Anal. Chem. **2001,** *73,* 310−314

Uniformity Test of Bias When the Reference Value Contains Experimental Error

Ruiguang Song,* Eugene Kennedy, and David Bartley

National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226

The uniformity test of biases for analytical methods must address uncertainties in the reference method. If the uncertainty associated with the estimates of true values is significant but ignored in the test of bias equality, the type I error can exceed the prespecified error rate. In general, when biases at each concentration level are confounded with a random component (confounding bias), the usual test of bias equality tests the uniformity of the combined bias, rather than the uniformity of fixed bias—the bias without the random component. Based on a confounding model that takes both the fixed and the confounding biases into account, the actual type I error rate of the uniformity test can be calculated. To eliminate the impact of confounding bias on the uniformity test of fixed biases, a new *F′*-test is proposed. The new *F′*-test is simply adding a correction factor to the conventional *F*-test. The correction factor is directly related to the uncertainty associated with the estimates of true values. A simulation study is conducted to show that the proposed test can bring the type I error rate down to the prespecified level. Data from two aldehyde methods are used to demonstrate how the proposed *F′*-test works. Recommendations on optimal sample allocation are also provided.

In analytical method evaluation, bias and precision are two important measures of method performance.[1,2] The bias is usually defined as the relative difference of the method mean response to the true value of sample concentration, and the precision is often measured by the relative standard deviation (RSD), which is the standard deviation divided by mean. In this article, our interest is focused on the evaluation of method bias and particularly the uniformity test of bias.

The performance of an analytical method can be affected by many environmental and sample conditions. Some of them can cause bias, and some may have impact on method precision. One of the most important factors is the sample concentration level itself. To evaluate the impact of sample concentration on method performance, samples from different concentration levels are used to test the method under consideration.[3,4] If the bias is constant over the concentration range considered, then the data from different concentration levels can be combined to get a better estimate of the bias. However, a hypothesis test is required to check the uniformity of bias before obtaining a pooled estimate of bias.

Factors that affect method bias include those that can be preset or controlled, such as the sample concentration. Bias determined by these factors is called fixed bias. Factors whose levels are not or can not be controlled may contribute to either method bias or precision depending on how bias and precision are defined. If these uncontrolled factors can stay constant under the repeatability condition[5] during the sampling and analysis period so they have equal effects on samples from the same group, then these factors are treated as random bias factors, and the bias derived from them is the random bias. Unlike the fixed bias, the random bias cannot be corrected and should be considered as a component of method precision.[6]

In this article, the method precision is defined as the relative standard deviation of measured results from samples that have the same concentration levels and are prepared and analyzed under repeatability conditions. That is, when a set of samples with the same concentration is prepared under the same condition and measured by the same analyst using the same measuring instrument, the uncertainty among measurement results of these samples is defined as the method precision.

In analytical method evaluation, various types of samples may be used to test the method depending on their availability. The test samples can be certified samples, spiked samples, or field samples. If the true values of test samples are known without error, then the uniformity test of bias is simply a test of equality of sample means relative to the true values. In this case, the following one-way analysis of variance (ANOVA) model can be used to describe the relationship between measurement results and biases.

Suppose that n samples from each of k concentration levels are used in the bias test. Results are expressed as

$$x_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, ..., k; \quad j = 1, ..., n$$

or in terms of bias

$$y_{ij} = x_{ij}/t_i - 1 = b_i + e_{ij}, \quad i = 1, ..., k; \quad j = 1, ..., n \quad (1)$$

where t_i is the true value, μ_i is the method mean, and $b_i = \mu_i/t_i -$

* Corresponding author: (e-mail) RSong@cdc.gov; (fax) 513 841-4545.

(1) Taylor, J. K. *Anal. Chem.* **1983**, *55*, 600A−608A.
(2) Keith, L. H.; Crummett, W.; Deegan, J., Jr.; Libby, R. A.; Taylor, J. K.; Wentler, G. *Anal. Chem.* **1983**, *55*, 2210−2218.
(3) Kennedy, E.; Fischbach, T.; Song, R.; Eller, P.; Shulman, S. *Guidelines for Air Sampling and Analytical Method Development and Evaluation*; DHHS-(NIOSH) publication 95-117, Cincinnati, 1995.

10.1021/ac000710n Not subject to U.S. Copyright. Publ. 2001 Am. Chem. Soc.
Published on Web 12/09/2000

1 is the bias relative to the true value at the ith concentration. Since measurement results often have a constant RSD, we assume that the relative error $e_{ij} = \epsilon_{ij}/t_i$ has a normal distribution with a mean zero and a constant variance σ_e^2.

