental Uncertainties” [1], approved in the Comité Interna-

---

tional des Poids et Mesures (CIPM) in 1981 [2] and reapproved 5 years after [3]. The CIPM suggested to the International Organization for Standardization (ISO) the development of a measurement uncertainty master document based on the Working Group recommendation. The ISO Technical Advisory Group on Metrology (TAG 4) assumed the responsibility to write the guideline. Seven organizations participated in the work of TAG 4: BIPM, International Electrotechnical Commission (IEC), ISO, International Organization of Legal Metrology (OIML), International Union of Pure and Applied Chemistry (IUPAC), International Union of Pure and Applied Physics (IUPAP), and International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), representing the medical laboratory. TAG 4 defined the Working Group 3 (ISO/TAG 4/WG 3) with a committee of experts from BIPM, OIML, IEC, and ISO. The guideline "Guide to the Expression of Uncertainty in Measurement" (GUM) was first published in 1993 and 2 years later was corrected and reprinted. Measurement uncertainty is defined as the "nonnegative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used" (entry 2.26 of [4]). It characterizes "the quality of a result of a measurement" expressed in uncertainty (quantitative indication). It was not intended to be applied to another quantity than numerical, for what it cannot be a standard to the determination of measurement uncertainty in all estimations. GUM is an open access document as it was republished with minor correction by BIPM in 2008 [5]. The terminology to the Uncertainty Approach is part of the International Vocabulary of Metrology (VIM), which is also freely available from the Bureau [4].

Measurement uncertainty provides information on the level of confidence on the measurement result. One example where the uncertainty is needed is when comparing a result to a clinical decision value. Measurement uncertainty is the quantifiable expression of the doubt related with the outcome. The expanded uncertainty  $U$  provides an interval within which the value of the measurand is assumed to be determined by a defined level of confidence. It is the result of multiplying the standard combined uncertainty  $u_c$  by a coverage factor  $k$ . The choice of the factor  $k$  is established on the level of confidence desired. For an approximate level of confidence of 95%,  $k$  is usually set to 2, and with a confidence higher than 99%,  $k$  is typically set to 3 when degrees of freedom for the combined uncertainty are more than 20.

Commonly, the measurement uncertainty result is expressed as value $\pm$ expanded uncertainty. For example, in a screening immunoassay, a ratio of  $1.0\pm 0.3$  (expanded uncertainty  $k=2$ ) corresponds to the interval 0.7 to 1.3 considering the clinical decision ("cutoff") ratio is equal to 1. It is interpreted as that the measured value $\pm$ expanded uncertainty covers the "cutoff" value, and the result is in a measurement uncertainty where the result positive or negative cannot be declared (i.e., it is considered indeterminate).

Pereira et al. [6] claimed that a valid uncertainty evaluation should consider:

- "A clear definition of the measurand (i.e., the quantity to be measured),
- A comprehensive specification of the measurement procedure and the measurement objects, and

- A comprehensive analysis of the effects impacting the measurement results. From a laboratory view, the main effects are intermediate precision (within-laboratory precision) and any residual measurement bias.”

According to the European Federation of National Associations of Measurement, Testing and Analytical Laboratories (Eurolab) Technical Report 1/2007 [7], there are four main approaches to estimating measurement uncertainty fulfilling Uncertainty Approach principles (see 2.1):

- Modeling,
- Single laboratory validation [including quality control (QC)],
- Interlaboratory comparisons, and
- External Quality Assessment (EQA) [Proficiency Testing (PT)].

Dybkaer [8], past IFCC President and a recognized supporter of GUM approach, argued in a paper that “this International Standard will outline the principles of estimating measurement uncertainty according to GUM (1993, 1995), ways of simplification in routine measurement, and possibilities of reporting. This International Standard applies to:

- Routine medical laboratories wishing or required to provide values with measurement uncertainty;
- Medical laboratories seeking accreditation according to ISO/IEC 17025 (testing) or ISO 15189;
- Organizations providing accreditation; and
- Manufacturers of *in vitro* diagnostic devices wishing to provide guidance on measurement uncertainty in their information documents” [8].

Eurachem/CITAC published in 2000 the “Quantifying Uncertainty in Analytical Measurement” (QUAM; revised for the third time in 2012) intended to be applied uniquely to measurement uncertainty in chemistry [9]. This document takes into account the practical experience of estimation of measurement uncertainty in chemistry laboratories. It emphasizes that procedures introduced by chemistry laboratories to estimate their measurement uncertainty should be integrated with existing quality assurance measurements, as these measures frequently provide much of the information required to evaluate the measurement uncertainty. It provides unequivocally for the use of validation and related data in the estimation of measurement uncertainty in full compliance with GUM principles. Different from GUM, its purpose is for not only modeling approaches (“bottom-up”) but also empirical approaches (“top-down”). The “top-down” approaches satisfy the need of chemistry laboratories looking to an alternative to modular models, given that these models are often inapplicable in laboratories’ methods such as medical laboratories’ methods. Its approach is consistent with the ISO/IEC 17025 and even with ISO 15189 requirements. Because medical laboratories methods are mainly chemical, QUAM is an adequate reference to support the estimation of measurement uncertainty in this field, and it is referred in this Chapter to describe the basic concepts of models to the estimation of measurement uncertainty (see 2.2).

The Nordtest TR 537 “Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories” [10], along with Eurolab TR 1/2007 “Measurement Uncertainty Revisited: Alternative Approaches to Uncertainty Evaluation” [7], proposes “top-down” approaches easily practicable in laboratories, including medical laboratories. Such as Eurachem/CITAC documents these are open access publications. The Finnish Environment Institute (SYKE) released a freeware, MUKit, featuring TR 537 mathematical models [11].

The National Pathology Accreditation Advisory Council (NPAAC) published in 2007 a public document entitled “Requirements for the Estimation of Measurement Uncertainty” to support the Australian accreditation of medical laboratories [12]. This guideline recommends a set of “top-down” models to use in this field.

The Clinical and Laboratory Standards Institute (CLSI)/IFCC C51-A “Expression of Measurement Uncertainty in Laboratory Medicine” was published in January 2012 (the code changed to EP29-A without text revision) [13] and it is the only global guideline for the determination of measurement uncertainty. The CLSI EP29 Working Group is chaired by Anders Kalnner, another recognized GUM expert in medical laboratories. Its application to medical laboratories is not consensual [14, 15]. According to this guideline, it “describes the principles of estimating measurement uncertainty and guides clinical laboratories and *in vitro* diagnostic device manufacturers on the specific issues to be considered for implementation of the concept in medical laboratory. This document illustrates the assessment of measurement uncertainty with both ‘bottom-up’ and ‘top-down’ approaches. The ‘bottom-up’ approach suggests that all possible sources of uncertainty are identified and quantified in an uncertainty budget. A combined uncertainty is calculated using statistical propagation rules. The ‘top-down’ approach directly estimates the measurement uncertainty results produced by a measuring system. The methods to estimate the precision and bias are presented theoretically and in worked examples.”

The ISO/PDTS 25680 “Medical Laboratories—Calculation and Expression of Measurement Uncertainty” [16] was written initially as an International Standard and later as a Technical Specification and was canceled on June 2011. The contents of the last draft were analogous to the CLSI C51-P.

Currently, the determination of measurement uncertainty is required in ISO/IEC 17025 and ISO 15189 [17]. It is mandatory in Australia, Latvia, and France after November 1, 2016.

The laboratorian should understand that several uncertainty components are immeasurable in Uncertainty Approach, for example, diagnostic uncertainty in tests with binary results (positive/negative). For further details about other sources of uncertainty, please refer to [18], and for diagnostic uncertainty, please refer to [19].

This Chapter presents a discussion of the estimation models of measurement uncertainty and its determination in a single medical laboratory test. Preferably, the reader should have basic statistical skills to understand the concepts and mathematical models. There are cited several references that should be revised for a deeper understanding of the Uncertainty Approach and also to review the examples of estimation in other medical laboratory tests.

## 2. Principles of GUM

### 2.1. Principles of uncertainty approach

GUM is recognized as the master document on measurement uncertainty throughout the field of metrology, for what it is also called “uncertainty bible.” The measurement uncertainty evaluation is recognized to be applied to all types of quantitative assay results in physics and chemistry. The theoretical modeling follows the Uncertainty Approach published in 1980. The principles are as follows:

- a. “The uncertainty in the result of a measurement generally consists of several components, which may be grouped into two categories according to the way in which their numerical value is estimated:

Those which are evaluated by statistical methods and

Those which are evaluated by other means.

There is not always a simple correspondence between the classification into categories A or B and the previously used classifications into “random” and “systematic” uncertainties. The term “systematic uncertainty” can be misleading and should be avoided.

Any detailed report of the uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

- b. The components in category A are characterized by the estimated variances (or the estimated standard deviation  $s_i$ ) and the number of degrees of freedom  $v_i$ . Where appropriate, the covariances should be given.
- c. The components in category B should be characterized by quantities  $u_j$ , which may be considered as approximations to the corresponding variances, the existence of which is assumed. The quantities may be treated like variances and the quantities  $u_j$  like standard deviations. Where appropriate, the covariance should be treated in similar ways.
- d. The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of “standard deviations.”
- e. If, for particular applications, it is necessary to multiply the combined uncertainty by a factor to obtain an overall uncertainty, the multiplying factor used must always be stated [1].

These principles provide a methodology where:

- Uncertainty evaluation should combine the major sources of uncertainty;
- Uncertainties of random and systematic error components should be treated identically;
- Major contributions coming from Type A and Type B errors should be identified and considered as standard uncertainty; and

- Uncertainty components are expressed as combined standard uncertainty or by the result of combined standard uncertainty multiplied by a defined coverage factor, designated expanded uncertainty [7].

