Pharmaceutical Research, Vol. 24, No. 6, June 2007 (© 2007)
DOI: 10.1007/s11095-007-9242-3

Research Paper

A Total Error Approach for the Validation of Quantitative Analytical Methods

David Hoffman[1,2,3] and Robert Kringle[1]

Received November 3, 2006; accepted January 9, 2007; published online March 21, 2007

Purpose. Typical acceptance criteria for analytical methods are not chosen with regard to the concept of method suitability and are commonly based on ad-hoc rules. Such approaches yield unknown and uncontrolled risks of accepting unsuitable analytical methods and rejecting suitable analytical methods. This paper proposes a formal statistical framework for the validation of analytical methods, which incorporates the use of total error and controls the risks of incorrect decision-making.
Materials and Methods. A total error approach for method validation based on the use of two-sided β-content tolerance intervals is proposed. The performance of the proposed approach is compared to the performance of current ad-hoc approaches via simulation techniques.
Results. The current ad-hoc approaches for method validation fail to control the risk of incorrectly accepting unsuitable analytical methods. The proposed total error approach controls the risk of incorrectly accepting unsuitable analytical methods and provides adequate power to accept truly suitable methods.
Conclusion. Current ad-hoc approaches to method validation are inconsistent with ensuring method suitability. A total error approach based on the use of two-sided β-content tolerance intervals was developed. The total error approach offers a formal statistical framework for assessing analytical method performance. The approach is consistent with the concept of method suitability and controls the risk of incorrectly accepting unsuitable analytical methods.

KEY WORDS: analysis of variance; bioanalytical assay; method validation; tolerance interval; total error.

INTRODUCTION

Analytical methods are utilized throughout the drug development process and the manufacturing of drug substances and drug products. Analytical results are used for decision-making regarding, for example, bioavailability, bioequivalence, shelf life, and batch release. The validation of these analytical methods is therefore critical to ensure the safety and efficacy of pharmaceuticals. Accordingly, method validation has been the focus of both scientific and regulatory interest for some time.

It is commonly accepted that the goal of method validation is to demonstrate that the method is "suitable" for its intended purpose (1,2). For any analytical method, performance characteristics which constitute desired "suitability" must be defined. Appropriately chosen acceptance criteria for these performance characteristics should then ensure the suitability of the method for its intended use.

However, typical acceptance criteria for analytical method precision and accuracy are not chosen with regard to the concept of method suitability and are commonly based on ad-hoc rules. For example, current pre-study acceptance criteria for bioanalytical methods require the observed mean to be within ±15% of the nominal value and the observed precision to be ≤15% coefficient of variation (%CV). Although such ad-hoc approaches may meet regulatory requirements (1), they yield unknown and uncontrolled risks of rejecting suitable bioanalytical methods (producer risk) and accepting unsuitable bioanalytical methods (consumer risk). Moreover, such acceptance criteria are incompatible with common recommendations for in-study acceptance criteria.

The inadequacy of such ad-hoc acceptance criteria has been clearly recognized (3). Alternate acceptance criteria emphasizing the use of total error have been discussed and/or proposed in the method validation literature (3–11). The use of a total error criterion which incorporates both systematic and random errors is a statistically and scientifically logical approach. However, a formal statistical framework for the validation of analytical methods, which incorporates the use of total error and controls the risks of incorrect decisions, has not yet been proposed or evaluated.

This paper proposes a total error criterion for assay validation through the use of two-sided β-content tolerance intervals. The performance of the proposed approach is compared to the performance of current ad-hoc approaches via simulation techniques. The use of the proposed approach

[1] Preclinical and Research Biostatistics, sanofi-aventis, Bridgewater, New Jersey, USA.
[2] 1041 Route 202-206, Mailstop M-203A, Bridgewater, New Jersey 08807-0800, USA.
[3] To whom correspondence should be addressed. (e-mail: david.hoffman2@sanofi-aventis.com)

Table I. Analysis of Variance Table for Balanced One-Way Random Effects Model

Source	Degrees of Freedom	Sums of Squares	Mean Square	EMS
Between-run	$df_B = I-1$	$SS_B = J\sum_{i=1}^{I}(\overline{Y}_i - \overline{Y})^2$	$MS_B = SS_B/df_B$	$J\sigma_B^2 + \sigma_E^2$
Within-run	$df_E = I(J-1)$	$SS_E = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_i)^2$	$MS_E = SS_E/df_E$	σ_E^2
Total	$df_T = IJ-1$	$SS_T = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y})^2$		

is demonstrated by application to actual assay validation data. The focus throughout will be on bioanalytical methods, though the proposed approach is readily applicable and relevant to other analytical methods (e.g. macromolecules, drug substance, etc.).

MATERIALS AND METHODS

One-Way Random Effects Model

During pre-study method validation, measurements are made over multiple independent assay runs with replicate determinations within each run. A statistical model to describe the measured values is given by:

$$Y_{ij} = \mu + b_i + \varepsilon_{ij}$$

where Y_{ij} is the jth (j=1,2,...,J) replicate observation from the ith (i=1,2,...,I) assay run, μ is the true (unknown) analytical mean for the method, b_i is the random error for the ith assay run, and ε_{ij} is the random error for the jth replicate observation from the ith assay run. The random errors b_i and ε_{ij} are assumed to be normally and independently distributed with means zero and variances σ_B^2 and σ_E^2, respectively. These variances, σ_B^2 and σ_E^2, correspond to the between-run (inter-batch) and within-run (intra-batch) variability of the method. The total analytical variability of the method is then given by $\sigma_{TOT}^2 = \sigma_B^2 + \sigma_E^2$.