Note that the above test is valid only if the true values of test samples are known without error. Unfortunately, it is often not the case in practice. The true values are rarely known and often substituted by estimates. For example, when spiked or field samples are used in the test, the true values of test samples are estimated by a reference method and the uncertainties associated with these estimates are usually ignored. Recently, it was found that many analytical methods failed the uniformity test. If the statistical test is legitimate, there should be only a small chance to observe a failure when an analytical method has a constant bias. To validate the test results, the impact of substituting the true value with an erratic reference value in the test must be evaluated. In this article, a confounding model that contains both the fixed effect of bias and the random effect due to the uncertainty associated with the estimates of true values is first introduced. A formula based on this model is provided to calculate the failure probability of the uniformity test so that the impact on rejecting rate due to the imprecision of the reference method can be assessed. A new F'-test that incorporates the current F-test statistic with a correction factor is then proposed. A simulation study was conducted to show the performance of the proposed procedure. Data from two aldehyde methods are used to compare results obtained from the proposed F'-test and the conventional F-test. Recommendations on optimal sample allocation are provided. Finally, how to eliminate the impact of confounding bias in the design stage is discussed.

A CONFOUNDING MODEL

When the true values of test samples are unknown and replaced with estimates from a reference method, the bias defined in the previous model is no longer the bias relative to a constant (the true value) but the bias relative to a variable—the estimate of the true value. In this case, the bias can be viewed as a sum of two components: a fixed component referring to the bias relative to the true value and a random component corresponding to a random deviation from the true bias due to the uncertainty associated with the estimate of true value. The bias relative to the true value is called fixed bias because it does not change when the sample concentration is fixed. In contrast, the random deviation from the true bias can change from one sample set to another even though the sample concentration of each set remains unchanged.

To describe the above situation and take both bias components into account the following confounding model is introduced:

$$y_{ij} = b_i + \delta_i + e_{ij}, \quad \delta_i \sim N(0,\sigma_\delta^2), \quad e_{ij} \sim \mathrm{N}(0,\sigma_e^{\,2}) \qquad (2)$$

where b_i is the fixed bias (relative to the true value) at the ith concentration and δ_i is the confounding bias that may contain the random bias from the test method and the bias deviation from the fixed bias due to the replacement of the true value with an estimate determined by a reference method. The bias deviation can be caused by the fixed and random biases of the reference method and also by the precision of the reference method. Normally, we assume that the reference method is unbiased but it may contain random bias. Here we should distinguish the random bias of the reference method from its contribution to the bias deviation due to its precision.

To see how the bias deviation is affected by the precision of the reference method, we assume that the reference method is unbiased and has no random bias with a nonzero constant RSD; that is

$$r_{ij} = t_i(1 + e'_{ij}), \quad i = 1, ..., k; \quad j = 1, ..., m \qquad (3)$$

$$e'_{ij} \sim N(0,\sigma_r^{\,2})$$

For a small x, we have $\mathrm{Ln}(1 + x) \approx x$, where Ln( ) is the function of natural logarithm. Therefore,

$$\mathrm{Ln}(\bar{r}_i) = \mathrm{Ln}(t_i) + \mathrm{Ln}(1 + \bar{e}'_i) \approx \mathrm{Ln}(t_i) + \bar{e}'_i$$

$$\bar{r}_i = \sum_{j=1}^{m} r_{ij}/m, \; \bar{e}'_i = \sum_{j=1}^{m} e'_{ij}/m \sim N(0,\sigma_r^{\,2}/m)$$

Let $z_{ij} = x_{ij}/\bar{r}_i$ and $l_{ij} = \mathrm{Ln}(z_{ij})$. Then

$$\begin{aligned} l_{ij} &= \mathrm{Ln}(x_{ij}/t_i) - \mathrm{Ln}(\bar{r}_i/t_i) \\ &= \mathrm{Ln}(1 + b_i + e_{ij}) - \mathrm{Ln}(1 + \bar{e}'_i) \\ &\approx b_i - \bar{e}'_i + e_{ij} \end{aligned}$$

Let $\delta_i = -\bar{e}'_i$ and $\sigma_\delta^2 = \sigma_r^2/m$. Then the above model can be represented by the confounding model 2. In this case, there is no random bias assumed from either the test method or the reference method. The confounding bias is solely derived from the imprecision of the reference method.