The approach is based on a model designed to consider the interrelation of all sources of uncertainty that significantly affect the measurand. All known systematic errors must be corrected if significant, and they are not included in the calculations. All the principal sources are combined according to the laws of propagation of uncertainty. The calculated result is the combined uncertainty  $\pm$  value associated with a quantitative measured value. The uncertainty budget expresses all measurement sources as well as the combined calculus. A Pareto chart can be used to display visually and compare different sources and their weight on combined uncertainty. This step requires the medical laboratory to define a mathematical model, which could require the competencies of a mathematician or a statistician with special training in measurement uncertainty. The combined uncertainty is multiplied by a Student's  $t$  value and its result, referred as "expanded uncertainty," is expressed by a confidence interval (e.g.,  $k=2$  to a confidence interval equal to 95%) [20, 21]. For further details about principles of Uncertainty Approach, please refer to [5].

## 2.2. Fundamental concepts of models to the estimation of measurement uncertainty

### 2.2.1. Stochastic mathematical equation of the measurement method

The equation should assure the stochasticity of the reaction. Nonstochastic equations cause an incorrect estimate of measurement uncertainty. This is a complex task when designing a modeling approaches, requiring principally in chemical expertise in medical laboratories' tests. It is uniquely applied to modeling approaches.

### 2.2.2. Distribution type in the estimation of standard deviations

For each uncertain variable, the probable results with a probability distribution should be defined. The conditions contiguous to the variable determine the category of certain distribution. The categories consist of the following:

- a. Discreet
  - Bernoulli (i.e., success/failure in a single experiment);
  - Binomial (i.e., the number of success/failure in some experiments);
  - Multi-nominal (i.e., the frequency registered a certain result in some experiments); and
  - Poisson (i.e., number of events of a certain occurrence from a period zero to a period  $t$ ).
- b. Continuous variables: exponential and normal.

Two types of estimations could be made: Type A or Type B. These different estimates are merged in the combined standard uncertainty. Before the combination, all uncertainty sources

must be expressed as standard uncertainties (i.e., as relative standard deviations). The standard deviation shall be measured according to data distribution, which could not be a normal distribution due to the number of different uncertainty sources that could include not only chemical but also physics. The following *rules* for distribution could be applied to most of the uncertainty sources in medical laboratory:

*Rule 1* for distribution: Normal distribution with known  $n$

“Where the uncertainty component was evaluated experimentally from the dispersion of repeated measurements, it can readily be expressed as a standard deviation. For the contribution to uncertainty in single measurements, the standard uncertainty is simply the observed standard deviation; for results subjected to averaging, the standard deviation of the mean is used.”

The standard deviation of the mean of  $n$  values taken from a population is given by  $n$ :

$$u(x_i) = s_{\bar{x}} = s / \sqrt{n} \quad (1)$$

*Rule 2* for distribution: Normal distribution with known confidence interval

“Where an uncertainty estimate is derived from previous results and data, it may already be expressed as a standard deviation. However, where a confidence interval is given with a level of confidence  $p\%$  (in the form  $\pm a$  at  $p\%$ ), then divide the value  $a$  by the appropriate percentage point of the normal distribution for the level of confidence given to calculate the standard deviation.”

$$u(x_i) = a / (t - \text{value}) \quad (2)$$

*Rule 3* for distribution: Rectangular distribution

“If limits of  $\pm a$  are given without a confidence level and there is reason to expect that extreme values are likely, it is normally appropriate to assume a rectangular distribution, with a standard deviation of”

$$u(x_i) = a / \sqrt{3} \quad (3)$$

*Rule 4* for distribution: Triangular distribution

“If limits of  $\pm a$  are given without a confidence level, but there is reason to expect that extreme values are unlikely, it is normally appropriate to assume a triangular distribution, with a standard deviation of”

$$u(x_i) = a / \sqrt{6} \quad (4)$$

*Rule 5 for distribution: Estimation made by judgment*

“Where an estimate is to be made on the basis of judgment, it may be possible to estimate the component directly as a standard deviation. If this is not possible, then an estimate should be made of the maximum deviation that could reasonably occur in practice (excluding simple mistakes). If a smaller value is considered substantially more likely, this estimate should be treated as descriptive of a triangular distribution. If there are no grounds for believing that a small error is more likely than a large error, the estimate should be treated as characterizing a rectangular distribution” (entry 8.1 of [9]).

### 2.2.3. Combined standard uncertainty according to the law of the propagation of uncertainty

VIM defines combined standard uncertainty as the “standard measurement uncertainty that is obtained using the individual standard measurement uncertainties associated with the input quantities in a measurement model. Note: In case of correlations of input quantities in a measurement model, covariances must also be taken into account when calculating the combined standard measurement uncertainty (...)” (entry 2.31 of [4]).

Combined uncertainty is the result of calculus according to the law of the propagation of uncertainty, which combines the sources of uncertainty in a single value of uncertainty. The law of the propagation of uncertainty is derived based on the Taylor series expansion of a functional relationship regularly used in differential calculus. A measurement process can be modeled mathematically using the function:

$$y = f(x_1, x_2, \dots, x_n) \quad (5)$$

When sources are not correlated (independent variables), the calculus is done according to the variance rules; when sources are correlated (dependent variables, share of variance), the calculus is done according to covariance [(entry Chapter 8 of [20], 22].

### 2.2.4. Independent variables

“The general relationship between the combined standard uncertainty  $u_c(y)$  of a value  $y$  and the uncertainty of the independent parameters  $x_1, x_2, \dots, x_n$  on which it depends on is

$$u_c(y) = (y(x_1, x_2, \dots)) = \sqrt{\sum_{i=1, n} c_i^2 u(x_i)^2} = \sqrt{\sum_{i=1, n} u(y, x_i)^2} \quad (6)$$

where  $y(x_1, x_2, \dots)$  is a function of several parameters,  $x_1, x_2, \dots, c_i$  is a sensitivity coefficient evaluated as  $c_i = \partial y / \partial x_i$ , the partial differential of  $y$  on  $x_i$ , and  $u(y, x_i)$  denotes the uncertainty in  $y$  arising from the uncertainty in  $x_i$ . Each variable's contribution  $u(y, x_i)$  is just the square of the

associated uncertainty expressed as a standard deviation multiplied by the square of the relevant sensitivity coefficient. These sensitivity coefficients describe how the value of  $y$  varies with changes in the parameters  $x_1, x_2$ , etc.” (entry 8.2.2 of [9]).

### 2.2.5. Dependent variables

“Where variables are not independent, the relationship is more complex:

$$u\left(y\left(x_{i,j,\dots}\right)\right)=\sqrt{\sum_{i=1,n}c_i^2u\left(x_i\right)^2+\sum_{\substack{i,k=1,n \\ i\neq k}}c_ic_ku\left(x_i,x_k\right)} \quad (7)$$

where  $u\left(x_i,x_k\right)$  is the covariance between  $x_i$  and  $x_k$ , and  $c_i$  and  $c_k$  are the sensitivity coefficients (...). The covariance is related to the correlation coefficient  $r_{ik}$  by

$$u\left(x_i,x_k\right)=u\left(x_i\right)u\left(x_k\right)r_{ik} \quad (8)$$

where  $-1 \leq r_{ik} \leq 1$ ” (entry 8.2.3 of [9]).

QUAM recommends the Kragten spreadsheet to the determination of measurement uncertainty [23].

### 2.2.6. Simpler expressions for independent variables

It is possible to avoid the use of complex calculus and to use simpler expressions taken from the expression for the general relationship of variables. Two rules are proposed [root-sum-of-squares (RSS)]:

*Rule 1 for combination of uncertainty: Sum or difference*

“For models involving only a sum or difference of quantities [e.g.,  $y=(p + q + r + \dots)$ ], the combined standard uncertainty  $u_c(y)$  is given by

$$u_c\left(y\left(p,q,\dots\right)\right)=\sqrt{u\left(p\right)^2+u\left(q\right)^2+\dots} \quad (9)$$

*Rule 2 for combination of uncertainty: Product or quotient*

“For models involving only a product or quotient [e.g.,  $y=(p \cdot q \cdot r \cdot \dots)$ ] or

$$y = p / (q \cdot r \cdot \dots) \quad (10)$$

The combined standard uncertainty  $u_c(y)$  is given by

$$u_c(y) = y \sqrt{(u(p)/p)^2 + (u(q)/q)^2 + \dots} \quad (11)$$

where  $u(p)/p$ , etc., are the uncertainties in the parameters, expressed as relative standard deviations. Note: Subtraction is treated in the same manner as addition, and division in the same way as multiplication" (entry 8.2.6 of [9]).

Theoretically, because RSS does not take into account the partial derivatives, it results in an inaccurate uncertainty result. However, the inaccuracy of determination is usually considered nonsignificant to the estimated result.

### 2.2.7. Expanded uncertainty

VIM defines expanded uncertainty as the "product of a combined standard measurement uncertainty and a factor larger than the number 1. Notes: (1) The factor depends on the type of probability distribution of the output quantity in a measurement model and on the selected coverage probability. (2) The term "factor" in this definition refers to a coverage factor. (3) Expanded measurement uncertainty is termed "overall uncertainty" in paragraph 5 of Recommendation INC-1 (1980) (see the GUM) and simply "uncertainty" in IEC documents" (entry 2.35 of [4]).

The final step in the evaluation of measurement uncertainty is the calculus of expanded uncertainty. Its purpose is the designation of an interval that may be expected to include a large fraction of the distribution of values, which could reasonably be attributed to the measurand. Its calculus is according to the formula:

$$U = k \cdot u \quad (12)$$

where  $k$  is a coverage factor according to the type of probability distribution and  $u$  is the combined uncertainty. The choice of  $k$  value should be done according to factors such as the level of confidence required, any knowledge of the underlying distributions, or any knowledge of the number of values used to estimate random effects.