The above is commonly referred to as a one-way random effects model (12). For convenience, we assume that the data are balanced (i.e. there are J replicates in each of the I runs). Denote the overall mean of the measurements by $\overline{Y} = \sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}/IJ$, and the mean for the ith assay run by $\overline{Y}_i = \sum_{j=1}^{J} Y_{ij}/J$. Table I gives the analysis of variance (ANOVA) table for the balanced one-way random effects model (where EMS denotes the expected mean square).

The mean squares MS_B and MS_E from Table I can be used to obtain estimates of the method within-run, between-run, and total variances. Table II gives the variance estimates obtained from the ANOVA mean squares.

Table II. Estimates of Within-Run, Between-Run, and Total Variance

Variance Component	Estimate
Within-run	$\widehat{\sigma}_E^2 = MS_E$
Between-run	$\widehat{\sigma}_B^2 = (MS_B - MS_E)/J$
Total	$\widehat{\sigma}_{TOT}^2 = \widehat{\sigma}_B^2 + \widehat{\sigma}_E^2$

Calculation of the quantities in Tables I and II is necessary to implement both the current ad-hoc acceptance criteria (for proper calculation of the observed coefficient of variation) and the proposed total error approach.

Current Acceptance Criteria

Current pre-study acceptance criteria for bioanalytical methods require the observed mean to be within ±15% of the nominal value and the observed precision to be ≤15% coefficient of variation (%CV), though these limits are both 20% at the lower limit of quantification (LLOQ). That is, $\overline{Y}$ must be within ±15% of the known nominal value, and $\widehat{\sigma}_{TOT}$ must be ≤15% relative to the mean value.

Now consider the current acceptance criteria for in-study monitoring: at least four of every six QC samples must be within 15% of their respective nominal concentration (1). Analytical runs failing this criterion must be rejected. This is the commonly known 4-6-15 rule.

The in-study acceptance criteria used for monitoring the method in routine use can be interpreted to define the suitability requirements of the method for its intended use. That is, it is reasonable to infer that a bioanalytical method is suitable for its intended use if at least 66.7% of the observed assay values (in the long run) are within 15% of the true value. This is a slight oversimplification, as the properties of a specific small sample (i.e. of six QC samples) are subject to random variation and may be different from the true long-run properties of the method.

Consider a method such that the true (long-run) proportion of observed values within 15% of nominal is exactly 0.667. Then application of the 4-6-15 rule will result in a rejection rate of approximately 32% (i.e. 32% of future analytical runs will be rejected). This approximate rejection rate is based on the assumption of six independent QC samples (and can be shown via simple binomial probability calculations), though the actual rejection rate will differ slightly depending on the proportion of total variability (σ_{TOT}^2) due to between-run variability (σ_B^2). The deficiencies of acceptance criteria such as the 4-6-15 rule are well-known (13). However, our long run interpretation above is likely in agreement with the original intent (if not application) of the 4-6-15 rule.

An appropriate choice of pre-study acceptance criteria should thus ensure that at least 66.7% of future assay values are within 15% of the true value. That is, the pre-study acceptance criteria should be consistent with the in-study criteria (i.e. the intended use of the method).

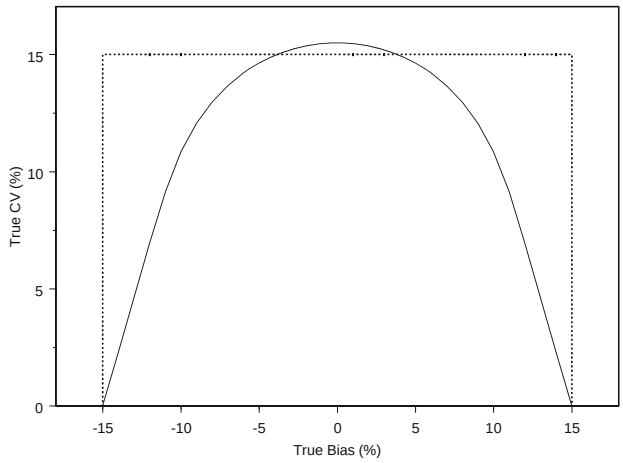

Fig. 1. Acceptance regions defined by current in-study and pre-study acceptance criteria. *Solid curve* gives current in-study acceptance region (i.e. 66.7% of observed values within 15% of nominal value). *Dashed line* gives current pre-study acceptance region (i.e. bias and %CV each within 15%).

