Upsala J Med Sci 98: 215–220, 1993

3. Truth, Accuracy, Error and Uncertainty

René Dybkær
Department of Clinical Chemistry Frederiksberg Hospital, Frederiksberg, Danmark

The year 1993 should witness the final version of some very important documents on metrology and statistics that are revising our understanding and vocabulary concerning the components of a result of measurement and its uncertainty (see Bibliography). The quite voluminous draft texts are neither easy reading nor fully harmonized, and they represent different philosophical approaches. The following may serve as an introduction to further study. A vocabulary, given in a separate paper, should ease the assimilation of any unfamiliar concepts and terms.

This discussion is limited to those properties of systems that can be expressed on a difference scale or a ratio scale, comprising the **measurable quantities,** hereinafter called quantities, where a value is a numerical value multiplied by a unit.

Due to residual uncertainty of any definition of a quantity in the real world, especially prominent in biology, the true magnitude of a quantity can only be represented by a **distribution of true values**, but definitions of theoretical metrological concepts usually select a true value, μ, as a representation of truth; in practice, a **conventional true value,** $\hat{\mu}$, is substituted.

A quantity being subjected to **measurement** is called a **measurand**, Y, and it can be considered as an **output quantity**, that is a function of a set of **input quantities,** X_g - such as calibrator quantities and influence quantities - so that $Y = H(X_1, X_2, ..., X_p)$, where Y and X_g are all considered to be random variables. Observations on any X_g yield an estimate, and all such estimates allow calculating an observed value of Y according to the function H. One or more observed values yield a **result of a measurement,** y, as described by the **measurement procedure.** It is assumed that any result has been corrected for known error.

The traditional analysis of error defines **error of a result** of a measurement as the difference $y - \mu$. Repeated measurements of the same measurand yield different results due to the variable error that forms a distribution having a location corresponding to the **systematic error** and a dispersion caused by the **random error.** The distribution of errors may be said to be convoluted with the true distribution to give the distribution of results for the measurand.

The systematic error is the expectation of all results minus a true value, $E(Y) - \mu$, whereas the random error is $y - E(Y)$. In practice $\bar{y} - \hat{\mu}$ and $y - \bar{y}$ respectively are used as estimates of the theoretical values.

The **accuracy** of a result is now defined as the closeness of agreement between the result and a true value; it therefore depends on a combination of random and systematic errors, not only on the latter. For the closeness of agreement between the expectation of a large series of results and a true value, the term **trueness** of results has been coined, whereas **precision** as before is the closeness of agreement between independent results under prescribed conditions. The three concepts can only be expressed on a few-value ordinal scale of measurement such as (poor average good).

The inverse measures of trueness and precision in the form of population parameters or their statistical estimates are **bias of results,** $E(Y) - \mu$ or $\bar{y} - \hat{\mu}$, and **variance of the distribution of results,** σ^2 or s^2, respectively. The variance is often replaced by the **standard deviation,** σ or s, or the **coefficient of variation,** $\sigma/E(Y)$ or $s/\bar{y}$.

The **theoretical statistical linear model of a result,** y, may be stated as

$y = E(Y) + B + S + M + e$

where

$E(Y) = m$ is the **overall expectation of all results** from all laboratories obtained by a given measurement procedure for a given measurand; it is equal to $\mu + \delta$ where δ is the inherent (unknown) **bias of the measurement procedure** under any set of precision conditions;

B is the **laboratory** *(i)* **deviation** under given precision conditions; it is equal to $E(Y_i) - E(Y)$;

another parameter may be defined on the above variables, namely

Δ which is the **laboratory bias of results;** it is equal to $E(Y_i) - \mu$ or $\delta + B$;

S is an **aberrant-sample bias of result** caused by an unusual analyte form or an unexpected influence quantity; its detection may require a reference measurement procedure, but this bias is not considered further here;

M is the **bias caused by an undetected mistake;** it may be detected by an outlying result, but is not considered further here;

e is a **random error of a result** occurring in every single result under repeatability conditions (r), *i.e.* intra-run; it is equal to $y - E(Y_i|\mathrm{r})$; the distribution of e is assumed to be approximately Gaussian with an expectation of zero.

Each of the parameters mentioned above has a variance. σ_μ^2 is often neglected and σ_δ^2 is often taken to be zero. Within a laboratory, i, the variance of the random error e is called the **within-laboratory repeatability variance**, σ_w^2, and if its magnitude does not vary much between laboratories, a mean is calculated, $\overline{\sigma_w^2}$, and called the **repeatability variance**, σ_r^2. The laboratory deviation, B, varies with the precision conditions as follows.

Under **repeatability conditions**, r, (one laboratory, operator, measuring system, and run) B is called the **laboratory component of bias of results**, $B_{i,\mathrm{r}}$, and considered to be constant, *i.e.* with a variance of zero. Consequently, the total variance of Y is σ_w^2.