The value for  $k$  in a medical laboratory usually is taken from a one- or two-tailed normal distribution for Student's  $t$ . When the effective degrees of freedom  $v_{eff}$  are higher than about 6, usually  $k$  is equal to 2, which correspond to 95% confidence; when  $v_{eff}$  are less than about 6, they shall be defined. The European Co-Operation for Accreditation (EA) recommends in EA4-02 "Expression of the Uncertainty of Measurement in Calibration," a formula for  $v_{eff}$  calculus [24]:

"Estimate the effective degrees of freedom  $v_{eff}$  of the standard uncertainty  $u(y)$  associated with the output estimate  $y$  from the Welch-Satterthwaite formula:

$$v_{eff} = \left( u^4(y) \right) / \left( \sum_{i=1}^N u_i^4(y) / v_i \right) \quad (13)$$

where  $u_i(y)$  ( $i=1, 2, \dots, N$ ), defined in the equation, are the contributions to the standard uncertainty associated with the output estimate  $y$  resulting from the standard uncertainty associated with the input estimate  $x_i$ , which are assumed to be mutually statistically independent, and  $v_i$  is the effective degrees of freedom of the standard uncertainty contribution  $u_i(y)$ ."

When a standard uncertainty is measured requiring Type A evaluation, the degrees of freedom  $v_i$  are measured according to a simpler formula:

$$v_i = n - 1 \quad (14)$$

The table with  $k$  values set from a  $t$  distribution evaluated for a coverage probability of 95% to be used for Type A evaluation is featured in (entry Annex G of [5]). If  $v_{eff}$  is not an integer, which is usually the case, truncate  $v_{eff}$  to the next lower integer. When a standard uncertainty is measured requiring Type B evaluation, the calculus is more complex. The common practice is to carry out such assessments in a mode that guarantees that any underestimation is avoided. Considering this practice is followed, the degrees of freedom of the standard uncertainty  $u(x_i)$  acquired from a Type B evaluation may be taken to be . The estimation of  $k$  could be easily done in Microsoft® Excel® using the function=TINV(probability;deg\_freedom) (note: for a 95% confidence, the probability is equal to the difference between 1 and 0.95). For further details about  $v_{eff}$  and levels of confidence in measurement uncertainty, please refer to (entry Annex G of [5]).

### 2.2.8. Reporting the measurement uncertainty

#### a. The report

The measurement uncertainty report should include:

- “A description of the methods used to calculate the measurement result and its uncertainty from the experimental observations and input data;
- The values and sources of all corrections and constants used in both the calculation and the uncertainty analysis; and
- A list of all the components of uncertainty with full documentation on how each was evaluated.”

Note: These elements could be also referred to documented sources.

The data and analysis should be presented in such a way that its important steps can be readily followed and the calculation of the result repeated if necessary (...).”

(...) “When reporting the results of routine analysis, it may be sufficient to state only the value of the expanded uncertainty and the value of  $k$ ” (entry 9.2 of [9]).

The combined standard uncertainty or expanded uncertainty is reported. The majority of assay and calibration certificates describe expanded uncertainty. QUAM proposes the following format for reporting:

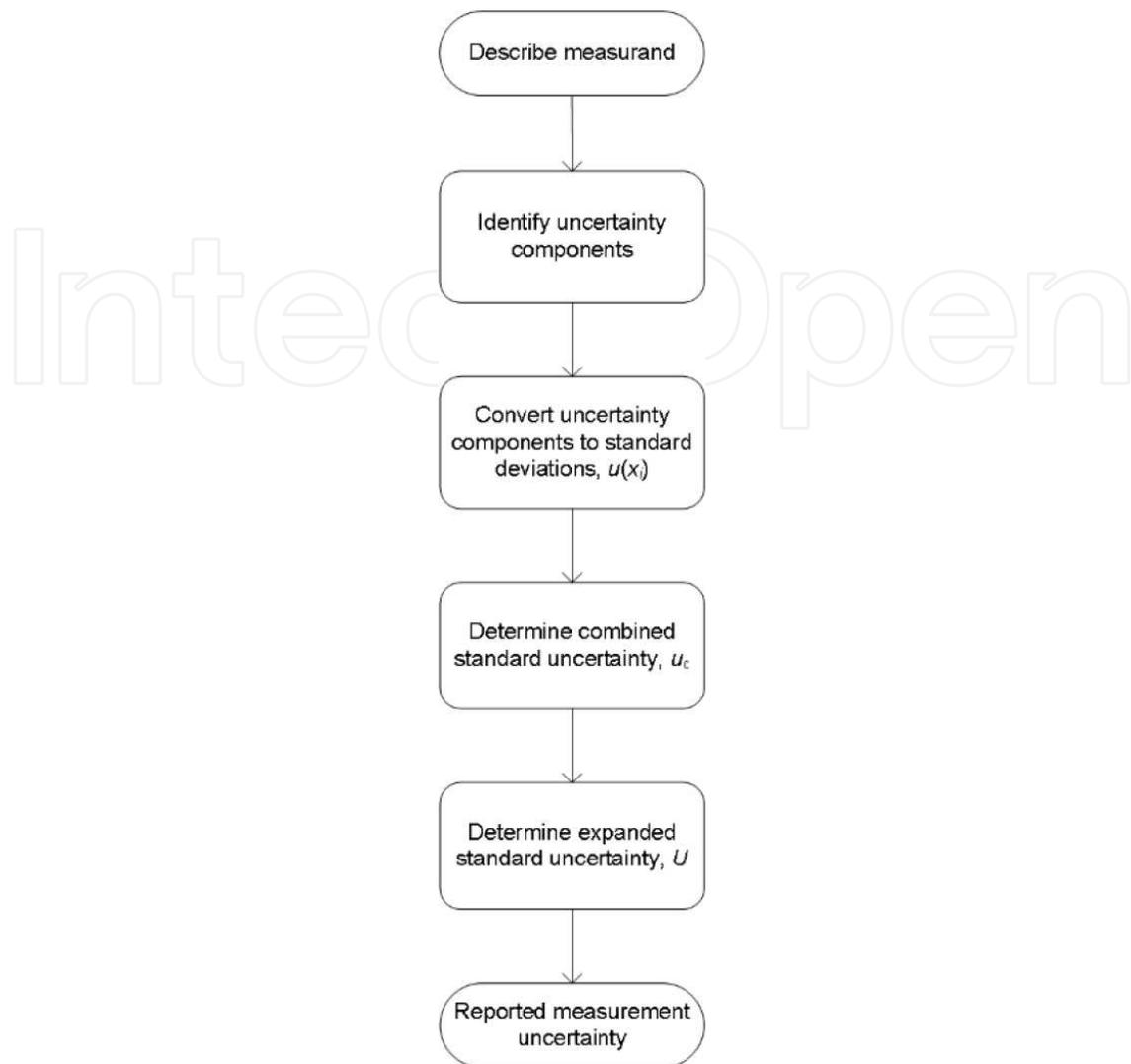
**Combined standard uncertainty:** “(Result):  $x$  (units) [with a] standard uncertainty of  $u_c$  (units) [where standard uncertainty is as defined in the ISO/IEC “Guide to the Expression of Uncertainty” and corresponds to 1 standard deviation].” Note: The use of the symbol  $\pm$  is not recommended when using standard uncertainty as the symbol is commonly associated with intervals corresponding to high levels of confidence. Terms in parentheses [] may be omitted or abbreviated as appropriate” (entry 9.3 of [9]).

**Expanded uncertainty:** “Unless otherwise required, the result  $x$  should be stated together with the expanded uncertainty  $U$  calculated using a coverage factor  $k=2$  (...). The following form is recommended: “(Result): ( $x$ )  $U$  (units) [where] the reported uncertainty is [an expanded uncertainty as defined in the International Vocabulary of Basic and General terms in Metrology, 2nd ed., ISO 1993 (note: this Chapter suggests the use the present VIM edition referred in [4]), calculated using a coverage factor of 2 [which gives a level of confidence of approximately 95%].” Terms in parentheses [] may be omitted or abbreviated as appropriate. The coverage factor should, of course, be adjusted to show the value actually used” (entry 9.4 of [9]).

The customers and consumers of this information must have the skills to understand the purpose of measurement uncertainty and to take appropriate actions based on this evaluation. Such skills are often lacking for the customers and consumers of laboratory test (physicians, patients, and agencies).

#### b. Reporting in medical laboratory scope

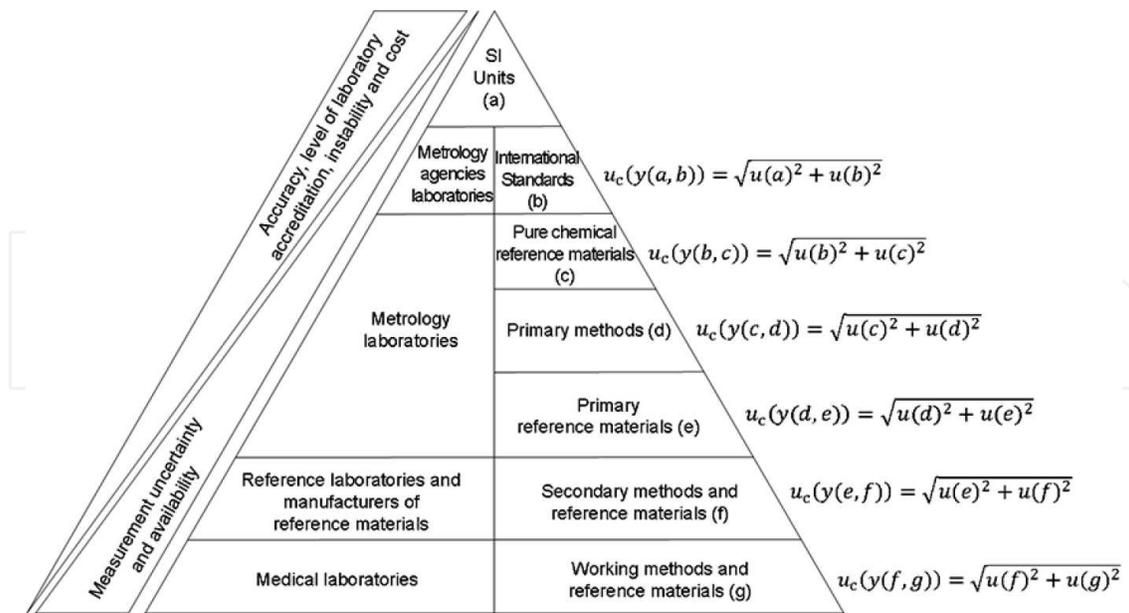
The information required to report the result of measurement depends on its intended use. In the medical laboratory, the final consumer is the patient, blood donor, or other, who is not responsible for the diagnosis and follow-up. The primary customer is the physician or someone else with the responsibility for the technical action (screening, diagnosis, follow-up, or other). The personnel who takes a decision on the result must understand the purpose of measurement uncertainty and its value for the judgment. If this does not happen, reports of measurement uncertainty could generate doubts that could compromise the clinical decision. These skills are rare in physicians or other healthcare professionals. It is not requested by the physician who probably does not understand its concept. Therefore, the majority of hospital laboratories do not report measurement uncertainty because it does not add value to clinical decisions and may only cause indecision. The report must be fit for its purpose [(entry 9.4 of [9]), 25]. **Figure 1** summarizes the stages of measurement uncertainty methodology.