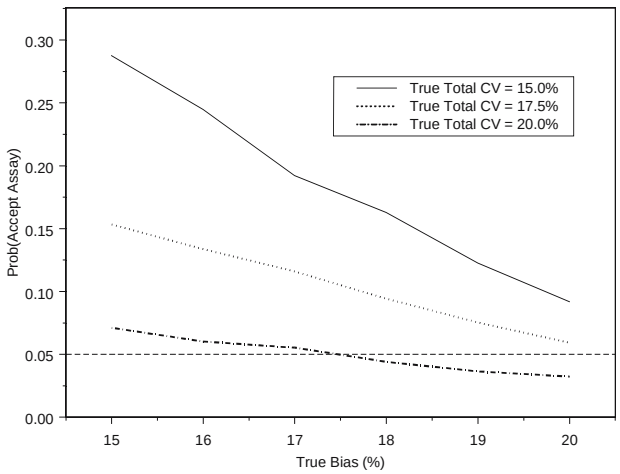

Fig. 3. Consumer risk for current pre-study acceptance criteria versus true bias, for various total %CV. Sampling design is (six runs, three replicates). Ratio $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_E^2)$ fixed at 0.50. Reference line at 5% error rate.

Figure 1 illustrates the acceptance regions defined by the pre-study and in-study criteria, respectively (under the assumption that observed assay values are normally distributed as given by the one-way random effects model described earlier).

Based on the in-study acceptance criteria, truly suitable assays have true means and coefficients of variation that lie within the solid curve in Fig. 1. These are assays such that at least 66.7% of observed values (in the long run) are within 15% of the true value. Truly unsuitable assays have true means and coefficients of variation that lie outside of the solid curve in Fig. 1. These are assays such that less than 66.7% of observed values (in the long run) are within 15% of the true value.

The acceptance region defined by the current pre-study criteria is given by the dashed line in Fig. 1. Assays which lie inside this region are such that true assay bias and %CV are each within 15%. Note that the pre-study acceptance region is not contained within the in-study acceptance region. Thus, some assays which are truly unsuitable (i.e. lie outside of the solid curve in Fig. 1) may be considered suitable according to the pre-study criteria.

Clearly, the current pre-study criteria are inconsistent with the in-study criteria, and will not ensure method suitability. Further, the current criteria are based on observed estimates of bias and variability, rather than on the true method bias and variability. The current pre-study criteria thus do not control consumer risks. That is, decisions based on the current pre-study criteria will result in incorrectly accepting assays that are truly unsuitable. This will be illustrated in the Results section.

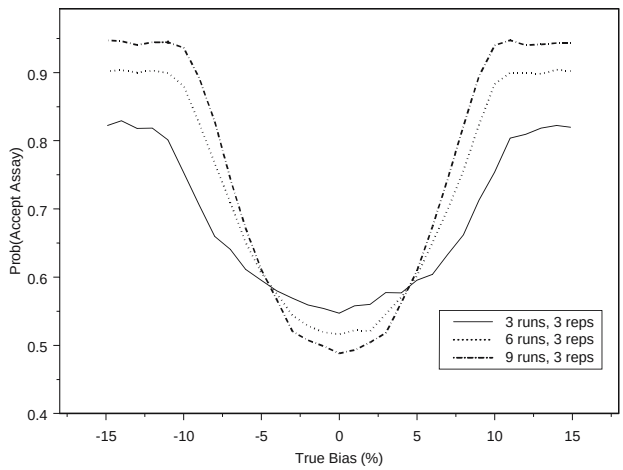

Fig. 2. Consumer risk for current pre-study acceptance criteria at in-study acceptance boundary (true total %CV not shown), for various sampling designs. Ratio $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_E^2)$ fixed at 0.50.

Proposed Total Error Approach

The use of confidence intervals and/or total error in application to method validation has been discussed or proposed in the literature (3–11). The use of total error is a statistically and scientifically sound approach which incorporates both systematic and random errors. A total error approach reflects how large a measurement error can be and is easily understood by analysts. Moreover, it is a single comprehensive measure of method performance, rather than an assessment of method bias and variability individually.

Ideal acceptance criteria would ensure that a high proportion (say β%) of future observations lie within acceptable limits (say ±15% of nominal), with a high degree of confidence (say γ%). Viewed in this manner, two-sided β-content tolerance intervals are an obvious choice.

A two-sided β-content tolerance interval is a statistical interval (L, U) such that at least a proportion β of a population will lie within the interval (L, U) with γ% confidence (14). Two-sided β-content tolerance intervals provide lower (L) and upper (U) limits so that we can claim a specified proportion β of measured assay values will lie within the interval (L, U), with specified confidence coefficient γ.

For any analytical method, we can define performance characteristics which constitute method suitability for its

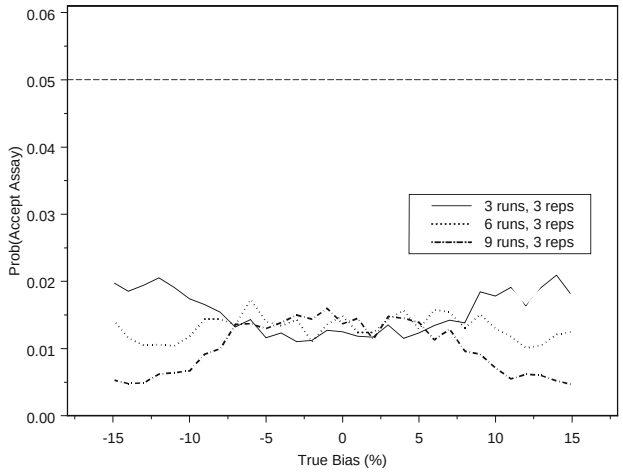

Fig. 4. Consumer risk for tolerance interval approach at in-study acceptance boundary (true total %CV not shown), for various sampling designs. Ratio $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_E^2)$ fixed at 0.50. Reference line at 5% error rate.

