Clinical Chemistry 44:4
839–848 (1998)

A stochastic approximation method for assigning values to calibrators

Brian Schlain*

A new procedure is provided for transferring analyte concentration values from a reference material to production calibrators. This method is robust to calibration curve-fitting errors and can be accomplished using only one instrument and one set of reagents. An easily implemented stochastic approximation algorithm iteratively finds the appropriate analyte level of a standard prepared from a reference material that will yield the same average signal response as the new production calibrator. Alternatively, a production bulk calibrator material can be iteratively adjusted to give the same average signal response as some prespecified, fixed reference standard. In either case, the outputted value assignment of the production calibrator is the analyte concentration of the reference standard in the final iteration of the algorithm. Sample sizes are statistically determined as functions of known within-run signal response precisions and user-specified accuracy tolerances.

The problem of transferring the analyte concentration value of a reference material to a new production calibrator is a common manufacturing activity in the in vitro medical diagnostic industry. Many popular approaches to this problem, whether they use consensus values from many laboratories, reference laboratories, or master lots, involve fitting calibration curves to assay responses of reference standards to determine the concentration values of the new production calibrators, which are assayed as unknowns *(1, 2)*. There are at least two potential problems with this approach: Any consistent systematic error in fitting the calibration curve is propagated to the value assignment of the production calibrator regardless of how many assay setups (or runs) are used, and the required sample sizes, even with a single instrument, are direct functions of the magnitudes of the random variabilities of fitted calibration curves and/or assay setups together with the within-run variabilities. With consensus value methods, the required sample sizes are further increased by the variabilities among laboratories, analyzers, and assay methods. The object of this paper is to provide a simpler, more economical method of value assignment that is robust to calibration curve-fitting errors and has its required sample size driven only by the smallest component of variability, which is the intraassay signal response SD, rather than by the interassay or interlaboratory sources of variabilities of recovered values, which drive up sample sizes.

For the case in which both the production calibrators and the reference materials have the same signal-response curve, which is linear in analyte concentration with a zero intercept, this problem has already been solved by the following well known equation *(1)*:

$$[V_t] = [V_r]\frac{S_t}{S_r},$$

where

$[V_t]$ is the estimated value assignment of the new production calibrator;

$[V_r]$ is the known concentration of the reference standard;

S_t is the average signal response of the new production calibrator; and

S_r is the average signal response of the reference standard.

Note that the above equation, which is restricted to a single linear signal-response curve with zero intercept, is prone to biases from nonlinearity, matrix, or specificity effects in either of the materials *(1)*.

A Stochastic Approximation Method of Value Assigning Production Calibrators

This paper provides a new method of transferring analyte concentration values from reference materials to production calibrators. As will be shown, the accuracy of this new method does not depend on fitted calibration curves. The stochastic approximation method of value assign-

Consulting biostatistician, ExperTech Affiliate, 46 Jamaica Road, Brookline, MA 02146.

* Address for correspondence. Fax 617-731-2456; e-mail bschlain@ultranet.com.

Received September 24, 1997; revision accepted January 16, 1998.

ment assumes that two standards that give the same average signal response when assayed on some specific assay system are functionally equivalent with respect to analyte level. Although this paper assumes that the analyte value of the reference material is known without error, in theory the stochastic approximation method extends to the case in which there is uncertainty in this value. As will be explained in the *Discussion*, judgment on the part of the user is required to determine when it is sensible to extend the methodology to this case.

In this paper, there are two distinct but related ways of using the stochastic approximation method. In the first way, a stochastic approximation algorithm together with a stopping rule iteratively finds the appropriate analyte level for a standard prepared from a reference material that will yield the same average signal response as the new production calibrator, whose concentration is not precisely known at the time of manufacture; the value assignment of the production calibrator is then the analyte level of the reference standard at the final iteration of the algorithm. In the second way, the same algorithmic method is used to alter the analyte concentration level of the new production calibrator bulk (by adding or diluting) so that it will yield the same average signal response as some prespecified reference standard. The value assignment of the altered production calibrator bulk material in the final iteration of the algorithm is obtained from the known analyte concentration of the fixed, reference standard.

Because there are two aforementioned ways of using the stochastic approximation method of value assignment, the standard which is iteratively adjusted (whether it be prepared from the reference or from calibrator bulk material) will be denoted as the "adjusted standard". The standard that is not iteratively adjusted will in turn be denoted as the "fixed standard". The corresponding materials from which these standards are prepared will similarly be denoted as the fixed and adjusted materials, respectively. Which of the standard preparations is to be iteratively adjusted by dilutions and/or additions is a matter of convenience, provided that matrix or antigen effects are not introduced in the process of the iterative adjustment.

At each iteration of the algorithm, the fixed and adjusted standards are run in statistically determined numbers of replicates and statistically compared to determine whether the algorithm should stop. The methodology for determining these sample sizes will be presented later.