**Figure 1.** Methodology to determine and report measurement uncertainty.

### 2.2.9. Metrological traceability

Metrological traceability is defined as the “property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” (entry 2.41 of [4]). Therefore, measurement uncertainty result should be metrologically traceable, assuring the comparability of outcomes in a metrological traceability chain, which is defined as the “sequence of measurement standards and calibrations that are used to relate a measurement result to a reference” (entry 2.42 of [4]). **Figure 2** shows an example of a metrological chain for a medical laboratory test. The measurement uncertainties and bias are determined according to the metrological traceability chain. The accuracy, level of accreditation laboratory, instability, and cost per material increase significantly from medical laboratories to the top of the hierarchy. On the contrary, the measurement uncertainty, bias, and availability of materials decrease from bottom to top.



**Figure 2.** A metrological traceability chain of measurement results in a medical laboratory test.

Although properly implemented in general metrology, it is not widely employed in most of medical laboratory tests due to the unavailability of reference materials and reference methods. Also, the “medical traceability” is hard to achieve due to the “physicochemical complexity” of human samples caused principally by the within-individual and interindividual biological variation [26]. For further details about metrological traceability, please refer to [27].

#### 2.2.10. Effect of bias in measurement uncertainty

Significant systematic effects should be corrected (entry Annex D of [5]). Where bias is statistically significant and is uncorrected, it should be reported with the measurement uncertainty (entry 7.16 of [9]).

Dybkaer [8] claimed that minimizing a diagnostic misclassification required that “trueness obtained through metrological traceability based on a calibration hierarchy.” On the article, he argued that the reduction of bias should happen according to seven approaches:

- “The type of quantity that is to be measured must be defined sufficiently well. This is particularly demanding when analyte isomorphs or speciation are involved.
- The principle and method of measurement must be carefully selected for analytical specificity.
- A practicable measurement procedure including sampling must be exhaustively described.
- A calibration hierarchy must be defined to allow metrological traceability, preferably to a unit of the International System of Units (SI). Traceability involves plugging into a reference measurement system of reference procedures and commutable calibration materials.
- An internal QC (IQC) system must be devised to reveal increases in bias.

- Any correction procedures must be defined and validated.
- Where possible, there should be participation in EQA (PT) using material with reference measurement values” [28].

Magnusson and Ellison [29] published a paper on the treatment of uncorrected measurement bias when determining measurement uncertainty, suggesting a correction process and determination of uncertainty for corrected results.

In some medical laboratory fields, such as virology, some biases cannot be corrected as seronegative “window period,” verification bias, spectrum bias, and length bias. Therefore, estimates of measurement uncertainty may not adequately describe the variability that is observed (entry 1.2.5 of [30]). For further details about dealing with bias in measurement uncertainty, please refer to [29].

## 2.3. Modeling approach

### 2.3.1. Fundamental concepts

This model is usually expressed in an equation accounting for the interrelation of the input quantities affecting the measurand pooled essentially with empirical data. Commonly, it is used a combination of different approaches to determine measurement uncertainty. The model incorporates a correction to minimize the effect of identified systematic error components. The law of propagation of uncertainty or propagation of distributions by Monte Carlo simulation method allows the computing of the combined standard uncertainty  $u_c$  of the result. The combined standard uncertainty is determined, and factor  $k$  is selected according to the chosen confidence level to determine expanded uncertainty. The partial derivative method requires a more complex model equation including partial derivatives of standard uncertainties. For further details about the modeling approach for the evaluation of uncertainty, please refer to (entry Chapter 8 of [5]).

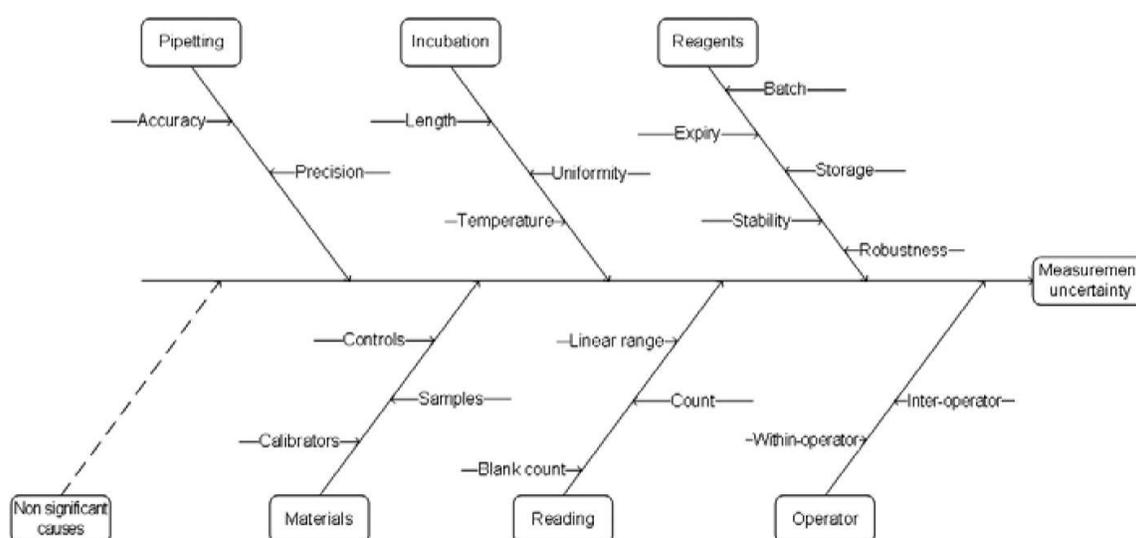
### 2.3.2. Identification of sources of uncertainty

#### a. Cause-and-effect diagram

The cause-and-effect diagram is valuable when the mathematical model is being formulated. It describes different groups or sources that may cause statistically significant contributions to measurement uncertainty (important sources of uncertainty), for example, sampling, storage conditions, instrument effects, reagent purity, assumed reaction stoichiometry, measurement conditions, sample effects, computational effects, blank correction, operator effects, and random effects (entry 6.7 of [9], entry Stage IV of [31]).

The groups of causes diverge in each assay, and the subcauses could be different even for the same commercial assays in different laboratories (i.e., some subcauses could cause significant uncertainty in one medical laboratory, and they should be expressed in mathematical module but, in a different medical laboratory, may not cause significant uncertainty and should be not expressed). This diagram should be carefully formulated for each measurement uncertainty

evaluation, and it could be a complex task. The diagram should not include bias measurement because it is not allowed in the calculation of measurement uncertainty, but when bias is statistically significant the contribution of bias to uncertainty (bias uncertainty) should be included. It should be clear that the mathematical model must be carefully formulated to include important causes but also avoid duplications of sources, which would increase estimates of uncertainty. The skills to formulate a correct mathematical model are not common in medical laboratory scientists or researchers. **Figure 3** shows an example of a cause-and-effect diagram that identifies known significant sources of uncertainty for a certain test statistically. This diagram describes the groups of sources (or standard uncertainty intended for measurement of combined standard uncertainty) as well as the causes of those sources. The result has a measurement uncertainty associated. The groups of causes (and subcauses) that contribute to this measurement uncertainty were defined after a depth research that included equipment process (workflow), service manual, reagents literature, assay development guidelines, manufacturer perspective of significant uncertainty causes, and scientific journals (entry Stage IV of [31]).

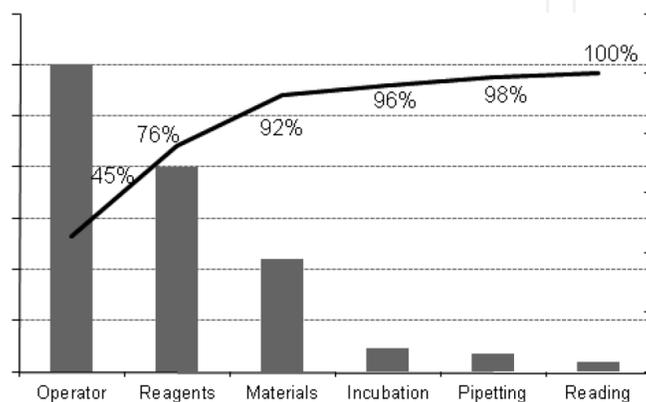


**Figure 3.** A cause-and-effect diagram for the determination of measurement uncertainty in a hypothetical medical laboratory test.