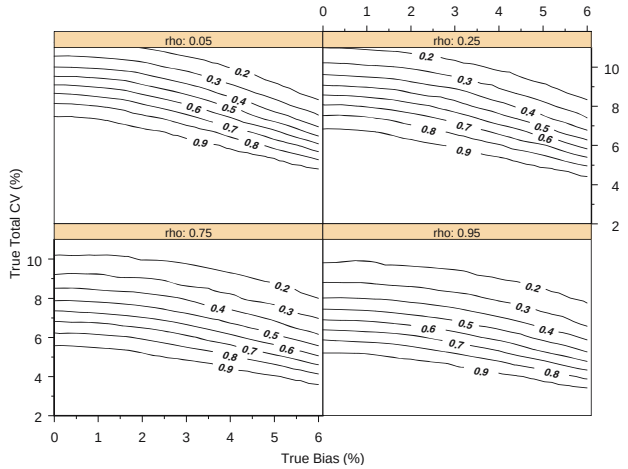

Fig. 6. Probability of passing tolerance interval criterion as a function of true bias and total %CV, for various ratio $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_E^2)$. Sampling design fixed at (six runs, three replicates).

intended use by appropriate choice of the proportion β and acceptable limits (A, B). That is, a method is suitable for its intended use if at least a proportion β of measured assay values lie within the specified acceptance limits (A, B). Two-sided β-content tolerance intervals provide a statistical framework for controlling the risk of incorrectly accepting methods that do not fulfill these suitability requirements.

The proposed total error approach is as follows:

1) Construct a two-sided β-content tolerance interval (L, U) with desired confidence level γ (say, 90%)
2) Compare the interval (L, U) to the acceptance limits (A, B)
3) If (L, U) falls completely within (A, B), the method is accepted; otherwise, the method is not accepted.

A similar approach incorporating the use of two-sided β-content tolerance intervals has also been proposed for evaluation of content uniformity (15,16). Note that this application of tolerance intervals has the structure of a statistical hypothesis test. The null hypothesis (H_0) is that less than a proportion β of measured assay values will fall within the acceptance limits (A, B), while the alternative (H_A) is that at least a proportion β fall within. The proposed total error procedure is to reject the null hypothesis (and therefore accept the analytical method) if the two-sided β-content tolerance interval falls completely within acceptance limits (A, B).

The construction of two-sided β-content tolerance intervals for the balanced one-way random effects model is straightforward and requires only the calculation of the quantities previously described in Tables I and II, as well as quantiles of the standard normal and chi-square distributions. Let $Z_{(1+\beta)/2}$ be the upper $(1+\beta)/2$ quantile of the standard normal distribution and $\chi^2_{1-\gamma,df}$ be the lower γ quantile of the chi-square distribution with df degrees of freedom.

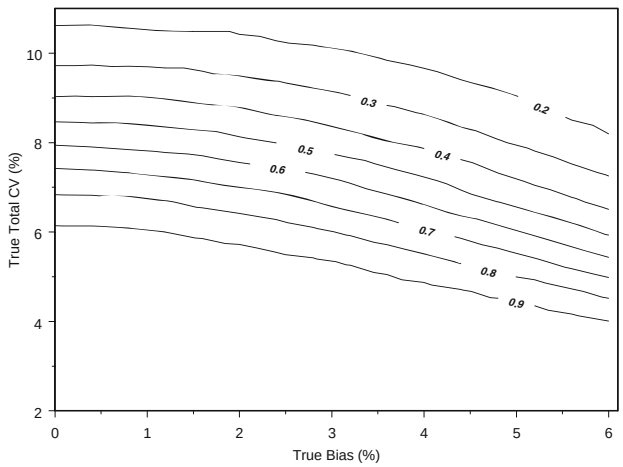

Fig. 5. Probability of passing tolerance interval criterion as a function of true bias and total %CV, for (six runs, three replicates) sampling design. Ratio $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_E^2)$ fixed at 0.50.

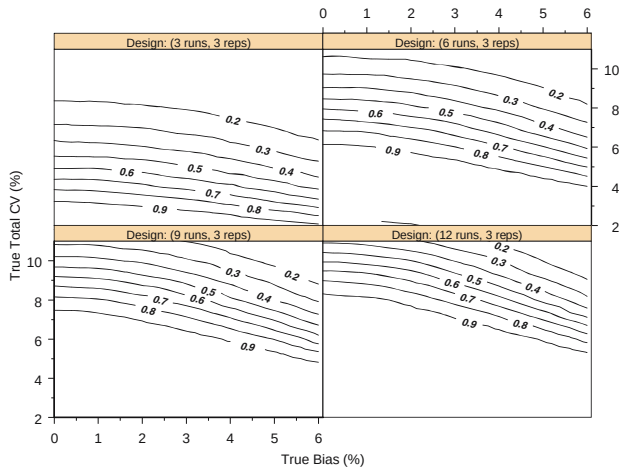

Fig. 7. Probability of passing tolerance interval criterion as a function of true bias and total %CV, for various sampling designs. Ratio $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_E^2)$ fixed at 0.50.