The following assumptions are made with the stochastic approximation method. Both the adjusted and fixed standards can, with sufficient specificity, be assayed within the same run on the same instrument, which can be used reliably for value assignment. The assay measurement system is in a state of control.

The underlying matrices and antigen species of the fixed and adjusted materials are similar enough that systematic differences in signal response can be assumed to reflect only differences in analyte concentrations. (As will be mentioned in the *Discussion*, there are some cases where the user might decide to waive this latter assumption.) The adjusted material has dilution and/or addition accuracy within some appropriate analyte range; for the case in which this latter assumption is not tenable, possible alternatives are presented in the *Discussion*.

The stochastic approximation algorithm for updating the analyte concentration correction factor in the preparation of the adjusted standard is given by *(3, 4)*:

$$c_{i+1} = c_i + \frac{D_i}{i \times \hat{\beta}}, \tag{1}$$

where

$i = 1, 2, \dots$ denotes the iteration;

c_i is the estimated concentration correction factor at the *ith* iteration for the adjusted standard;

c_1 is always set to zero;

D_i is the estimated average difference in log signal response units of the fixed minus the adjusted standard at the *i*th iteration; and

$\hat{\beta}$ is the approximate slope of the log signal response vs analyte concentration in the vicinity of the fixed standard.

Note that in Eq. 1, the term ($D_i/i \times \hat{\beta}$) provides the analyte concentration correction to the adjusted standard of the current iteration and that c_{i+1} provides the correction to the adjusted standard that was prepared for iteration 1. To determine by which factor to effectively dilute or add to the first adjusted standard of any iteration, it is necessary to have an initial estimate of the value assignment of the production calibrator. Based on this initial estimate, the adjusted standard of the first iteration is prepared to approximate the concentration of the fixed standard.

The factor i in Eq. 1 is necessary to guarantee that the algorithm will converge *(3)*. Even if $\hat{\beta}$ and the initial estimated value assignment are poor estimates, the algorithm is still expected to converge, but will do so more slowly compared with having good estimates *(3–5)*. Thus, for the first iteration, the user should try to prepare the adjusted standard to be close to the fixed standard. The closer the initial adjusted standard in iteration 1 is to the fixed standard with respect to expected signal response, the faster the algorithm is expected to converge *(5)*. More will be mentioned later on obtaining these two aforementioned initial estimates.

A Stopping Rule for the Stochastic Approximation Algorithm

For ease of explication, a stopping rule is first presented for the case in which the data are not adjusted for within-assay-run time trends. For each iteration i of the stochastic approximation algorithm, the following limits of a two-sided 95% confidence interval are first constructed *(6)*:

$$l_i = Y_{fi.} - Y_{ai.} - S_{pi}\ \sqrt{1/n_{fi} + 1/n_{ai}}\ t_{\alpha/2}{}^{(v_i)} \tag{2}$$

$$u_i = Y_{fi.} - Y_{ai,} + S_{pi}\ \sqrt{1/n_{fi} + 1/n_{ai}}\ t_{\alpha/2}{}^{(v_i)} \tag{3}$$

where

$Y_{fi.}$ is the average natural log signal response of the replicates for the fixed standard of the *i*th iteration;

$Y_{ai.}$ is the average natural log signal response of the replicates for the adjusted standard of the *i*th iteration;

$S_{pi} = \sqrt{[(n_{ai} - 1)S_{ai}{}^2 + (n_{fi} - 1)S_{fi}{}^2]/(n_{ai} + n_{fi} - 2)}$;

$S_{ai}{}^2 = \Sigma_j(Y_{aij} - Y_{ai.})^2/(n_{ai} - 1)$;

$S_{fi}{}^2 = \Sigma_j(Y_{fij} - Y_{fi.})^2/(n_{fi} - 1)$;

Y_{aij} is the natural log signal response for the *j*th replicate of the adjusted material for iteration *i*;

Y_{fij} is the natural log signal response for the *j*th replicate of the fixed material for iteration *i*;

$t_{\alpha/2}(v_i)$ is the upper $1 - \alpha/2$ percentile of the Student's *t*-distribution with v_i degrees of freedom; the parameter α is customarily fixed at 0.05;

$v_i = n_{ai} + n_{fi} - 2$;

n_{ai} is the sample size for the adjusted standard at the *i*th iteration; and

n_{fi} is the sample size for the fixed standard at the *i*th iteration.

The following is one possible stopping rule for the stochastic approximation algorithm:

If $u_i \leq$ UL and $l_i \geq$ LL, then output the known analyte concentration of the reference standard at the *i*th iteration as the estimated value assignment of the new production calibrator, where LL and UL are the lower and upper manufacturing specifications, respectively, for the average log signal difference between the two standards; otherwise, update the concentration correction factor by Eq. 1 to prepare a new adjusted standard and proceed to iteration $i + 1$.