### b. Pareto diagram

The purpose of a Pareto diagram (bar diagram) is to highlight the critical sources of uncertainty, recognized as those with a significant contribution to measurement uncertainty. Usually, it is linked to the cause-and-effect diagram, allowing an easy detection of the primary sources of uncertainty. Typically, it is considered that approximately 80% of the effects come from 20% of the causes (Pareto principle also known as the “80-20 rule”), illustrating that a small part of the causes has the significant contribution to the effect. **Figure 4** shows an example of a Pareto diagram showing the “operator,” “material,” and “reagents” as the components with a significant contribution to the measurement uncertainty of a certain test, contributing to 92%

of the uncertainty. When it is using a modular approach, the selection of the uncertainty components that will be combined according to a model equation should be selected according to this principle. In this condition, the unmeasured components are statistically considered nonsignificant. Thompson and Ellison [32] claimed that the mostly determined measurement uncertainty was statistically significantly lower when compared to the standard deviation in reproducibility conditions, indicating a lower estimation of the uncertainty. The uncertainty not recognized was invoked to as “dark uncertainty” [32]. The mathematical models of empirical models already consider this principle, for what the Pareto diagram is useless for the laboratorian using these models.



**Figure 4.** A Pareto diagram for the determination of measurement uncertainty in a hypothetical medical laboratory test.

### 2.3.3. Partial derivative method

The partial derivative method was used during the years close to GUM publication, for what is recognized as the “GUM method.” The laboratorian must identify the critical sources of uncertainty and combine in a stochastic model equation. This required depth skills in chemistry, mathematics, and statistics, which are rarely available in medical laboratories, represent a high cost. Usually, the output of the modeling approach is an “uncertainty budget” summarizing the determination of the combined standard uncertainty and the uncertainty components. Pareto diagrams are also used to compare input values with the combined standard uncertainty. Usually, model equations are not developed in medical laboratories but in a reagent manufacturer’s laboratory during the test research and development (R&D). Even when medical laboratories are using “in-house” tests with published model equations, generally it is not practical for the determination of measurement uncertainty because, at least, it will require staff with advanced statistical skills of Uncertainty Approach, which is also not common in medical laboratories. It is easy to recognize the role of this approach in the manufacturer, because it is possible to identify and control the major sources of uncertainty, allowing the manufacturer to reduce certain uncertainty sources. This allows the new test to have a smaller measurement uncertainty (i.e., a high probability of the results cannot be significantly different from *in vivo* results). However, this role is not intended to a medical

laboratory, where the major uncertainty sources are associated principally with good laboratory practices (e.g., training of staff and storage conditions). The manufacturer considers this a useful method for assessing the influence of reference value uncertainties to the pooled uncertainty related to the final result of measurement.

Some other well-known disadvantages should be considered in the decision to use this method and in the interpretation of measurement uncertainty result. Preanalytical sources are not usually considered, including biological sources, which causes a misestimation of measurement uncertainty. In medical laboratory tests, measurement uncertainty is composed uniquely of analytical components. This is a critical lack of the accuracy of the determination. The application of inadequate model equation could lead to misestimation or overestimation; for example, unsuspected covariance could give rise to an overestimation of measurement uncertainty. Also, the statistical distribution might not be the same distribution than the one associated with method, which makes the evaluation inaccurate. The complex calculation of partial derivatives is time-consuming and expensive, requiring skilled statisticians. For further details about the partial derivative method, please refer to [5].

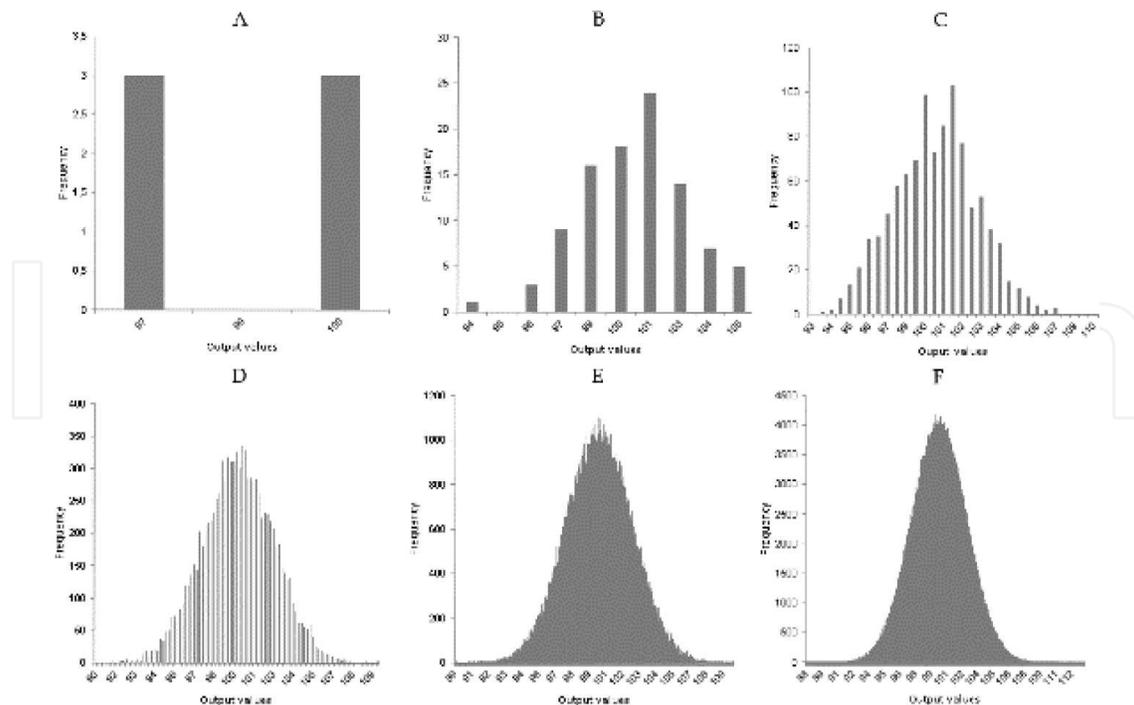
#### *2.3.4. Propagation of distributions by Monte Carlo simulation method*

The propagation of distribution method is an alternative to partial derivative method, and it is expected to be recommended in the next revision of GUM as the primary option when a modeling approach is considered. Currently, GUM proposes the propagation of uncertainties through the mathematical model measurement. The propagation of distribution substitutes the propagation of uncertainties method, correcting some obvious lacks of the propagation of uncertainties methods such as the linearity of the model and the normal distribution of the random variable representing the possible values of the measurand. This limitation is associated with the application of the partial derivative method in some tests. The propagation of distribution is determined commonly using Monte Carlo simulation method (Monte Carlo). BIPM published in 2008 a Supplement 1 to the GUM "Propagation of Distributions Using a Monte Carlo Method" [33], considering a correct application of Monte Carlo in the determination and evaluation of uncertainties. This method does not request a complex calculation of partial derivatives different from the partial derivative method. Monte Carlo simulation involves a random sampling of a probability distribution.

Contrary to the partial derivative method, Monte Carlo can be used in linear and nonlinear equations. It is used to simulate measurement results based on the input quantities of the probability density function (PDF). PDF describes the relative likelihood for a continuous random variable to take on a given value. Monte Carlo produces the propagation of PDF of the input through the mathematical model measurement, providing results in a PDF describing the values of the measurand consistent with the available data. Nonlinear model equations, asymmetric distributions, and other problems to the partial derivative method are not significant. Another difference to the propagation of uncertainties is the use of Welch-Satterthwaite formula in the estimation of expanded uncertainty, which is unnecessary. The limits of a symmetrical range of coverage are estimated by the values of the 2.5th and 97.5th percentiles for  $p=95\%$ . An extreme asymmetry could indicate the need to increase the  $M$ . Monte

Carlo is not adequate when the output distribution is not symmetrical. The confidence of the results is based on the demonstration of the model equation and the number of simulated measurements.

The effect of the number of simulated measurements  $M$  in the sampling error for estimates is major. Supplement 1 to the GUM request at least  $10^5$  repetitions/trials of the model equation to achieve a statistically acceptable result. At present, it is easy to obtain results with higher  $M$ , such as  $10^6$ , due to the available software and hardware. A simple test could be performed in Microsoft® Excel® 2007 or later release using =NORMINV(RAND();mean;standard\_dev) function and histograms. The reader will observe that  $M=10$  graph does not show a bell-shaped curve (see Figure 5, histogram A), simulations with  $M \geq 10$  show a bell-shaped curve (see Figure 5, histograms A–F), and  $M=10^5$  (see Figure 5, histogram E) and  $M=10^6$  (see Figure 5, histogram F) bell-shaped curves are closer from what is represented in the infinite number of samples. This task was not possible for most of the users a decade ago. Model equations with unreliable quantities can produce a sampling error that cannot be reduced increasing  $M$ . Monte Carlo estimates are considered reasonably accurate when repeated simulations deliver values of  $u_c(y)$  that do not diverge from each other in the second significant number. Some other known limitations of Monte Carlo are that the calculated uncertainties vary from one run to the next because of the intentionally random nature of the simulation and that is hard to identify the most significant contributions to the combined uncertainty without repeating the simulation.



**Figure 5.** Histograms from six simulations with data normally distributed: (A) 10, (B)  $10^2$ , (C)  $10^3$ , (D)  $10^4$ , (E)  $10^5$ , and (F)  $10^6$  trials.

Because, in the case of most medical laboratory tests, any model equation to describe the measurement is not known, the modeling approach cannot be applied in this case using partial derivative or Monte Carlo for what their role should be principally at the reagent manufacturer level. For further details about propagation of distributions by Monte Carlo simulation method, please refer to [33].