A two-sided β-content tolerance interval with confidence coefficient γ is given by (17):

$$\overline{Y} \pm Z_{(1+\beta)/2}\sqrt{1+N_e^{-1}}\sqrt{\hat{\sigma}^2_{\mathrm{TOT}} + \left\{H_1^2(1/J)^2\mathrm{MS}_\mathrm{B}^2 + \mathrm{H}_2^2((\mathrm{J}-1)/\mathrm{J})^2\mathrm{MS}_\mathrm{E}^2\right\}^{1/2}} \quad (1)$$

where

$$N_e = \frac{I(\mathrm{MS_B} + (\mathrm{J}-1)\mathrm{MS_E})}{\mathrm{MS_B}},$$

$$H_1 = \frac{\mathrm{df_B}}{\chi^2_{1-\gamma,\mathrm{df_B}}} - 1,$$

and

$$H_2 = \frac{\mathrm{df_E}}{\chi^2_{1-\gamma,\mathrm{df_E}}} - 1.$$

Slight modifications to Eq. 1 are suggested for unbalanced designs (see Appendix).

Implementation of the total error approach requires appropriate choices of content level (β), confidence level (γ) and acceptance limits (A, B). For bioanalytical assays, we propose that 66.7% content, 90% confidence, and ±15% acceptance limits are logical choices. That is, the total error approach consists of constructing a two-sided β=66.7% content, γ=90% confidence tolerance interval. If the resulting tolerance limits are completely within ±15% of nominal, the assay is accepted; otherwise, it is not.

The choice of 66.7% content with ±15% acceptance limits is the only choice consistent with the long run interpretation of the current in-study acceptance criteria discussed earlier. However, other choices for content level, confidence level, and acceptance limits are clearly possible. The deficiencies of the 4-6-15 rule noted earlier may motivate the choice of a content level greater than 66.7% (i.e. the 4-6-15 rule will reject approximately 32% of analytical runs even for a method with true content level of 66.7%, as noted earlier). For example, the 4-6-15 rule will reject approximately 10 and 5% of analytical runs for methods with true content levels of 80 and 85% within 15% of nominal, respectively (this can be shown via simple binomial probability calculations). Yet such choices for content level would likely be chosen solely to mitigate the deficient performance of the 4-6-15 rule. A more satisfactory alternative would be implementation of a statistically sound procedure to replace the 4-6-15 rule for in-study monitoring. Statistical quality control procedures (such as the Shewart procedure) may be appropriate alternatives for such in-study monitoring (13).

RESULTS

Current Acceptance Criteria

Consider again the acceptance region defined by the in-study acceptance criteria, previously illustrated by the solid curve in Fig. 1. Appropriate pre-study acceptance criteria should control the risk of accepting assays which have true bias and total %CV which lie on or outside this solid curve. Pre-study acceptance criteria should ensure that the probability of accepting such assays is small, say 5%.

The performance of the current pre-study acceptance criteria was evaluated via simulation techniques. Simulated assay values were assumed to follow the one-way random effects model described earlier, with normally distributed random errors. The simulation considered various combinations of true assay bias and total %CV which lie on or outside the in-study acceptance boundary. Several sampling designs were also considered (I=3, 6, or 9 runs with J=3 replicates per run). The proportion (ρ) of total variability due to between-run variability was assumed to be 0.50 (i.e. equal between-run and within-run variability). For each combination of true bias, true total %CV, and sampling design, 10,000 datasets were simulated and the probability of passing the current pre-study acceptance criteria was estimated. All simulations were performed using SAS (version 8.2) software.

Figure 2 gives the probability of passing the current pre-study acceptance criteria for assays with true bias and total %CV which lie directly on the in-study acceptance boundary (i.e. true bias and total %CV lie on the solid curve in Fig. 1), for various sampling designs and fixed ρ=0.50. Note that while Fig. 2 does not explicitly indicate the true total %CV, it is implicit from Fig. 1.

The results in Fig. 2 indicate that the current acceptance criteria have a high probability of accepting assays which lie on the in-study acceptance boundary. This probability is slightly greater than 50% for assays with no bias (and true total %CV of approximately 15%), and as large as 90% for assays with approximately ±10% bias (and true total %CV of approximately 11%).

Figure 3 gives the probability of passing the current pre-study acceptance criteria for assays with true bias and total %CV which lie outside the in-study acceptance boundary (i.e. true bias and total %CV lie outside the solid curve in Fig. 1).

The results in Fig. 3 indicate that the current acceptance criteria have an undesirably high probability of accepting even assays which lie well beyond the in-study acceptance boundary. For example, the probability of passing the current acceptance criteria is over 10% even for assays with true bias of ±20% and total %CV of 15%.

Total Error Approach

The performance of the proposed tolerance interval approach was also evaluated via simulation techniques. Simulated assay values were assumed to follow the one-way random effects model described earlier, with normally distributed random errors. The simulation considered various combinations of true assay bias and total %CV which lie on or inside the in-study acceptance region. Several sampling designs were considered (I=3, 6, 9, or 12 runs with J=3 replicates per run), as were several proportions (ρ) of total variability due to between-run variability. For each combination of true bias, true total %CV, sampling design and ρ, 10,000 datasets were simulated and the probability of passing the proposed tolerance interval criteria

Table III. Calculated Concentrations for Example No. 1 (ng/ml)

	Run					
Replicate	1	2	3	4	5	6
1	0.969	0.952	0.989	1.000	0.959	1.020
2	0.976	0.993	0.883	0.969	0.989	1.090
3	0.938	0.956	0.981	0.954	0.998	1.020

was estimated. All simulations were performed using SAS (version 8.2) software.