When a one-way ANOVA test is applied to the above confounding model, it is not testing the null hypothesis H_0, all b_i's are equal or $\Delta^2 = \sum_i (b_i - \bar{b})^2/k = 0$, but actually testing a composite hypothesis H'_0, $\Delta^2 = 0$ and $\sigma_\delta^2 = 0$. So the nominal type I error rate of the ANOVA test is the error rate for testing H'_0, not for testing H_0. The type I error rate for testing H_0 is always higher than the nominal rate for testing H'_0 and the extra rate is proportional to the variability of the random bias.

IMPACT OF CONFOUNDING BIAS

To evaluate the impact of the confounding bias on the uniformity test of fixed biases in model 2, the actual type I error rate β must be calculated from the uniformity test of fixed biases. Under the condition $\Delta^2 = 0$, the ANOVA F-test is actually testing the hypothesis that $\sigma_\delta^2 = 0$. When the data fail to pass the test, it is telling us that $\sigma_\delta^2 \neq 0$. In this case, the probability of failure depends on the standard deviation ratio of the two random

(4) Kennedy, E.; Fischbach, T.; Song, R.; Eller, P.; Shulman, S. *Analyst* **1996**, *121*, 1163−1169.
(5) Taylor, B. N.; Kuyatt, C. E. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297, 1994 ed.; U.S. Government Printing Office, Washington, DC, 1994.
(6) Kane, J. S. *Analyst* **1997**, *122*, 1283−1288.

Table 1. Probabilities of Rejecting the Null Hypothesis Given Samples from Four Concentration Levels with Seven Samples Per Level

	Δ/σ_e (bias difference/standard deviation of error)					
σ_δ/σ_e	0.0	0.2	0.4	0.6	0.8	1.0
0.0	0.050	0.113	0.343	0.687	0.920	0.990
0.2	0.098	0.170	0.395	0.697	0.909	0.985
0.4	0.262	0.330	0.511	0.727	0.890	0.969
0.6	0.478	0.522	0.634	0.770	0.883	0.952
0.8	0.654	0.677	0.738	0.817	0.889	0.943
1.0	0.771	0.783	0.816	0.860	0.904	0.942

components[7] $\theta = \sigma_\delta/\sigma_e$:

$$\beta(\theta) = \text{Prob}(F \geq F_{1-\alpha;\nu_1,\nu_2}) = \text{Prob}(F_{\nu_1,\nu_2} \geq F_{1-\alpha;\nu_1,\nu_2}/(1 + n\theta^2)) \quad (4)$$

where $F = \text{MS}_\delta/\text{MS}_e$ is the test statistic—the mean squares of between groups over the mean squares of within groups:

$$\text{MS}_e = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2; \quad \bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$$

$$\text{MS}_\delta = \frac{n}{k-1}\sum_{i=1}^{k}(\bar{y}_i - \bar{y}_{..})^2; \quad \bar{y}_{..} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_i$$

F_{ν_1,ν_2} is a random variable with an F-distribution, $F_{1-\alpha;\nu_1,\nu_2}$ is the $(1 - \alpha) \times 100\%$ percentile of the F-distribution, $\nu_1 = k - 1$, $\nu_2 = k(n - 1)$. The second column of Table 1 gives the actual type I error rate $\beta(\theta)$ for $k = 4$, $n = 7$, and $\alpha = 0.05$ when a random bias δ is present with the value of $\theta = \sigma_\delta/\sigma_e$ shown in the first column. When there is no random bias ($\theta = 0$ or $\sigma_\delta = 0$), the F-test gives the nominal type I error rate 5%. When the random bias has a standard deviation equal to the standard deviation of the error term ($\theta = 1$), the rejecting probability of the F-test increases to 0.771 even though the fixed bias is a constant ($\Delta^2 = 0$). To control the error rate at 10%, one should not allow the standard deviation of the random bias to be greater than 20% of the standard deviation of the error term.

To see how the F-test performs when $\Delta^2 \neq 0$, the sampling distribution of the test statistic F must be known. It can be derived from the noncentral χ^2 distribution[8] that $(k - 1)\text{MS}_\delta/[\sigma_e^2(1 + n\theta^2)]$ has a noncentral χ^2 distribution with degrees of freedom $k - 1$ and a noncentrality parameter $\lambda = nk(\Delta^2/\sigma_e^2)/(1 + n\theta^2)$. Therefore, $F/(1 + n\theta^2)$ has a noncentral F-distribution with the same noncentrality parameter λ and

$$\beta(\theta,\Delta) = \text{Prob}(F \geq F_{1-\alpha;\nu_1,\nu_2}) = \text{Prob}(F_{\nu_1,\nu_2;\lambda} \geq F_{1-\alpha;\nu_1,\nu_2}/(1 + n\theta^2)) \quad (5)$$

where $F_{\nu_1,\nu_2;\lambda}$ is a random variable with a noncentral F-distribution with a noncentral parameter λ.