The natural logarithmic transformation is used because the ratio of the fixed to the adjusted standard with respect to signal response has proven from experience to generally be a meaningful comparison metric, and the exponentiated average difference in natural log signal responses ($\exp[Y_{fi.} - Y_{ai.}]$) is known to be an excellent first order approximation of this ratio *(7)*. Note that if δ is between -0.05 and $+0.05$, then $\exp(\delta) \approx 1 + \delta$. The manufacturing specifications, LL and UL, can thus for practical purposes be derived from lower and upper specifications on the ratio of the fixed to the adjusted standard with respect to average untransformed signal response. See the *Discussion* for further considerations in the setting of these specifications for the case in which there is uncertainty in the assigned value of the reference material.

To ensure that this procedure be unbiased in the presence of within-run time trends, the user should use a different randomly selected sample sequence for assaying the adjusted and fixed standards for each run of each iteration. Random number generators such as SAS RANUNI can be used to construct random sample sequences. These methods will be further elucidated later by a hypothetical data example.

The following two assumptions, which, unless always true, should be checked statistically, are important to the validity of the value assignment experiments. The adjusted and fixed standards do not have statistically different within-run time trends, and the ratio of the average signal responses of the fixed to the adjusted standards remains constant across assay runs within expected statistical variation. In the author's consulting experience with value assignment problems, it has always been important to check these assumptions. Possible examples of laboratory blunders that could lead to violations of these assumptions are volumetric or gravimetric errors during sample and/or reagent preparations, not allowing sufficient time for a refrigerated sample to equilibrate to room temperature, a contaminated calibrator vial, or an improperly filled or mixed calibrator or reagent vial. By using a trend-free design and multiple regression analysis with a linear model that adjusts for time trends *(8)*, these aforementioned assumptions can be statistically checked (as will later be illustrated with an actual data example). Within-run time trends occur not infrequently with many analyzers and can be due to assay drift and/or various instrument effects. Further advantages of the multiple regression method with a trend-free design are that the random sample sequences for each iteration can be replaced by a single trend robust design, and the required sample sizes for each iteration can be minimized, because the error variance is minimized by removing the within-run time trends.

With the multiple regression approach, the term $Y_{fi.} - Y_{ai.}$ in Eqs. 2 and 3 is replaced by D_i, which would be obtained from fitting a linear model that adjusts for time trends to the data. For a completely orthogonal design, in which the estimated parameters are uncorrelated, D_i would correspond to $Y_{fi.} - Y_{ai.}$; the pooled standard deviation S_{pi} in Eqs. 2 and 3 is replaced by the standard error of D_i, which would be outputted by a standard multiple regression program. The degrees of freedom v_i for the critical value $t_\alpha(v_i)$ in Eqs. 2 and 3 would correspond to the *df* for error, which would also be outputted by any standard multiple regression software package. Thus, for the general case, Eqs. 2 and 3 are replaced with:

$$l_i = D_i - SE(D_i)t_{\alpha/2}(v_i) \tag{4}$$

$$u_i = D_i + SE(D_i)t_{\alpha/2}(v_i) \tag{5}$$

For determining the required sample sizes for each iteration of the algorithm, one possible criterion is the following: If at some iteration, the two standards are identical with respect to their average signal responses, there should be at least 0.95 certainty that the algorithm will stop iterating [which occurs when the confidence interval constructed by Eqs. 2 and 3 or Eqs. 4 and 5 is completely contained by the specification interval (LL, UL)]. The statistical methodology for determining these required sample sizes is provided in *Appendix I*.

Obtaining Initial Estimates of the Value Assignment and Calibration Curve Slope for the Stochastic Approximation Algorithm

The speed of the algorithm can be improved by preparing the initial adjusted standard in iteration 1 to be close to the fixed standard and having a good approximation to the calibration curve slope β in the vicinity of the fixed standard *(4, 5)*. If the assay response curves are reasonably stable, β can be estimated from historical data. If there is not excessive curvature in the assay log-response curve in the vicinity of the fixed standard, a simple linear approximation based on two standards, which span this local region, will be adequate for estimating the slope.

The initial adjusted standard for iteration 1 could be determined by first assaying a range of standards prepared from the adjusted material along with the fixed standard using a calibration curve based on (or traceable to) the reference material. If β cannot be estimated from historical data, a pilot experiment should be performed with several dilutions and/or additions of the reference material in the region of the fixed standard. Some trend-free designs that could be used for the pilot experiment are given in references *8* and *9*.

A Hypothetical Value Assignment Example Using Random Sample Sequences

A hypothetical data example is based on an actual value assignment data set provided in Table 1, which was generated using the stochastic approximation algorithm with a trend-free design. For the hypothetical and for the actual data examples, the fixed standard was a reference calibrator equivalent to 1 μg/L, and the initial adjusted standard, which was prepared from the production calibrator bulk material, was estimated to also have a value of 1 μg/L at the first iteration of the algorithm.