Figure 4 gives the probability of passing the proposed tolerance interval criterion for assays with true bias and total %CV which lie directly on the in-study acceptance boundary (i.e. true bias and total %CV lie on the solid curve in Fig. 1), for various sampling designs and fixed ρ=0.50. Note that while Fig. 4 does not explicitly indicate the true total %CV, it is implicit from Fig. 1.

The results in Fig. 4 indicate that the tolerance interval criterion controls the risk of accepting assays which lie on the in-study acceptance boundary. This probability is less than 5% across the entire in-study acceptance boundary, regardless of sampling design.

Other values of ρ were also considered in the simulation (results not shown). For all values of ρ considered (0.05–0.95), the probability of passing the tolerance interval criterion is less than 5% across the entire in-study acceptance boundary.

The tolerance interval criterion clearly controls consumer risk. It is also of interest to assess the producer risk associated with the tolerance interval approach. That is, the risk of incorrectly rejecting assays which are truly suitable.

Figure 5 gives the probability of passing the proposed tolerance interval criterion as a function of true assay bias and total %CV, for a (six runs, three replicate) sampling design and fixed ρ=0.50. Figure 6 gives the probability of passing the proposed tolerance interval criterion as a function of true assay bias and total %CV, for various ρ and a fixed sampling design of (six runs, three replicates). Figure 7 gives the probability of passing the proposed tolerance interval criterion as a function of true assay bias and total %CV, for various sampling designs and fixed ρ=0.50.

The results in Fig. 5 indicate that the tolerance interval approach has good power (≥80%) to accept assays with small bias and total %CV of approximately 7% or less, with a sampling design of (six runs, three replicates). As the true total %CV increases, the power to pass the tolerance interval criterion decreases accordingly.

Figures 6 and 7 indicate that the power of the tolerance interval approach to accept suitable assays depends largely on the 1) true total %CV of the assay and 2) the sampling design. Clearly, the true assay bias also has an impact on the power to accept suitable assays, but this impact is relatively low unless the true bias is fairly large (>5%). The ratio ρ also has some impact on the power to accept suitable assays, though generally only for extreme values near 0 or 1 (i.e. ρ values near 0 tend to increase power, while ρ values near 1 tend to decrease power).

With small sample sizes, such as a (three run, three replicate) design, the power to accept suitable assays is relatively low unless the true assay bias and total %CV are small (≤3%). However, with moderate sample sizes, the power to accept suitable assays is reasonable. For example, there is at least 90% power for small bias (≤2%) and total %CV≤6% with a (six run, three replicate) sampling design.

Tiered or multistage sampling designs may also be considered. That is, if the calculated tolerance interval is not contained within the specified acceptance limits, additional analytical runs may be performed to increase the sample size. Due to the relative conservatism of the tolerance interval approach (i.e. false acceptance rate less than 5%), an additional stage of sampling will not result in objectionably high false acceptance rates. However, the overall false acceptance rates of a multistage sampling design has not been formally investigated here.

EXAMPLES

The proposed total error approach is illustrated by application to data from two actual pre-study validation experiments conducted at sanofi-aventis. The data are calculated concentrations (ng/ml) of an analyte in human plasma and are shown in Tables III and V. For both examples, the sampling design consisted of six independent runs with three replicates per run.

The proposed total error approach entails the calculation of a two-sided β-content tolerance interval with content β=0.667 and confidence coefficient γ=0.90. A method will be judged suitable if the entire tolerance interval is within ± 15% of the nominal value.

Example 1

The data are shown in Table III. The nominal concentration is 1 ng/ml. Thus, the method will be judged suitable if the entire two-sided β-content tolerance interval is within (0.85, 1.15) ng/ml.

Table IV. Analysis of Variance Table for Example No. 1

Source	Degrees of Freedom	Sums of Squares	Mean Square
Between-run	$df_B = 5$	$SS_B = 0.016254$	$MS_B = 0.003251$
Within-run	$df_E = 12$	$SS_E = 0.014009$	$MS_E = 0.001167$
Total	$df_T = 17$	$SS_T = 0.030263$	

Table V. Calculated Concentrations for Example No. 2 (ng/ml)

	Run					
Replicate	1	2	3	4	5	6
1	0.1260	0.0969	0.0888	0.0991	0.1070	0.1150
2	0.1220	0.0963	0.0958	0.0921	0.0961	0.0989
3	0.1330	0.0931	0.0945	0.0982	0.1080	0.0961

Table VI. Analysis of Variance Table for Example No. 2

Source	Degrees of Freedom	Sums of Squares	Mean Square
Between-run	$df_B = 5$	$SS_B = 0.002327$	$MS_B = 0.000465$
Within-run	$df_E = 12$	$SS_E = 0.000422$	$MS_E = 0.000035$
Total	$df_T = 17$	$SS_T = 0.002749$	

Table VII. Unweighted Sums of Squares for Unbalanced One-Way Random Effects Model

Source	Degrees of Freedom	Sums of Squares	Mean Square	EMS
Between-run	$df_B = I-1$	$SS_{BU} = J_H \sum_{i=1}^{I} (\overline{Y}_i - \overline{Y})^2$	$MS_{BU} = SS_{BU}/df_B$	$J_H\sigma_B^2 + \sigma_E^2$
Within-run	$df_E = \sum_{i=1}^{I} J_i - I$	$SS_E = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_i)^2$	$MS_E = SS_E/df_E$	σ_E^2
Total	$df_T = \sum_{i=1}^{I} J_i - 1$	$SS_T = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y})^2$		

To calculate the interval, we construct the analysis of variance table as previously shown in Table I. We have I=6 runs, J=3 replicates per run, and overall mean concentration of $\overline{Y} = 0.9798\,\text{ng/ml}$. Table IV gives the analysis of variance.