(7) Scheffe, H. *The Analysis of Variance*; John Wiley & Sons: New York, 1959.
(8) Johnson, N. L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wiley & Sons: New York, 1995; Vol. 2, p 434.

Table 2. Simulation Results of Uniformity Test of Bias When True Values Are Estimated by a Reference Method Given That $k = 4$, $n = 7$, $m = 3$, and $\boldsymbol{\Delta} = 0$[a]

		type I error rate		
RSD_r/σ_e	σ_δ/σ_e	F-test expected	F-test observed	F'-test observed
0.2	1.154 7	0.064 71	0.064 31	0.046 66
0.4	2.309 4	0.115 28	0.114 72	0.044 88
0.6	3.464 1	0.207 56	0.207 91	0.048 53
0.8	4.618 8	0.328 62	0.329 14	0.054 10
1.0	5.773 5	0.454 38	0.453 77	0.061 04

[a] Results are based on 100 000 replicates.

If $\Delta = 0$, then $\lambda = 0$ and $\beta(\theta,0) = \beta(\theta)$. For $\Delta \neq 0$, values of $\beta(\theta,\Delta)$ are calculated for θ and $\Delta/\sigma_e = 0$, 0.2, 0.4, 0.6, 0.8, and 1.0 (see Table 1). From this table, it can be seen that, given the value of θ, the rejecting probability $\beta(\theta,\Delta)$ of the F-test always increases as Δ/σ_e increases. However, given the value of Δ/σ_e, the rejecting probability increases only if $\Delta/\sigma_e < 0.65$.

F'-TEST

To control the type I error when a random bias is present, apparently a new test statistic is needed, and the new test statistic will rely on the variability of the random bias. If the true value of θ is known, then on the basis of eq 4, the test statistic given by $F' = F/(1 + n\theta^2)$ can be used. Given the type I error rate α, the null hypothesis H_0 is rejected if and only if $F' \geq F_{1-\alpha;\nu_1,\nu_2}$. This test, the "$F'$-test", gives the exact type I error that is wanted because under the hypothesis H_0, $F/(1 + n\theta^2)$ has a central F-distribution.

However, θ is usually unknown. In this case, the parameter must be estimated on the basis of best knowledge and replaced in the test statistic with the estimate. In this case, the actual type I error rates may not be exactly equal to the target level, but they should be very close. The more accurate the estimate of θ, the closer the actual type I errors to the target level.

To see the performance of the proposed F'-test, consider a typical situation in analytical method evaluation where the true concentration is determined by a reference method. Suppose that the reference method is unbiased and a group of m samples from each concentration level is measured by the reference method to estimate the true value. If the reference method has a constant RSD, then the test statistic is given by

$$F' = F/(1 + n\hat{\theta}^2) \quad (6)$$

where $\hat{\theta} = \hat{\text{RSD}}_r/(m\text{MS}_e)^{1/2}$ and $\hat{\text{RSD}}_r$ is the pooled RSD estimate of the reference method. A simulation study was conducted for $k = 4$, $n = 7$, $m = 3$, and selected values of the ratio RSD_r/σ_e. Data are generated based on model 1 for the test method and model 3 for the reference method. Results based on 100 000 replicates are listed in Table 2.

In the simulation study, the true type I error rates are calculated using the formula 4 and presented in Table 2 under "F-test expected". The observed type I error rates from the simulation data are given under "F-test observed". All of them are within the 1% range from the true type I error rates. This indicates

Table 3. Uniformity Test of Bias on Two Analytical Methods

air concentration level			
1	2	3	statistic & *p*-value
	Test Method (Furfural), mg/m^3		
2.48	9.85	38.18	
2.59	10.15	35.72[a]	$F = 11.71$
2.52	9.8	40.35	$p = 0.0009$[b]
2.71	9.73	42.3	
2.71	9.92	41.5	$F' = 3.54$
2.64	10	42.34	$p = 0.0549$
	Independent Method, mg/m^3		
2.65	9.59	45.91	
2.87	9.42	44.64	
	Test Method (Glutaraldehyde), mg/m^3		
0.66	1.55	8.12	
0.82	1.45	8.14	$F = 11.28$
0.84	1.62	8.13	$p = 0.0010$[b]
0.85	1.48	8.34	
0.84	1.42	8.3	$F' = 2.27$
0.83	1.51	6.83	$p = 0.1375$
	Independent Method, mg/m^3		
0.83	1.55	6.8	
1.01	1.53	7.08	