For simplicity of explication, the hypothetical data example does not use regression analysis with a trend-free design but supposes that the data in Table 1 were instead generated from the random sample sequences given in Table 2. The three random 30-sample sequences in Table 2, which correspond to the runs in Table 1, were generated using the SAS program given in *Appendix II*.

Table 1. Assay system data.

		Iteration 1	Iteration 2	
Time	Std[a]	Run 1 signal	Run 1 signal	Run 2 signal
1	Test	2757.7	2677.1	2713.8
2	Ref	2745.7	2778.8	2814.1
3	Ref	2855.4	2738.8	2735.6
4	Test	2749.3	2745.2	2829.0
5	Ref	2836.9	2750.6	2800.4
6	Test	2736.9	2792.3	2796.9
7	Test	2851.2	2694.8	2759.2
8	Ref	2788.2	2757.2	2777.2
9	Ref	2865.7	2686.9	2811.5
10	Test	2699.3	2654.7	2824.6
11	Test	2735.3	2799.8	2781.6
12	Ref	2732.7	2755.7	2822.1
13	Test	2624.0	2689.3	2850.8
14	Ref	2750.2	2821.4	2762.8
15	Ref	2841.0	2761.9	2747.1
16	Test	2729.7	2727.2	2769.8
17	Test	2715.9	2707.4	2783.6
18	Ref	2714.1	2732.5	2772.0
19	Ref	2799.7	2799.8	2722.6
20	Test	2840.6	2819.5	2829.8
21	Ref	2934.7	2806.2	2781.8
22	Test	2799.0	2774.0	2788.7
23	Test	2755.8	2709.5	2731.5
24	Ref	2841.2	2788.4	2811.7
25	Ref	2888.1	2712.3	2752.7
26	Test	2827.7	2748.2	2732.5
27	Test	2701.7	2670.5	2686.2
28	Ref	2860.7	2696.5	2836.4
29	Test	2780.7	2598.5	2720.9
30	Ref	2836.0	2685.6	2721.5

[a] Std, Standard; Ref, reference.

By the argument that follows, it will be shown how the sample sizes given in Table 1 were adequate for the hypothetical example. It was determined by examining historical calibration curves that the lower and upper specifications of the ratio of the fixed to the adjusted standard with respect to the average signal response should be 0.975 and 1.025, respectively, which is approximately equivalent to a specification of ±0.025 in log-signal units. Thus, for the stopping rule, LL and UL were set at -0.025 and $+0.025$, respectively. From a large amount of historical signal-response data, the standard deviation of the log signal response (σ) was estimated to be 0.022. The α-level for the stopping rule of the stochastic approximation algorithm was set at 0.05. For the random design, the *df* for two runs of 15 replicates for each standard (corresponding to iteration 2 of Table 1) are given by $v = 56$. If at some iteration of the algorithm, the fixed and adjusted standards were identical, we would have liked to have been at least 0.95 certain that the algorithm would stop. With two runs of 15 replicates per standard (corresponding to iteration 2 of Table 1) and with the aforementioned values of LL, UL, σ, α, and v together with the methodology in *Appendix I*, it was calculated to be at least 0.982 certain that the algorithm would stop when the two standards were identical with respect to average signal response. For only one run of 15 replicates per standard (corresponding to iteration 1 of Table 1), this certainty was only guaranteed to be at least 0.699. Thus, for the hypothetical example, it can be argued that two runs of 15 replicates per standard (or 30 replicates per standard) were more than adequate, and that, for purposes of economy, only 15 replicates per standard were run at iteration 1.

The data results for the hypothetical example are summarized in Table 3A. At iteration 1 of the simulated

Table 2. Random sample sequences for simulated example.

	Iteration 1[a]	Iteration 2[a]	
Time	Run 1 sample sequence	Run 1 sample sequence	Run 2 sample sequence
1	Ref	Test	Test
2	Ref	Ref	Test
3	Test	Test	Test
4	Ref	Test	Ref
5	Test	Test	Test
6	Test	Test	Ref
7	Test	Ref	Ref
8	Ref	Ref	Ref
9	Ref	Test	Test
10	Ref	Ref	Test
11	Ref	Ref	Test
12	Ref	Test	Ref
13	Ref	Ref	Test
14	Test	Ref	Test
15	Test	Test	Test
16	Ref	Ref	Ref
17	Ref	Test	Test
18	Ref	Test	Ref
19	Test	Ref	Ref
20	Test	Ref	Ref
21	Ref	Ref	Ref
22	Test	Test	Ref
23	Test	Ref	Test
24	Ref	Test	Test
25	Test	Ref	Ref
26	Test	Test	Ref
27	Test	Test	Test
28	Test	Test	Test
29	Test	Ref	Ref
30	Ref	Ref	Ref

[a] Ref, reference.

example, both the adjusted and unadjusted standards were each assayed in 15 replicates as per the data in Table 1 and the supposed corresponding sample sequence in Table 2. The difference between the two standards with respect to the log signal was estimated to be +0.0235 with a SE of 0.0081 (Table 3A). Using Eqs. 2 and 3, the lower and upper limits of a 95% confidence interval around this estimate were 0.0069 and 0.0401 log signal units, respectively. Because the confidence interval was not completely contained within the specification interval, [−0.025, +0.025], the algorithm proceeded to iteration 2.