From the mean squares in Table IV, we have that $\hat{\sigma}^2_{TOT} = 0.00186$ and N_e=10.308. The appropriate standard normal and chi-square quantiles can be easily obtained from tabulated values or from a statistical software package, and are as follows: $Z_{0.8335}$=0.96809, $\chi^2_{0.10,5} = 1.61031$, and $\chi^2_{0.10,12} = 6.30380$. From the degrees of freedom in Table IV and the chi-square quantiles above, we have H_1=2.1050 and H_2=0.9036. A two-sided β-content tolerance interval can then be calculated using Eq. 1:

$$0.9798 \pm 0.96809\sqrt{1+10.308^{-1}}\sqrt{0.00186 + \left\{\begin{array}{l} 2.1050^2(1/3)^2 0.003251^2 \\ +0.9036^2(2/3)^2 0.001167^2 \end{array}\right\}^{1/2}}$$

The resulting two-sided β-content tolerance interval is given by (0.914, 1.046) ng/ml. Equivalently, the interval is (-8.6%, 4.6%) from the nominal concentration. Thus, the assay performance is judged suitable at this nominal concentration.

Note that the observed estimates of the bias and total %CV are −2.02 and 4.40%, respectively. Since the observed bias is within ±15% of the nominal and the %CV is ≤15%, the assay also passes the current acceptance criteria.

Example 2

The data are shown in Table V. The nominal concentration is 0.1 ng/ml. Thus, the method will be judged suitable if the entire two-sided β-content tolerance interval is within (0.085, 0.115) ng/ml.

For this example, we have I=6 runs, J=3 replicates per run, and overall mean concentration of $\overline{Y} = 0.10316\,\text{ng/ml}$. Table VI gives the analysis of variance.

From the mean squares in Table VI, we have that $\hat{\sigma}^2_{TOT} = 0.000179$ and N_e=6.907. The appropriate standard normal and chi-square quantiles can be obtained as in the above example. A two-sided β-content tolerance interval can then be calculated using Eq. 1:

$$0.10316 \pm 0.96809\sqrt{1+6.907^{-1}}\sqrt{0.000179 + \left\{\begin{array}{l} 2.1050^2(1/3)^2 0.000465^2 \\ +0.9036^2(2/3)^2 0.000035^2 \end{array}\right\}^{1/2}}$$

The resulting two-sided β-content tolerance interval is given by (0.0799, 0.1265) ng/ml. Equivalently, the interval is (−20.1%, 26.5%) from the nominal concentration. Thus, the assay has failed to demonstrate suitable performance at this nominal concentration.

Note that the observed estimates of the bias and total %CV are 3.16 and 12.95%, respectively. Thus, the assay passes the current acceptance criteria but fails the proposed total error criterion.

Table VIII. Unweighted Sums of Squares for Unbalanced Example No. 1 Data (Replicates 1 and 2 from Run no. 1 Missing)

Source	Degrees of Freedom	Sums of Squares	Mean Square
Between-run	$df_B = 5$	$SS_B = 0.015126$	$MS_B = 0.003025$
Within-run	$df_E = 10$	$SS_E = 0.013191$	$MS_E = 0.001319$
Total	$df_T = 15$	$SS_T = 0.030119$	

CONCLUSION

Current pre-study acceptance criteria for the validation of analytical methods are based on ad-hoc rules and are inconsistent with ensuring suitable method performance. The current criteria yield high risks of falsely accepting assays which are truly unsuitable.

A total error approach incorporating the use of two-sided β-content tolerance intervals is proposed. The approach offers a formal statistical framework by which to assess method performance. The calculation of the intervals is straightforward and requires only quantities from a simple analysis of variance. The proposed approach is consistent with the concept of method suitability and controls the risk of falsely accepting truly unsuitable assays. The approach has good power to accept truly suitable assays with moderate sample sizes.

APPENDIX

For sampling designs which are unbalanced (i.e. the number of replicates is not identical for each run), slight modifications to the tolerance interval given in Eq. 1 are required. For unbalanced designs, the usual analysis of variance sums of squares (given in Table I) are replaced by unweighted sums of squares.

Let J_i be the number of replicates in the ith assay run (i=1,2,...,I). Denote the mean for the ith assay run by $\overline{Y}_i = \sum_{j=1}^{J_i} Y_{ij}/J_i$ the overall mean of the measurements by $\overline{Y} = \sum_{i=1}^{I} \overline{Y}_i/I$, and the harmonic mean of the replicates by $J_H = I \Big/ \sum_{i=1}^{I} (1/J_i)$. Table VII gives the unweighted sums of squares for the unbalanced one-way random effects model.

Estimates of within-run, between-run, and total variance are obtained as in Table II, with MS_{BU} and J_H replacing MS_B and J, respectively.