Under **intermediate precision conditions**, I, where one or more sources of random variation are added in either or all of the groups time, operator, and calibration of equipment and reagents, the term B may be divided into elements that remain constant and other that become variables with expectations and variances; for, e.g., element 2 we have $E(B2_i|\mathrm{I})$ and $e2_{i,\mathrm{I}}$ respectively. The former changes the constant B_r and thus $E(Y_i|\mathrm{r})$, the latter adds to the total random variance so that $\sigma^2_{y\,i,\mathrm{I}} = \sigma_w^2 + \sigma^2_{e2_{i,\mathrm{I}}}$. With each new source of variation being introduced, the residual B is being further transformed, acquiring another value, $B_{i,\mathrm{I}}$, where I has to be specified, and the total variance is increased.

Under **reproducibility conditions**, R, several laboratories are involved in applying the same measurement procedure, increasing the number of sources of variation. The

individual laboratory component of bias, $B_{i,r}$ or $B_{i,I}$, is substituted by the **between-laboratory variation,** B_R. With a sufficient number of laboratories, its distribution is assumed to be approximately Gaussian with an expectation, $E(B|R)$, equal to zero and a **between-laboratory variance,** $\sigma_{B_R}^2$, also symbolized σ_L^2. The **overall expectation** of the population of results from all laboratories may be taken to be $E(Y) = \mu + \delta$ and the total variance of the random error, called the **reproducibility variance,** is $\sigma_R^2 = \sigma_L^2 + \sigma_r^2$.

The effects of additional sources of variation can be seen and treated analogously when changes in measurement procedure, method of measurement, and even principle of measurement are introduced during laboratory proficiency testing or certification exercises for reference materials.

The description above concerns populations and their parameters, θ. In practice, sample estimators, $\hat{\theta}$, are used.

The classical description of **uncertainty of measurement** of a result was based on an analysis of error with systematic errors being summed linearly and random errors quadratically; the first sum was added to the positive square root of the latter.

The newer procedure establishes an **uncertainty budget** giving an estimate of the uncertainty - a socalled **standard uncertainty,** u - for each input quantity. Some are estimated as standard deviations, $u = s$, calculated from the observed number fraction distribution of values or stated on the certificate of a calibrator (type A evaluation). Others are estimated non-statistically, but are also characterized by standard deviations, assuming some specified type of distribution, such as for influences of environment, algorithms, inhomogeneous sample, imperfect measurement procedures (type B evaluation). The former group corresponds to random errors, the latter to both random and systematic ones. The **combined standard uncertainty,** u_c, is calculated from the sum of the estimated variances, $u_{x_g}^2$, with due consideration of any correlation between input quantities, X_g, and the type of function, H, providing the value of the output quantity. Multiplying the combined standard uncertainty, u_c, by a **coverage factor,** k, yields an **expanded uncertainty,** U. This allows an interval to be calculated, $[(y - k \cdot u_c); (y + k \cdot u_c)]$, such that a stated number fraction of all intervals obtained by calculation from repeated measurement results will contain a true value of the measurand; the

number fraction (or probability) depends on k.

In any case, the conditions under which a result and its uncertainty are obtained must be specified; terms such as "inter-run" and "reproducibility conditions" need explanation in view of different definitions being offered by various standards.

BIBLIOGRAPHY

Abbreviations of international organizations

BIPM	International Bureau of Weights and Measures
IEC	International Electrotechnical Commission
IFCC	International Federation of Clinical Chemistry
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
IUPAP	International Union of Pure and Applied Physics
OIML	International Organization of Legal Metrology

BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML. International vocabulary of basic and general terms in metrology (VIM). 2 ed. Geneva: ISO, 1993 (in press).

ISO. Statistics - Vocabulary and symbols. Part 2. Statistical quality control. ISO/DIS 3534-2. Geneva: ISO, 1986:iii + 30 pp.

ISO. Statistics - Vocabulary and symbols. Part 1. Probability and general statistical terms. ISO/DIS 3534-1. Geneva: ISO, 1990:80 pp.

ISO. Accuracy (trueness and precision) of measurement methods and results. Part 1. General principles and definitions. ISO/DIS 5725-1. Geneva: ISO, 1991:24 pp.

Idem. Part 2. A basic method for the determination of repeatability and reproducibility of a standard measurement method. ISO/DIS 5725-2. Geneva: ISO, 1991:67 pp.

Idem. Part 4. Basic methods for estimating the trueness of a test method. ISO/DIS 5725-4. Geneva: ISO, 1991:35 pp.

Idem. Part 3. Intermediate measures of the precision of a measurement method. ISO/DIS 5725-3.2. Geneva: ISO, 1992:49 pp.

BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML. Guide to the expression of uncertainty in measurement. Geneva: ISO, 1993 (in press).

Correspondence:
René Dybkær
Department of Clinical Chemistry
Frederiksberg Hospital
DK-2000 Frederiksberg, Danmark