## 2.4. Empirical approaches

### 2.4.1. Single laboratory validation (including QC)

Single laboratory validation and QC, the major components of variance, can regularly be calculated by an “in-house” validation study pooled with internal QC using repeated determinations of reliable control samples. The evaluation of bias uncertainty is commonly done using a certified reference material (CRM) or comparing using a reference method. The laboratorian easily associates this approach to a method evaluation model. For further details about single laboratory validation for the evaluation of measurement uncertainty, please refer to Nordtest TR 537 [10] and ISO 11352 [34]. These technical guidelines use general approaches that can also be applied to results of medical laboratories’ tests.

The within-laboratory reproducibility uncertainty  $s_{Rw}$  is calculated by pooling the repeatability standard deviation  $s_r$  arising from replicate measurements of human samples and the intermediate standard deviation  $s_i$  from between runs as in Equation (15) (entry 4 of [10]).

$$s_{Rw} = \sqrt{s_r^2 + s_i^2} \quad (15)$$

Equation (16) (entry 5 of [10]) is used to compute the bias uncertainty  $u_b$  if a CRM is available. Bias  $b$  is the result of the mean deviation of measurement results of replicates from the corresponding reference value;  $s_b$  is the bias standard deviation, is the reference value standard uncertainty, and  $m$  is the number of replicate determinations:

$$u_b = \sqrt{b^2 + \left(s_b / \sqrt{m}\right)^2 + u_{c_{ref}}^2} \quad (16)$$

To obtain the combined standard uncertainty, the uncertainty due to precision and that due to bias are combined as in Equation (17) (entry 1.2.2 of [7]):

$$u_c(y) = \sqrt{s_{Rw}^2 + u_b^2} \quad (17)$$

This approach considers the within-laboratory reproducibility standard deviation according to two different methods:

(a) Method validation protocol intended for validating the precision of numerical quantity tests. It is recommended for most of the quantitative tests—the approach described in CLSI EP15-A3 protocol to evaluate the precision (and bias) [34]. A series of five analytical runs with three replicates per run is suitable. For further details about precision evaluation, please refer to [(entry Chapter 2 of [35]), (entry 1.2.2 of [7])].

(b) Using data from between-run variation. The data arise from long-term determination, usually using IQC data. For some tests, such as IQC sample's batches, the expiration date is only 1 month; for others, the expiration could be 1 year. Data coming from batches using a larger period are preferable, when available. The statistical power of the determination is critically influenced by this period. IQC should also be chosen when the mobile mean and variance are stable (test in control). For further details about precision evaluation, please refer to (entry 1.2.2 of [7]).

#### 2.4.2. Interlaboratory comparisons

In the interlaboratory validation approach, the principal sources of variability can often be assessed by interlaboratory studies performed and evaluated according to ISO 5725 [36]. This approach to estimating uncertainty is fully described in ISO 21748 [37].

The approach requires the determination of the between-laboratories reproducibility standard deviation  $s_R$  from the results of an interlaboratory trial according to ISO 5725. In a standardized method, these precision data are usually given in an Appendix. For further details about interlaboratory comparisons approach, please refer to (entry 1.2.3 of [7]).

#### 2.4.3. EQA (PT)

EQA programs are proposed to verify periodically the performance of a laboratory test based on data of a laboratory group using proficiency tests [36]. Medical laboratory could also use the comparison data to determine the measurement uncertainty. EuroLab Technical Report 1/2007 [7] presents an approach to laboratories to evaluate measurement uncertainty. Accordingly, the standard uncertainty measured from the results of the group's participants could to a common test be treated as an early evaluation of the combined standard uncertainty. The reproducibility standard deviation could be determined using the results of the group's laboratories with a combined standard uncertainty. For further details about EQA (PT) approach, please refer to (entry 1.2.4 of [7]).

### 2.5. Practical application

#### 2.5.1. Test and measurand

To exemplify the determination of measurement uncertainty, an *in vitro* chemiluminescent immunoassay was chosen to determine the concentration of antibodies to the hepatitis C virus (HCV), Abbott Prism® HCV (Abbott Diagnostics, Abbott Park, IL, USA) [38], in the Transmissible Agents Laboratory, Portuguese Institute of Blood and Transplantation (IPST; blood establishment/blood bank). The measurand is the immunoglobulin concentration in the serum

or plasma samples, which binds to solid-phase particles attached to recombinant antigens of the Core, NS3, NS4, and NS5 regions of the HCV genome.

The results of samples were derived from the photons emitted from the chemiluminescent reaction. They are expressed in a number of photons over a certain period. The number of photons is proportional to the number of antibody-antigen complexes released. This count is corrected by subtracting the number of photons in the dark. The equation to determine the test's results is defined by the manufacturer, validated using true negative and true positive human samples, and permitted by the national agencies. The "cutoff"  $n_c$  is estimated per analytical run using positive and negative calibrators considering , where  $n_c$  is the number of photons of "cutoff" value, is the number of photons of the negative calibrator expressed as the average result of the two lowest replicates out of three, is number of photons of the positive calibrator expressed as the average result of the two highest replicates out of three, and 0.55 is a factor critical for the false-negative rate. The results are verified in an ordinal scale using the ratio of sample corrected number of photons divided by "cutoff" corrected number of photons. Considering ratio value, the "cutoff" is a constant equal to 1. Ratios equal or higher than the "cutoff" are classified as positive, and ratios lower than the "cutoff" are classified as negative. A multiparametric positive sample is used at the end of each analytical run to control the run. This sample is supplied uniquely by the manufacturer and the acceptance criterion for anti-HCV is very large with ratio value from 1.02 to 6.00.

According to reagent manufacturer, "no qualitative performance differences were observed" for the immunoassay in controlled studies using anti-HCV nonreactive and reactive specimens when testing the following potentially interfering substances at the specified levels: bilirubin ( $\leq 20$  mg/dL), hemoglobin ( $\leq 500$  mg/dL), red blood cells ( $\leq 0.4\%$  v/v), triglycerides ( $\leq 3.000$  mg/dL) or protein ( $\leq 12$  g/dL) [38].

Other possible sources of false results are not stated in reagent insert and could be identified by review of the scientific literature, particularly Young's extensive summaries of the effects of preanalytic and analytic variables:

**a. Preanalytical components**

- Hepatitis C IgG antibodies (Serum no effect analytical): Stability of specimen: "No effect observed when serum stored at ambient temperature for 14 days, refrigerated for 31 days, or frozen for 2 months."
- HCV antibodies (Serum no effect analytical): Stability of specimen: "No effect observed when serum stored at ambient temperature for 14 days, refrigerated for 31 days, or frozen for 2 months."
- HCV core protein (Serum positive physiological): Test association: "In patients with chronic HCV infection, no significant positive correlation of  $r=0.081$  with serum tumor necrosis factor receptor p55; in patients with chronic HCV infection, no significant positive correlation of  $r=0.141$  with serum tumor necrosis factor receptor p75" [39].
- Disease, drug, and herbs and natural medicine effects

**b. There is no related association to disease effects [40].**

HCV antibodies (Serum positive):

- Hepatocellular carcinoma: “In 100 patients with hepatocellular carcinoma, 92 were positive for HCV antibodies.”
- Chronic hepatitis: “In 21 patients with chronic hepatitis, 18 were HCV antibody positive.”
- Cirrhosis of liver: “In 24 patients with cirrhosis of liver 22 were HCV antibody positive” [41].

There is no related association with natural medicine effects [42].

When the medical laboratory receives samples from patients having potentially interfering substances, it is desirable to make a new measurement using an assay that has not affected the potential cause or to request a collection of a sample during a period when the potential cause is not present. Correlation with past results, as well as other assays, could also be helpful. Results affected by preanalytic and analytic variables could be unrealistic, and the probability of incorrect clinical decision could be critical. In IPST, during the interview of candidates as blood donors, candidates with hepatocellular carcinoma and chronic or cirrhosis of the liver were rejected.

Because the model equation is unknown, uniquely empirical methods are used in this example. Interlaboratory comparisons were not determined for Abbott Prism® immunoassay results due to the method validation to be normally and uniquely intralaboratorial for screening immunoassays.

### 2.5.2. Bias uncertainty

Data for laboratory bias evaluation are lacking due to the unavailability of a CRM or a reference laboratory. The mean difference between two laboratories of the IPST was used as an approximation for bias estimation. The bias due to the different usage of reagent batches is already included in the within-laboratory reproducibility standard deviation. For further details about single laboratory validation approach, please refer to (entry 1.2.2 of [7]).

### 2.5.3. Results

**Table 1** summarizes the measurement uncertainty determinations for the intralaboratory and EQA approaches. It is evidenced that the standard deviation decreases according to the ratio, and the highest relative standard deviation is observed for samples close to the “cutoff” value. Standard deviation results under repeatability conditions used replicate testing from 0.5 to 1.5.

### 2.5.4. Discussion

The first intralaboratory approach (EP15) has precision estimates similar to what is claimed in a manufacturer’s precision study [38] tested in five runs over 5 days, where  $s_{Rw}$  is from 5.7%

(average ratio equal to 3.17) to 8.6% (average ratio equal to 0.17) under intralaboratory conditions (within-laboratory reproducibility). On the contrary, in the second intralaboratory approach using replicated results and between-run “cutoff” data, the estimated  $s_{Rw}$  is 16.5%. The contribution of precision of the replicate results was the major uncertainty component, as it has selected unique results with a ratio close to the “cutoff” value. The relative standard deviation is lower at higher ratios, as the EP15 results demonstrate it with an average ratio equal to 3.70.

Approach	$u_{Rw}$							Bias method	$u_b$	$u_c$	$U$
	Method	Ratio	$s_r$ (method)	$s_r$	$s_l$ (method)	$s_l$	$s_{Rw}$ (Equation 3)				
Intralab validation data <sup>a</sup>	EP15	2.2–	NA	NA	NA	NA	8.5%	Control sample measured at two laboratories <sup>a</sup>	6.0%	10.4%	21%
	Validation and QC data	0.5– 2.8	Replicates from ratio 0.5 to 1.5	14.4%	Between-run “cutoff” <sup>b</sup>	8.5%	16.5%				
EQA <sup>c</sup>		6.2– 10.1	NA	NA	NA	NA	NA	NA	NA	13.9 %	28 %

NA, not applicable.