[a] Adjusted value since tubing came off sampler during run. [b] Indicates significant differences at 95% confidence level in bias between levels.

that the simulation is quite accurate. The type I error rates based on the proposed F'-test are listed in the last column under "F'-test observed". They are fairly close to the 5% nominal type I error rate.

APPLICATION TO ALDEHYDE METHODS

Performance evaluation data from two aldehyde methods was used to compare the proposed F′-test results and the conventional F-test results. The two methods were compared using air-generated concentrations of two different aldehydes. The test method used a reactive sorbent sampler with gas chromatography analysis.[9] The reference method used an impinger sampler with liquid chromatography analysis.[10] Each method was tested at three concentrations with each aldehyde. Six samples were taken with the test method, and two samples were collected with the reference method. The precisions assumed for the F′-test with the test and reference methods were 0.034 and 0.035 for furfural and 0.072 and 0.082 for glutaraldehyde. The experimental data for all the methods are listed in Table 3. Uniformity test statistics of the bias and the corresponding probability (p) values are also provided in the table.

When looking at the methods from the analytical chemistry perspective, there was no apparent source of bias in the results of the studies. However, on the basis of the F-test, in which true values are simply replaced with corresponding estimates, the two methods fail the uniformity test of bias at the 1% significant level. Apparently, this is a type I error. The uncertainty associated with the reference method must be the main contributor to the error since the proposed F'-test does not show the significance, even at the 5% significance level.

OPTIMAL SAMPLE ALLOCATION

In this section, we consider the case where the true values of test samples are determined by an unbiased reference method. If the total cost of method evaluation is limited, one would like to know how many samples should be used to test the method under consideration and how many samples should be analyzed using the reference method to determine the true values. To solve this problem, we need to know the cost for analyzing a single sample by each method, the precision of each method, and the difference among biases from different concentration levels that we would like to detect.

If the cost of analyzing a sample using the test (reference) method is c_t (c_r, respectively) and the total cost of the evaluation is limited by C, then the number of concentration levels k, the number of samples analyzed by the test method at each level n, and the number of samples analyzed by the reference method at each level m are limited by the following restriction condition:

$$k(nc_t + mc_r) \le C \tag{7}$$

Suppose that RSD_t and RSD_r are the relative standard deviations of the test method and the reference method, respectively, and Δ_0 is the difference to be detected. Given the type I error α, we can calculate the power of detecting the difference with $\theta = RSD_r/(m^{1/2} \times RSD_t)$ and $\Delta = \Delta_0$. The optimal combination of k, n, and m is the one that maximizes the power among all combinations satisfying (7)

$$\max\{\beta(\theta,\Delta_0)|(k,n,m):k \times (nc_t + mc_r) \le C\}$$

As an example, we set $c_t = c_r = c$ and $C/c = 40$. Given $RSD_r/RSD_t = 0.1$, 0.5, and 1.0 and $\Delta_0/RSD_t = 0.5$ and 0.1, we calculated the power for various combinations of (k, n, m) with $k(n + m) = 40$ (see Table 4). Probability values in boldface type correspond to the optimal combination of (n, m) for a given k.

From Table 4, we can see that when k or $n + m$ is fixed, the optimal ratio m/n is positively related to the precision ratio RSD_r/RSD_t. In other words, the more precise the reference method, the less number of samples needed to estimate the true value.

In many cases, the number of concentration levels is predetermined. If this number can be varied, then the above example suggests that we should keep the number of concentration levels minimum to increase the power of detecting the bias difference Δ_0 among these levels.

In the last column of Table 4, we listed the probability values calculated under the assumption that true sample concentrations are known. This assumption is equivalent to the condition $RSD_r = 0$ or $m = \infty$. When these probabilities are compared with results listed in the previous columns, we can see how much power gets lost when true sample concentrations are unknown and estimated by a reference method with different precision and sample size.

DISCUSSION

Frequently, analytical methods are evaluated using standard reference materials (SRMs). A sample made up with SRMs comes

(9) Kennedy, E. R.; Gagnon, Y. T.; Okenfuss, J. R.; Teass, A. W. *Appl. Ind. Hyg.* **1988**, *3*, 274−279.
(10) Lipari, F.; Swarin, S. J. *J. Chromatogr.* **1982**, *247*, 297−306.

