Medical Utility Frequency
Kicab Castañeda-Méndez

Judging the medical acceptability and comparing the medical utility of laboratory methods requires standards for reproducibility with regard to medical performance. A completely general and flexible performance standard for reproducibility consists of the reproducibility criterion, the allowable error, and the concentration value. Neither precision, accuracy, nor total error correctly addresses this standard. The medical utility frequency of a method does do so, however, its use making method evaluation and comparison simple, completely flexible, and exact.

Additional Keyphrases: method evaluation · statistics · quality control

The user of a diagnostic device is concerned about its reproducibility (1); how frequently a result will occur within a certain tolerance about the true value. Reproducibility being a function of precision and accuracy (2), method evaluation and comparison have traditionally focused on these two characteristics.

For many uses it does not matter whether a deviation from the true value is due to imprecision (RE) or inaccuracy (SE). Thus the concept of total analytical error (TE) (3, 4), as a function of random and systematic error, was derived. By applying total error to medical performance standards, defined by "...two terms: a) the allowable analytical error (E_A), which is defined as a 95 percent limit of allowable error, and b) the medical decision level (X_C)" (5, IV), the ability to characterize diagnostic devices was advanced significantly (5).

The "95 percent limit" is considered the minimum proportion, P, of a method's results that must occur within E_A about X_C. Thus it has been recommended that TE = SE + 1.96RE (3-6) for a symmetrical allowable error about the medical decision level, where 1.96RE corresponds to the 97.5th percentile of the gaussian distribution.

The "performance standard" states that a method is acceptable if, and only if, at least 95% of its distribution is within E_A of X_C. Method evaluation and comparison can then be accomplished with two decision rules (4-6).

Rule 1: A method is acceptable if, and only if, TE < E_A.
Rule 2: Method 1 is better than Method 2 if, and only if, TE_1 < TE_2.

General Performance Standards

The performance standard actually consists of three terms, not two: (a) the concentration value, X_C, (b) the allowable error, E_A, and (c) the reproducibility criterion, P. Fixing P at 95% detrimentally reduces the number of terms considered to two. Because of different uses of medical methods (7-12), P = 95% may not always be desirable; e.g., P = 99% may be preferred (6). For certain changes in various factors, e.g., analyte, state of the art, medical criticality, concentration value, seriousness of misdiagnosis, possible treatment, etc., it might be more appropriate to have an asymmetrical allowable error, let P \neq 95%, and assume the distribution of results to be non gaussian.

All these possibilities imply a "general performance standard":

A minimum proportion, P, of a method's results must fall within E_A = (X_C - E_A) of X_C of E_A, where P is the reproducibility criterion, E_A is the medical allowable error, and X_C is the value for the analyte concentration.

Because total error specifies a symmetrical allowable error, a gaussian distribution, and a fixed reproducibility criterion, TE is not applicable. However, the frequency of a method's results that occur within E_A of X_C is applicable for general performance standards. I call this the medical utility frequency, F_U. The decision scheme for method evaluation and comparison using the medical utility frequency and the general performance standard consists of two rules.

Rule 3: For a fixed X_C and E_A, a method is acceptable if, and only if, F_U \geq P.
Rule 4: For a fixed X_C and E_A, Method 1 is better than Method 2 if, and only if, F_U_1 > F_U_2.

Utility Frequency vs Total Error

How does the medical utility frequency compare with total error? F_U does not require the gaussian assumption. The binomial distribution (13) of F_U for any distribution of results makes the concept applicable for any distribution, reproducibility criterion, and asymmetrical allowable error.

When a gaussian distribution is supportable (14), estimating F_U is also simple, with use of invariant formulas, regardless of the value of P.3

Unfortunately, total error cannot be made applicable by removing the restrictions. As an approximation (15-17), even with the restrictions, this approach can yield incorrect comparisons and evaluations.

For example, let P = 95%, E_A = 12 symmetrically about X_C = 100 and the distribution of each method be gaussian with mean, \mu, and standard deviation, \sigma, denoted N(\mu, \sigma^2). Assume N(105, 4^2) characterizes Method 1 (Figure 1). Then TE_1 = 4 \sigma = 1.96 \sigma = 12.84 > E_A = 12. By Rule 1, Method 1 is unacceptable; but this conclusion is incorrect because 95% of its results are within E_A of X_C.

It can be shown that any method characterized by N(\mu, \sigma^2) with 1.645 \sigma < (E_A - SE) < 1.96 \sigma and only one tail beyond E_A, will have more than 95% of its results within E_A

---

1 Nonstandard abbreviations: TE, total analytical error; RE, random analytical error or imprecision; SE, systematic analytical error or inaccuracy; X_C, analyte concentration value; E_A, medical allowable error; P, reproducibility criterion as a percent or proportion; F_U, medical utility frequency; N(\mu, \sigma^2), gaussian distribution with mean \mu and standard deviation \sigma.
2 Received September 17, 1966; accepted October 13, 1966.

---

1 For a gaussian distribution, the limits are converted to z-scores and F_U is determined by using a standard gaussian distribution table. Only a lower confidence interval needs to be estimated; if it is greater than P, then the method is acceptable. A lower confidence interval estimate can be approximated rather well for large sample sizes by using F_U - z_{1-\alpha} \sqrt{F_U (1-F_U)/n}, where F_U is the single-value estimate.
Because of various factors, e.g., criticality of concentration value, analyte, user, method, state of the art, treatment, etc., a more general performance standard is recommended that would allow for different reproducibility criteria, asymmetrical allowable errors, and nongaussian distributions of a method's results.

Reproducibility is a function of both precision and accuracy; separately they do not correctly represent reproducibility. Total error, by being a fixed combination of precision and accuracy, is an approximation of the function. This makes it inapplicable for general performance standards and often too inaccurate for the specific performance standards to which it was intended to apply.

The medical utility frequency of a method is its reproducibility. It is always applicable for any performance standard, regardless of the distribution of the results, the reproducibility criterion, or whether the allowable error is asymmetrical. Exact formulas exist for the medical utility frequency, it is a simpler concept, and the decision-rules are easily understood and consistent, because they characterize the method's reproducibility directly.

References

222 CLINICAL CHEMISTRY, Vol. 33, No. 2, 1987