<sup>a</sup>Sample Accurun-1 Series 2400 batch no. 10017751 (Seracare Life Sciences, Inc., Milford, MA, USA); average ratio is equal to 3.70 (15 replicate measurements per run).

<sup>b</sup>Average ratio equal to 1 (299 results per run).

<sup>c</sup>For this approach, the obtained reproducibility standard deviation is set as the combined uncertainty (sample UK NEQAS no. 9316); average ratio is equal to 2.67 (results from nine laboratories).

**Table 1.** Estimated measurement uncertainty values for the Abbott Prism<sup>®</sup> HCV method using data from intralaboratory and EQA approaches.

The between-run data are more reliable due to much higher degrees of freedom of this estimation. This is evidence comparing the usage of EP15 precision data and the usage of the between-run data. **Table 1** shows that the between-run data are derived from approximately 300 results determined during a long period, and EP15 data were calculated uniquely using 15 results for 5 days from December 16 to 20, 2014.

The EQA results from nine laboratories tested in the Abbott Prism<sup>®</sup> gave an estimate of a relative expanded uncertainty of 28% [35] compared to the second intralaboratory approach with an estimated uncertainty of 36%. The higher measurement uncertainty was caused principally by the heterogeneity of the group’s laboratories. It used a sample with an average ratio equal to 8.2, for what if it used a sample with a ratio close to the “cutoff” value, the

measurement uncertainty should be higher, closer to the second intralaboratory approach percentage.

Supposedly, it is an alternative to the between-run precision using “cutoff” raw data for the use of a long-term IQC data, given that the whole analytical process is covered. Nevertheless, using a QC sample with an average ratio equal to 1 makes it a second suggestion. For example, the usage of IQC data from March 23, 2010 to June 11, 2011 determined from sample Seracare Accurun-1 Series 2400 batch no. 116406 (Seracare Life Sciences Inc. Milford (MA)) was not a primary choice considered the between-run “cutoff” data because the average ratio was statistically significantly higher. The average ratio of 293 samples was equal to 2.42, and the intermediate standard deviation was equal to 11.1%.

The example should not be understood as applicable to all tests in the medical laboratory. Although the mathematical models could be applied to any quantitative test, some adaptations must be done in different tests. For example, in clinical chemistry, one of the major concerns is the lack of biological components when determining measurement uncertainty. This problem happened already with the Westgard et al. [43] total analytical error (TAE) [43], which is the sum of the absolute value of bias with a  $k$  multiplied by the standard deviation. Fraser [26] expanded the TAE concept to the total biological error ( $TE_{ba}$ ) combining not only analytical error components but also biological precision and accuracy. These models are based in the Error Approach (also recognized as the Traditional Approach or True Value Approach). The Australian NPAAC guideline [12] considers not only analytical sources of precision but also the individual biological variation ( $CV_1$ ) when it is known using GUM empirical estimations and Fraser’s proposal. The Westgard QC provides quality requirements tables with  $CV_1$  for most of the chemistry tests [44]. For further details about the empirical determinations of measurement uncertainty combining  $CV_1$ , please refer to [12].

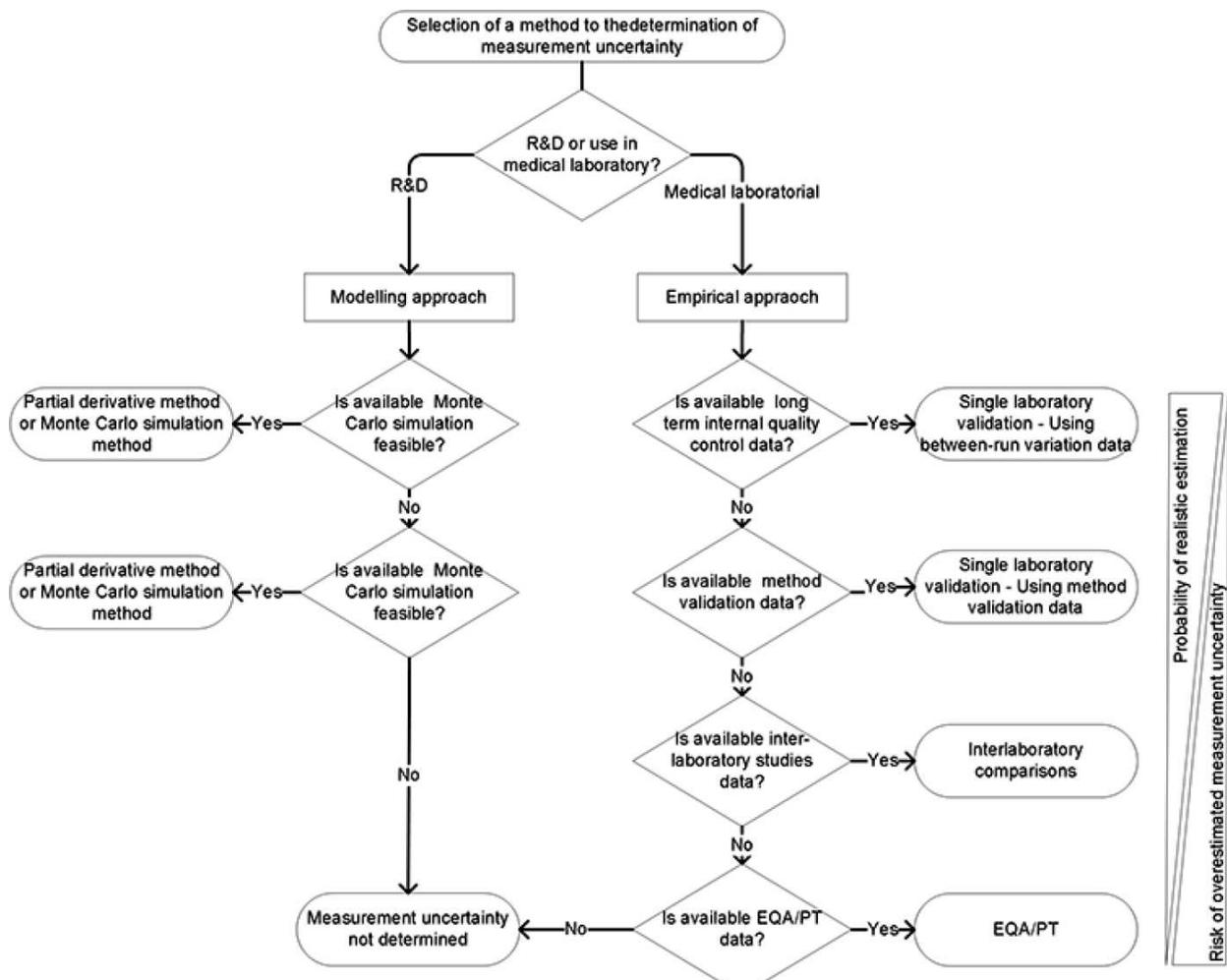
Although the determination of measurement uncertainty is practical in a medical laboratory, as it was already demonstrated [5], there is a serious lack of consensus on the use of the Uncertainty Approach, and the medical laboratory staff rarely understands it. Although TAE does not fulfill part of the Uncertainty Approach principles (i.e., it cannot be a model for the determination of measurement uncertainty), it is accepted in medical laboratory practice and recognized by some as a role analogous to measurement uncertainty [15, 45].

Medical laboratory results affect 60% to 70% of clinical decisions, affecting 100% of validation of blood donations, cells, and tissues. Consequently, results with higher measurement uncertainty have a significant probability of being unrealistic arising from a high risk of the uncorrected clinical decision. Measurement uncertainty has been demonstrated to be a tool to verify the level of confidence in medical laboratory results. Results with large measurement uncertainty intervals have a significant chance to be distant from the true value (i.e., *in vivo* value). Therefore, measurement uncertainty could also be used as a method validation estimation. The laboratorian could consider it as analogous to measurement uncertainty but not confusing for both concepts, differing principally in the use of bias.

### 3. Recommended models in medical laboratories

According to what was explained in Section 2.3, the use of modeling approach is intended principally to the R&D of medical laboratory tests. Its use in medical laboratories is complex and expensive and the measurement uncertainty results could be misestimated because not all uncertainty components are determined, different what happens using empirical data. The reagent manufacturer should choose the Monte Carlo simulation instead of partial derivative method because the propagation of distribution result is considered more reliable.

The application of empirical models is recommended for medical laboratory tests, as most of its theory and practice are already known and the results are more realistic when compared to the modeling approach. However, the laboratorian should select the model according to the chance to produce more realistic results. **Figure 6** represents in a flow chart the steps to the selection of measurement uncertainty models in medical laboratories. The first model to select is the intralaboratory approach using repeatability and between-run precision using long-term data. This is usually applied to tests used in the long-term. The second choice is the intrala-



**Figure 6.** Flow chart to select a model for the determination of measurement uncertainty in a medical laboratory.

laboratory approach using EP15 data. It is expected to be used in initial uncertainty estimation in a brand-new test. Interlaboratory comparisons are the third choice, and it is principally used when the comparator test is external. The fourth and last option is the EQA (PT). Due to the heterogeneity of the group's data, this estimation should be carefully evaluated. There is a high risk of an unrealistic estimation with the overestimation of measurement uncertainty and in some cases could be useless.