The updated analyte concentration correction to the adjusted standard from iteration 1 was calculated by first solving Eq. 1 (given below), which required an estimate of the slope of the log signal calibration curve ($\hat{\beta}$) and the estimate D_1 (or $Y_{f1.} - Y_{a1.} = 0.0235$; Table 3A):

$$c_{i+1} = c_i + \frac{D_i}{i \times \hat{\beta}}.$$

From historical data, the slope of the calibration curve (β) in the vicinity of the fixed standard was approximated to be −0.44 log signal units/μg/L. Substituting these values of D_1 and $\hat{\beta}$ into Eq. 1 yielded the updated correction factor estimate c_2 (−0.0534 μg/L). Thus, for iteration 2, the original preparation of the adjusted standard from the first iteration, which for purposes of the algorithm was assumed to be 1 μg/L, was effectively diluted by the factor 1.0564 [1/(1 μg/L − 0.0534 μg/L)/1 μg/L].

For iteration 2 of the hypothetical example, both the adjusted and fixed standards were each run in 30 replicates (or 2 runs) as if the data in Table 1 had been generated from the corresponding random sample sequences in Table 2. The difference between the two standards with respect to log signal ($Y_{f2.} - Y_{a2.}$) was estimated to be 0.0066 log signal units with a SE of 0.0045 (Table 3A). Using Eqs. 2 and 3, the lower and upper limits of a 95% confidence interval around this estimate were −0.0024 and 0.0156 log signal units, respectively. Because the entire confidence interval (in log signal units) was contained between −0.025 and +0.025, the stochastic approximation algorithm stopped iterating and outputted 1 μg/L, the analyte concentration of the reference standard, as the estimated value assignment of the new production calibrator.

Table 3. Data summary of examples of random and trend-free design.

A. Hypothetical example with the random design

	Fixed-adjusted standard			95% confidence interval, log signal units		95% confidence interval, exponentiated	
Iteration	Log signal	SE	*df*	LL	UL	LL	UL
1	0.0235	0.0081	28	0.0069	0.0401	1.0069	1.0409
2	0.0066	0.0045	56	−0.0024	0.0156	0.9976	1.0157

B. Data example with the trend-free design

	Fixed-adjusted standard			95% confidence interval, log signal units		95% confidence interval, exponentiated	
Iteration	Log signal	SE	*df*	LL	UL	LL	UL
1	0.0234	0.0079	26	0.0072	0.0396	1.0072	1.0404
2	0.0068	0.0042	53	−0.0016	0.0152	0.9984	1.0153

Although the hypothetical example with the random design is useful for explication, in general it has certain weaknesses: A different random sample sequence must be constructed for each run; the experimental error can be inflated by within-run time trends and thus drive up the required number of replicates at each iteration; and the important assumptions of no standard-by-time-trend interactions within runs and no run-by-standard-effect interactions are not statistically checked. The presence of these interactions, which generally expresses some laboratory blunder in the value assignment procedures, can be rigorously checked using multiple regression analysis *(6)*, which in theory could be performed with the random design but with less statistical efficiency compared with using a trend-free design. Because of the aforementioned weaknesses with the random design, the actual data example was performed using multiple regression analysis with the quadratic trend-free design given in Table 1.

A Value Assignment Example Using Multiple Regression Analysis with a Trend-Free Design

This section assumes a familiarity with multiple regression analysis *(6)*. Readers without this preparation may want to skip to the next section.

An actual value assignment data set, which has intentionally been made anonymous, is provided in Table 1. These data were generated using the stochastic approximation algorithm on an in vitro medical diagnostic device with the first and second iteration consisting of one and two runs, respectively. To adjust the data for time trends, a nearly orthogonal, quadratic trend-free, 30-sample sequence design was used with each run (Table 1). The design was limited to a sequence of 30 by the constraints of the instrument. The near-orthogonality property of the sample sequence for fitting quadratic trend equations will be discussed later in this section.

Because this assay has historically shown quadratic time trends within assay runs, the following quadratic model was initially fitted by least-squares multiple regression *(6)* to each run:

$$Y_i = \alpha + \delta\, x_i + \beta_1\, t_i + \beta_{11}\, t_i^2 + \gamma_1\, t_i\, x_i + \gamma_{11}\, t_i^2\, x_i \qquad i = 1, \ldots, 30, \qquad (6)$$

where

Y_i is the log signal response of *i*th observation;
$x_i = 1$ if fixed standard;
$x_i = 0$ if adjusted standard; and
t_i is the sample order.