For unbalanced designs, a two-sided β-content tolerance interval with confidence coefficient γ is then given by [17]:

$$\overline{Y} \pm Z_{(1+\beta)/2}\sqrt{1+N_e^{-1}}\sqrt{\hat{\sigma}^2_{\mathrm{TOT}} + \left\{H_1^2(1/J_H)^2\mathrm{MS}^2_{\mathrm{B}U} + H_2^2((J_H-1)/J_H)^2\mathrm{MS}^2_E\right\}^{1/2}} \quad (2)$$

where all quantities are as in Eq. 1, with MS_{BU} and J_H replacing MS_B and J, respectively.

For illustration, consider the example given earlier in Table III. Assume the first two replicates in Run no. 1 are missing (i.e. the values 0.969 and 0.976), resulting in an unbalanced design. This yields an overall mean concentration of $\overline{Y} = 0.9759$ and replicate harmonic mean of J_H=2.25. Table VIII gives the unweighted sums of squares for the unbalanced example data.

From the unweighted mean squares in Table VIII, we have that $\hat{\sigma}^2_{\mathrm{TOT}} = 0.00208$ and N_e=9.270. The appropriate standard normal and chi-square quantiles can be easily obtained from tabulated values or from a statistical software package, and are as follows: $Z_{0.8335}$=0.96809, $\chi^2_{0.10,5} = 1.61031$, and $\chi^2_{0.10,10} = 4.86518$. From the degrees of freedom in Table VIII and the chi-square quantiles above, we have H_1=2.1050 and H_2=1.0554. A two-sided β-content tolerance interval can then be calculated using Eq. 2:

$$0.9759 \pm 0.96809\sqrt{1+9.270^{-1}}\sqrt{0.00208 + \left\{\begin{array}{l} 2.1050^2(1/2.25)^2\,0.003025^2 + \\ 1.0554^2(1.25/2.25)^2\,0.001319^2 \end{array}\right\}^{1/2}}$$

The resulting two-sided β-content tolerance interval is given by (0.904, 1.048) ng/ml, or (−9.6%, 4.8%) from the nominal concentration. The assay performance is judged suitable at this nominal concentration.

REFERENCES

1. Guidance for industry: bioanalytical method validation. Food and Drug Administration. May 2001.
2. Guidance for industry: analytical procedures and methods validation (Draft Guidance). Food and Drug Administration. August 2000.
3. R. O. Kringle and R.C. Khan-Malek. A statistical assessment of the recommendations from a conference on analytical methods validation in bioavailability, bioequivalence, and pharmacokinetic studies. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, Alexandria, VA, 510–514 (1994).
4. J. W. A. Findlay, W. C. Smith, J. W. Lee, *et al*. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J. Pharm. Biomed. Anal.* **21**:1249–1273 (2000).
5. K. Miller, R. Bowsher, A. Celniker, *et al*. Workshop on bioanalytical methods validation for macromolecules: summary report. *Pharm. Res.* **18**:1373–1383 (2001).
6. B. DeSilva, W. Smith, R. Weiner, *et al*. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm. Res.* **20**:1885–1900 (2003).
7. B. Boulanger, P. Chiap, W. Dewe, *et al*. An analysis of the SFSTP guide on validation of chromatographic bioanalytical methods: progresses and limitations. *J. Pharm. Biomed. Anal.* **32**:753–765 (2003).
8. P. Hubert, J. J. Nguyen-Huu, B. Boulanger, *et al*. Validation of quantitative analytical procedures, harmonization of approaches. *STP Pharma Pratiques* **13**:101–138 (2003).
9. P. Hubert, J. J. Nguyen-Huu, B. Boulanger, *et al*. Harmonization of strategies for the validation of quantitative analytical procedures: a SFSTP proposal—part 1. *J. Pharm. Biomed. Anal.* **36**:579–586 (2004).
10. J. Smolec, B. DeSilva, W. Smith, *et al*. Bioanalytical method validation for macromolecules in support of pharmacokinetic studies. *Pharm. Res.* **22**:1425–1431 (2005).
11. P. Hubert, J. J. Nguyen-Huu, B. Boulanger, *et al*. Quantitative analytical procedures: harmonization of the approaches part II—statistics. *STP Pharma Pratiques* **16**:28–58 (2006).
12. R. Burdick and F. Graybill. *Confidence Intervals on Variance Components*, Marcel Dekker, New York, 1992.
13. R. Kringle. An assessment of the 4-6-20 rule for acceptance of analytical runs in bioavailability, bioequivalence, and pharmacokinetic studies. *Pharm. Res.* **11**:556–560 (1994).
14. A. Wald and J. Wolfowitz. Tolerance limits for a normal distribution. *Ann. Math. Stat.* **17**:208–215 (1946).
15. W. Hauck and R. Shaikh. Sample sizes for batch acceptance from single- and multistage designs using two-sided normal tolerance intervals with specified content. *J. Biopharm. Stat.* **11**:335–346 (2001).
16. W. Hauck and R. Shaikh. Modified two-sided normal tolerance intervals for batch acceptance of dose uniformity. *Pharm. Statist.* **3**:89–97 (2004).
17. D. Hoffman and R. Kringle. Two-sided tolerance intervals for balanced and unbalanced random effects models. *J. Biopharm. Stat.* **15**:283–293 (2005).