## Author details

Paulo Pereira

Address all correspondence to: paulo.pereira@ipst.min-saude.pt

Quality Management Department, Portuguese Institute of Blood Transplantation, Lisbon, Portugal

## References

- [1] Bureau International des Poids et Mesures. Rapport BIPM-80/3, Report on the BIPM Inquiry on Error Statements. Sèvres: BIPM; 1980. Retrieved from: <http://www.BIPM.org/utills/common/pdf/rapportBIPM/1980/03.pdf>. Accessed: January 4, 2016.
- [2] Giacomo P. News from the BIPM. *Metrologia* 1982; 18(1): 43–44.
- [3] Giacomo P. News from the BIPM. *Metrologia* 1987; 24(1): 45–50.
- [4] Bureau International des Poids et Mesures. JCGM 200:2012 International Vocabulary of Metrology—Basic and General Concepts and Associated Terms, 2008 Version with Minor Corrections. 3rd ed. Sèvres: BIPM; 2012. Retrieved from: [http://www.BIPM.org/utills/common/documents/jcgm/JCGM\\_200\\_2012.pdf](http://www.BIPM.org/utills/common/documents/jcgm/JCGM_200_2012.pdf). Accessed: January 4, 2016.
- [5] Bureau International des Poids et Mesures. Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008, GUM 1995 with Minor Corrections. Sèvres: BIPM; 2008. Retrieved from: [http://www.BIPM.org/utills/common/documents/jcgm/JCGM\\_100\\_2008\\_E.pdf](http://www.BIPM.org/utills/common/documents/jcgm/JCGM_100_2008_E.pdf). Accessed: January 4, 2016.
- [6] Pereira P, Magnusson B, Theodorsson E, Westgard J, Encarnação P. Measurement uncertainty as a tool for evaluating the “grey-zone” to reduce the false negatives in immunochemical screening of blood donors for infectious diseases. *Accred Qual Assur* 2015; 21(1): 25–32. DOI: 10.1007/s00769-015-1180-x
- [7] European Federation of National Associations of Measurement, Testing and Analytical Laboratories. Technical Report No. 1/2007 Measurement Uncertainty Revisited:

- Alternative Approaches to Uncertainty Evaluation. EuroLab; 2007. Retrieved from: <http://www.eurolab.org/documents/1-2007.pdf>. Accessed: January 4, 2016.
- [8] Dybkaer R. From total allowable error via metrological traceability to uncertainty of measurement of the unbiased result. *Accred Qual Assur* 1999; 4(9): 401–405. DOI: 10.1007/s007690050395
- [9] Ellison SLR, Williams A. *Quantifying Uncertainty in Analytical Measurement*. 3rd ed. Eurachem/CITAC; 2012. Retrieved from: [https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012\\_P1.pdf](https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf). Accessed: January 4, 2016.
- [10] Magnusson B, Näykk T, Hovind H, Krysell M. *NordTest NT TR 537 Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories*. 31st ed. Oslo: Nordic Innovation; 2011. Retrieved from: <http://www.nordtest.info/index.php/technical-reports/item/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-3.html>. Accessed: January 4, 2016.
- [11] SYKE. MUKIT—Measurement Uncertainty Kit Finnish Environmental Institute; 2014. Retrieved from: [http://www.syke.fi/en-us/Services/Calibration\\_services\\_and\\_contract\\_laboratory/MUKIT\\_\\_Measurement\\_Uncertainty\\_Kit](http://www.syke.fi/en-us/Services/Calibration_services_and_contract_laboratory/MUKIT__Measurement_Uncertainty_Kit). Accessed: January 4, 2016.
- [12] National Pathology Accreditation Advisory Council. *Requirements for the Estimation of Measurement Uncertainty*. Barton, ACT: NPAAC; 2007. Retrieved from: [http://www.health.gov.au/internet/main/publishing.nsf/Content/B1074B732F32282DCA257BF0001FA218/\\$File/dhaeou.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/B1074B732F32282DCA257BF0001FA218/$File/dhaeou.pdf). Accessed: January 4, 2016.
- [13] Clinical and Laboratory Standards Institute. *EP29-A Expression of Measurement Uncertainty in Laboratory Medicine*. Wayne, PA: CLSI; 2012.
- [14] Westgard J. Update on Measurement Uncertainty: New CLSI C51A Guidance. Retrieved from: <http://www.westgard.com/CLSI-c51.htm>. Accessed: January 4, 2016.
- [15] Westgard J. Managing quality vs. measuring uncertainty in the medical laboratory. *Clin Chem Lab Med* 2012; 48(1): 31–40. DOI: 10.1515/CCLM.2010.024
- [16] International Organization for Standardization. *ISO/PDTS 25680 Medical Laboratories – Calculation and Expression of Measurement Uncertainty*. Geneva: ISO; 2008.
- [17] Pereira P, Westgard J, Encarnação P, Seghatchian J, Sousa G. Quality management in the European screening laboratories in blood establishments: a view on current approaches and trends. *Transfus Apher Sci* 2015; 52(2): 245–251. DOI: 10.1016/j.transci.2015.02.014
- [18] Pereira P, Westgard J, Encarnação P, Seghatchian J, de Sousa G. The role of uncertainty in results of screening immunoassays in blood establishments. *Transfus Apher Sci* 2015; 52(2): 252–255. DOI: 10.1016/j.transci.2015.02.015
- [19] Pereira P, Westgard J, Encarnação P, Seghatchian J. Evaluation of the measurement uncertainty in screening immunoassays in blood establishments: Computation of

- diagnostic accuracy models. *Transfus Apher Sci* 2015; 52(1): 35–41. DOI: 10.1016/j.transci.2014.12.017
- [20] Kimothi S. *The Uncertainty of Measurements*. Milwaukee, WI: ASQ; 2002.
- [21] Hibbert D. Measurement uncertainty. In: *Quality Assurance for the Analytical Chemistry Laboratory*. Oxford: Oxford University Press; 2007, pp. 161–202.
- [22] Struik D. *A Source Book in Mathematics 1200–1800*. Cambridge, MA: Harvard University Press; 1969.
- [23] Kragten J. Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique (Tutorial review). *Analyst* 1994; 119 (10): 2161–2165. DOI: 10.1039/AN9941902161
- [24] European Co-operation for Accreditation. EA-4/02 M: 2013 Expression of the Uncertainty of Measurement in Calibration. 2nd ed. Europe: EA; 2013. Retrieved from: <http://www.european-accreditation.org/publication/ea-4-02-m-rev01-september-2013>. Accessed: January 4, 2016.
- [25] Petersen P, Jørgensen L, Jensen E, Kynde K, Brandslund I, et al. Er GUM skadelig? - Eller blot overflødig? (Is GUM harmful? - Or become redundant?). *Klinisk Biokemi: Norden* 2003; 15(2): 27–31.
- [26] Fraser C. *Biological Variation: From Principles to Practice*. Washington, DC: AACC Press; 2009.
- [27] Vesper H, Thienpont L. Traceability in laboratory medicine. *Clin Chem* 2009; 55(6): 1067–1075. DOI: 10.1373/clinchem.2008.107052
- [28] Dybkaer R. Setting quality specifications for the future with newer approaches to defining uncertainty in laboratory medicine. *Scand J Clin Lab Invest* 1999; 59(7): 579–584. DOI: 10.1080/00365519950185391
- [29] Magnusson B, Ellison SL. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. *Anal Bioanal Chem* 2008; 390(1): 201–213. DOI: 10.1007/s00216-007-1693-1
- [30] M. Pepe. *Oxford Statistical Science Series 28. The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press; 2003.
- [31] Krouwer J. *Assay Development and Evaluation*. Washington, DC: AACC Press; 2002.
- [32] Thompson M, Ellison S. Dark uncertainty. *Accred Qual Assur* 2011; 16(10): 483–487. DOI: 10.1007/s00769-011-0803-0
- [33] Bureau International des Poids et Mesures. JCGM 101 Evaluation of Measurement Data – Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement” – Propagation of Distributions using a Monte Carlo Method. Sèvres: BIPM; 2008.

Retrieved from: [http://www.BIPM.org/utis/common/documents/jcgm/JCGM\\_101\\_2008\\_E.pdf](http://www.BIPM.org/utis/common/documents/jcgm/JCGM_101_2008_E.pdf). Accessed: January 4, 2016.

- [34] International Organization for Standardization. ISO 11352 Water Quality—Estimation of Measurement Uncertainty Based on Validation and Quality Control Data. Geneva: ISO; 2012.
- [35] Clinical and Laboratory Standards Institute. EP15-A3 User Verification of Precision and Estimation of Bias. 3rd ed. Wayne, PA: CLSI; 2014.
- [36] International Organization for Standardization. ISO 5725-2 Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method. Geneva: ISO; 1994.
- [37] International Organization for Standardization. ISO 21748 Guidance for the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation. Geneva: ISO; 2010.
- [38] Abbott Diagnostics Division. Abbott Prism HCV. Code: 84-5342/R9. Wiesbaden: Abbott Laboratories; 2008.
- [39] Young D. Effects of Preanalytical Variables on Clinical Laboratory Tests. 3rd ed. Washington, DC: AACC Press; 2007.
- [40] Young D. Effects of Disease on Clinical Laboratory Tests. Volumes 1 & 2. 4th ed. Washington, DC: AACC Press; 2001.
- [41] Young D. Effects of Drugs on Clinical Laboratory Tests. Volumes 1 & 2. 5th ed. Washington, DC: AACC Press; 2000.
- [42] Narayanan S, Young D. Effects of Herbs and Natural Products on Clinical Laboratory Tests. Washington, DC: AACC Press; 2007.
- [43] Westgard J, Neil Carey R, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974; 20(7): 825–833.
- [44] Westgard J. Quality Requirements. Retrieved from: <http://www.westgard.com/quality-requirements.htm>. Accessed: January 4, 2016.
- [45] Pereira P, Westgard J, Encarnação P, Seghatchian J. Analytical model for calculating indeterminate results interval of screening tests, the effect on seroconversion window period: A brief evaluation of the impact of uncertain results on the blood establishment budget. *Transfus Apher Sci* 2014; 51(2): 126–131. DOI: 10.1016/j.transci.2014.10.004