The above model in Eq. 6 was challenged by alternative models which permitted more complex time trends than quadratics such as the quadratic spline model with knot points at the 9th, 17th, and 25th time points and the linear spline model with knot points at the 5th, 9th, 13th, 17th, 21st, and 25th time points. Neither with respect to the root-mean-square error of the fitted equation nor with the SE of the estimate of the δ-coefficient did either of these spline models show any improvement over the fitted quadratic model in Eq. 6. In fact, the three models were comparable with respect to their estimates of δ and its SE.

An important assumption for the validity of the value assignment is that the test and reference standards do not have different time trends within a run, which could make the γ_1 and/or γ_{11} coefficients in Eq. 6 statistically significant. To check this assumption, the following reduced model was statistically tested against the full model in Eq. 4 using the reduction sum-of-squares *F*-statistic with numerator and denominator *df* of 2 and 24, respectively *(6)*:

$$Y_i = \alpha + \delta\, x_i + \beta_1\, t_i + \beta_{11}\, t_i^2. \qquad (7)$$

None of the runs in Table 1 showed a statistically detectable time-trend-by-standard-type interaction effect at the 0.05-level of significance. Thus Eq. 7 was used as the final fitted model for each run. The coefficient δ in Eq. 7 is the expected difference between the two standard types with respect to log signal response, whereas the coefficients β_1 and β_{11} account for quadratic time trends.

The design given in Table 1 is considered nearly orthogonal for fitting Eq. 6 to the data, because all of the model coefficients have design efficiencies ≥0.994, where a value of 1 indicates complete orthogonality *(8)*. The design efficiency for the model coefficient δ in Eq. 7, whose estimate is used to update the concentration correction factor, is 0.9997.

The following argument was used to determine that two runs of the trend-free design in Table 1 would be required for each iteration. As with the hypothetical example, the lower and upper specifications, LL and UL, were set to −0.025 and +0.025, respectively; σ was based on the historical estimate of 0.022; for the stopping rule, α was set at 0.05. To expand the basic quadratic trend model given by Eq. 7 to multiple runs, the degrees of freedom *v* in Eq. 11 (in *Appendix I*) were given by (number of runs) × *n* − 4 − (number of runs − 1) × 3, where *n* was the number of replicates per standard per iteration. If at some iteration the two standards were identical with respect to average signal response, we would want to be at least 0.95 certain that the algorithm would stop iterating. With the aforementioned values of LL, UL, σ, and α together with the methodology in *Appendix I*, this probability was calculated to be at least 0.982 for a sample size $n = 30$ for each standard (or two runs per iteration). For $n = 15$ replicates per standard (or one run per iteration), this probability was calculated to be at least 0.699. From these calculations, it was concluded that two runs of the trend-free 30-sample sequence were more than adequate for each iteration. Although two runs of the trend-free design were performed at iteration 2, for purposes of economy, only one run of the trend-free design was performed at iteration 1 (Table 1).

The results, which parallel the hypothetical example, are given in Table 3B. Because for iteration 1 the 95% confidence interval for the difference between the two standards with respect to average log signal response did not fall within the specification interval [−0.025, +0.025], the algorithm proceeded to iteration 2 with an updated correction of the adjusted standard. The updated concentration correction factor was calculated by substituting the least-squares estimate of δ (from fitting Eq. 7 to the iteration 1 data) for D_1 and the slope estimate of −0.44 for $\hat{\beta}$ into Eq. 1 to yield −0.053 μg/L as the updated concentration correction factor for the adjusted standard that was prepared from the calibrator bulk material. Thus, the updated, adjusted standard for iteration 2 was a 1.0562 [1/(1 μg/L − 0.0532 μg/L)/1 μg/L] dilution of the adjusted standard for iteration 1.

For iteration 2, both the new adjusted standard and the fixed standard were each run in 30 replicates (or two runs) using the same trend-free sample sequence that was used in iteration 1 (Table 1). The stochastic approximation algorithm assumes that the average difference between the fixed and adjusted standards with respect to log signal response is constant across runs within an iteration. Using the following fitted model, the two runs were tested for homogeneity with respect to the difference in log signal response between the fixed and adjusted standards:

$$Y_i = \alpha + \delta\, x_i + \delta_r\, r + \delta_{rx}\, rx_i + \beta_1\, t_i + \beta_{11}\, t_i^2 + \theta_1\, rt_i + \theta_{11}\, rt_i^2 \qquad (8)$$

where

$r = 1$ if run 2 and

$r = 0$ if run 1.

The data from iteration 2 did not statistically refute the assumption that the average difference between the adjusted and fixed standards with respect to log signal was constant across the two runs; from a two-sided *t*-test of the standard-type-by-run interaction, which was outputted from the multiple regression program SAS REG, the coefficient δ_{rx} was not judged to be statistically different from zero at the 0.05-level of significance. In this example, the coefficient δ_{rx} was properly considered a fixed (rather than random) effect that would only be nonzero if the experiment were performed inconsistently between the two runs. In other applications, it might be more appropriate to analyze this term as a random effect using a statistical software program such as SAS GLM. This issue will be addressed further in the *Discussion*.