Table 4. Probabilities of Rejecting the Null Hypothesis in the F'-Test When $\Delta_0 = RSD_t$ and $k(n + m) = 40$

			number of samples per level analyzed by a reference method (m)										
RSD_r/RSD_t	k	$n + m$	1	2	3	4	5	6	7	8	9	10	$RSD_r = 0$
0.1	2	20	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.996	0.993	0.988	**1.000**
	4	10	**0.998**	0.996	0.989	0.970	0.924	0.820	0.614	0.298			**1.000**
	5	8	**0.995**	0.986	0.958	0.878	0.687	0.346					**0.999**
	8	5	**0.962**	0.841	0.481								**0.995**
	10	4	**0.899**	0.561									**0.988**
	20	2											**0.837**
0.5	2	20	0.706	0.898	0.953	0.972	0.979	**0.981**	0.981	0.979	0.974	0.966	**1.000**
	4	10	0.752	0.894	**0.917**	0.903	0.856	0.755	0.569	0.284			**1.000**
	5	8	0.757	**0.871**	0.865	0.791	0.622	0.323					**0.999**
	8	5	**0.716**	0.696	0.420								**0.995**
	10	4	**0.650**	0.461									**0.988**
	20	2											**0.837**
1	2	20	0.269	0.454	0.591	0.687	0.753	0.797	0.825	0.842	**0.849**	0.848	**1.000**
	4	10	0.291	0.486	0.599	**0.648**	0.644	0.586	0.459	0.247			**1.000**
	5	8	0.301	0.487	**0.569**	0.565	0.470	0.268					**0.999**
	8	5	0.308	**0.409**	0.298								**0.995**
	10	4	**0.292**	0.290									**0.988**
	20	2											**0.837**

with a certified value and specified uncertainty.[5] The F'-test is particularly applicable to this situation. If u_c is the standard error of the certified value, then the parameter θ can be estimated by $\hat{\theta} = u_c/MS_e$.

In the bias evaluation of analytical methods, the random bias should not be treated as the same as the fixed bias. For fixed bias, its actual values and how these values depend on concentration level and other factors are of interest. But for a random bias, its variability is required. A random bias can originate from two different sources, one from the test method itself and the other one from the reference method used to determine the true value. In the former case, the random bias should be considered as a component of method precision in analytical method evaluation. In the latter case, the random bias should not be ignored in the uniformity test of fixed bias.

If the random bias is due to the test method itself, it cannot be estimated and hence cannot be separated from fixed bias using data with only one sample group per concentration. In this case, the type I error rate of uniformity test of fixed bias is unknown but higher than the nominal level. However, if the random bias is due to the reference method, then the derived random bias can be estimated and the type I error rate can be controlled in the desired range by applying the proposed F'-test.

If multiple sample groups per concentration level are allowed in the bias evaluation (e.g., two independently prepared standard solutions), one can eliminate the confounding effect by applying the following two-way mixed effect model to test the uniformity of fixed biases:

$$y_{ijk} = b_i + \delta_{ij} + e_{ijk}, \quad \delta_{ij} \sim N(0,\sigma_\delta^2), \quad e_{ijk} \sim N(0,\sigma_e^2) \qquad (8)$$

With this model, one can separate the random bias from the fixed bias and control the type I error of testing the null hypothesis H_0 at the nominal level. Particularly for a balanced design, the test statistic is given by $F = MS_b/MS_\delta$, where

$$MS_\delta = \frac{m}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{ij.} - \bar{y}_{i..})^2; \quad \bar{y}_{i..} = \frac{1}{n}\sum_{j=1}^{n}\bar{y}_{ij.}, \quad \bar{y}_{ij.} = \frac{1}{m}\sum_{l=1}^{m}\bar{y}_{ijk}$$

$$MS_b = \frac{nm}{k-1}\sum_{i=1}^{k}(\bar{y}_{i..} - \bar{y}_{...})^2; \quad \bar{y}_{...} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_{i..}$$

Under the null hypothesis H_0, the test statistic F has an F-distribution with degrees of freedom $k - 1$ and $k(n - 1)$. Given the type I error rate α, the null hypothesis H_0 is rejected if and only if $F \geq F_{1-\alpha;k-1,k(n-1)}$.

Received for review June 20, 2000. Accepted October 27, 2000.

AC000710N