Having not statistically refuted the assumption that the average difference between the adjusted and fixed standards with respect to log signal units was constant across runs, the final iteration 2 model was then:

$$Y_i = \alpha + \delta\, x_i + \delta_r\, r + \beta_1\, t_i + \beta_{11}\, t_i^2 + \theta_1\, rt_i + \theta_{11}\, rt_i^2 \qquad (9)$$

By way of least-squares multiple regression analysis (SAS REG), δ was estimated by 0.0068 log signal units with a SE of 0.0042 (Table 3B). The lower and upper limits of a 95% confidence interval for δ were −0.0016 and 0.0152, respectively (Table 3B). Because the entire confidence interval was contained within the specification interval [−0.025, 0.025], the algorithm stopped and outputted 1 μg/L, the known value of the fixed reference standard at iteration 2, as the value assignment of the production calibrator.

Summary of the Stochastic Approximation Method of Value Assigning Production Calibrators

Following is a sequential summary of the principal steps of the stochastic approximation method of value assigning production calibrators.

(*a*) Select a reference material that is compatible with the matrix and analyte species of the production calibrator material with respect to a specific assay system.

(*b*) Determine whether the production calibrator or reference material will be used to prepare the adjusted standard according to convenience and whether dilution and/or addition accuracy in the region of the fixed standard can be ensured.

(*c*) At the analyte level of the fixed standard, use historical data to estimate the within-run log signal standard deviation σ and the slope of the log assay response curve β, and determine the specification limits LL and UL.

(*d*) Prepare fixed and adjusted standards from the reference and production calibrator materials.

(*e*) Try to prepare the initial adjusted standard to be close to the fixed standard; set the concentration factor $C_1 = 0$.

(*f*) Using the relevant assay system, assay the fixed and adjusted standards in statistically determined replicates (*Appendix I*) using a trend-free design. For the first iteration, it might, for reasons of economy, be decided to run fewer than the statistically required number of replicates.

(*g*) Using multiple regression analysis, test each run of each iteration for within-time-trend-by-standard-type interactions and each iteration for standard-type-by-run interactions; if either of these statistical checks fail, troubleshoot the experiment and return to steps (*d*) or (*f*) as appropriate.

(*h*) Using the data from assaying the fixed and adjusted standards, construct a 95% confidence interval around their average difference in log signal units.

(*i*) Determine if the confidence interval is completely contained within the specification interval [LL, UL]. If the confidence interval is not contained within the specification interval, update the concentration correction factor according to Eq. 1 to prepare a new adjusted standard and return to step (*f*); if the confidence interval is contained within the specification interval, stop iterating and output the value assignment of the production calibrator as the known concentration of the reference standard for the current iteration.

Discussion

The stochastic approximation methodology permits value assignment to be accomplished on a single instrument with a single set of reagents. The author has verified this latter point on a wide variety of consulting problems where the stochastic approximation method was successfully used and implemented into manufacturing operations. With respect to setting the specifications LL and UL, an implicit assumption here is that the ratio of two standards with respect to average signal response does not vary across instruments. The validity of this assumption will depend on how constant the calibration curve slope function is across instruments. Experience has shown that for many applications, this is a reasonable assumption. For cases in which this assumption is not expected to hold, the user is advised to sufficiently narrow the accuracy specification and to set it according to the worst case (e.g., the instrument with the shallowest slope of the log assay response curve in the vicinity of the fixed standard).

If neither the reference nor the production calibrator material has adequate dilution accuracy and/or addition recovery, it might in some cases be convenient to transfer the value of the reference material to some transfer material that does have such accuracy and also has a matrix and analyte species that are functionally equivalent to those of the production calibrators.

If the production calibrators have dilution and/or addition accuracy, then, in theory, only the highest (lowest) calibrator needs to be value-assigned by way of the stochastic approximation algorithm.

There are cases in the medical device industry in which severe matrix effects occur with every production calibrator lot. One possible strategy for this difficult problem would be to use the stochastic approximation algorithm to force the production calibrators to mimic specific reference standards. The resulting inaccuracies that might necessarily occur in the production calibration curve could be minimized by narrowing the analyte concentration mesh between calibrators.

Whether the average difference between the fixed and adjusted standards with respect to log signal should be considered a fixed or random effect across assay runs depends upon the assay process. If there are factors that affect this average difference and typically vary across runs, then it might be more appropriate to analyze this difference as a random effect using a mixed linear model formulation *(10)* by way of some statistical software program such as SAS GLM. An example might be with a lyophilized material, in which new vials must be used with each run. The random variability in this case might enter from the volumetric step or from among-vial variability. If possible, steps should be taken to make such sources of random variability negligible (e.g., pooling of a statistically determined number of randomly selected vials, using large volumes for volumetric steps, and/or using calibrated pipettes). If the random variability of the standard effect across assay runs cannot be made negligible, this variability would need to be estimated and entered into the sample size calculations, in which case the sampling unit would be an assay run. (Methodology for determining required sample sizes when the sampling unit is an assay run and the variability is unknown is given in reference *11*.) If it can be assumed that such random factors do not exist when the assay is performed correctly, then the average difference should be analyzed as a fixed effect (as was done in the data example). In the case of the fixed-effect formulation, it is assumed that an interaction of the standard effect with runs would only be caused by the experiment having been performed incorrectly (e.g., an incorrect volumetric or gravimetric step).

For the laboratory that is interested in experimentally validating the stochastic approximation method, the following is proposed. In place of the production calibrator, a test material of known analyte concentration could be used; the final result of the algorithm would be compared with the known concentration of the test material. The experiment could be repeated with various concentration differences between the adjusted and fixed standard at the first iteration. Care should be taken that both the matrix and antigen species of the test and reference materials be functionally equivalent with respect to the specific assay system.

If a laboratory wants to compare the stochastic approximation method with some value assignment method that is based on fitting calibration curves, the following potential sources of bias should first be considered before the experiment is performed: systematic errors in fitting the calibration curve, errors in the value assignments of the calibrators that are used to fit the calibration curves, and within-run time trends. Even without systematic sources of bias, a proper statistical comparison with a value assignment method based on fitting calibration curves should account for interassay and possibly among-instrument variabilities *(11)*.

This paper has treated the ideal case in which the analyte value of the reference standard is known without error. In theory, the stochastic approximation method extends to the case in which the assigned value of the reference material has uncertainty: The uncertainty in the value of the reference standard is transferred to the production calibrator value assignment. Extending the methodology to this case is sensible, if and only if the manufacturing specification interval, [LL, UL], can be sufficiently narrowed to control the uncertainty propagated by the assigned value of the reference standard. If the lower and upper uncertainty limits of the reference standard assigned value are known (or can be estimated), the maximum propagated uncertainty interval induced by the prespecified values of LL and UL on the production calibrator value assignment can be estimated using the log signal calibration curve.

I thank Herbert Robbins for his helpful comments. I am deeply indebted to Karl Kennedy for his tireless help with computer programming, data analyses, and data gathering, without which the computerized implementation of this methodology would not have been possible. I am also grateful to William Present, who provided an industrial test site for the early development of the enclosed methodology.

Appendix I: Sample Size Determination

One possible criterion for determining the required number of replicates for each iteration of the algorithm is that if, at some iteration, the fixed and adjusted standards are identical with respect to their expected signal responses, there should be at least a 0.95 certainty that the algorithm will stop iterating (i.e., the 95% confidence interval constructed by Eqs. 2 and 3 or Eqs. 4 and 5 will be completely contained by the specification interval [LL, UL]). A lower bound on this probability is calculated by the following equation *(12)*:

$$\gamma = 1 - Pr\{u_\mathrm{i} > UL\} + Pr\{l_\mathrm{i} \leq LL\} \tag{10}$$

where

$$Pr\{u_\mathrm{i} > UL\} = 1 - PROBT(-t_\mathrm{i}, v_\mathrm{i}, ncu_\mathrm{i}) \tag{11}$$

$$Pr\{l_\mathrm{i} \leq LL\} = PROBT(t_\mathrm{i}, v_\mathrm{i}, ncl_\mathrm{i}). \tag{12}$$

PROBT, which can be calculated using the SAS statistical software package among others, is the cumulative distribution function of the noncentral *t*-distribution with v_i, degrees of freedom, and with noncentrality parameters, ncu$_i$ and ncl$_i$ *(13)*:

$$ncu_\mathrm{i} = \frac{-UL}{\sigma\sqrt{1/n_\mathrm{ai} + 1/n_\mathrm{fi}}} \tag{13}$$

$$ncl_\mathrm{i} = \frac{-LL}{\sigma\sqrt{1/n_\mathrm{ai} + 1/n_\mathrm{fi}}}. \tag{14}$$

σ is the historical, large sample estimate of the within-run SD of the log signal (approximately the CV of the untransformed signal response *(7)*); and

t$_i$ is the upper $1 - \alpha/2$ percentile of the Student's *t*-distribution with v_i degrees of freedom.

Thus, by setting $\gamma = \gamma_o \geq 0.95$ and $\alpha = \alpha_o \leq 0.05$, Eqs. 10–12 can be solved for the required number of replicates n (n$_{ai}$, n$_{fi}$) per standard per iteration. In all of the examples of this paper, $\gamma_o = 0.95$ and $\alpha_o = 0.05$.

Critical to the validity of the sample size calculations is an accurate estimate of σ, which should be obtained with the help of one's local statistician. The specification limits LL and UL should be chosen by careful study of appropriate historical log assay response curves in the region of the fixed standard; specifically, the first derivative function of the fitted calibration curve function, which relates the change in log signal response to change in analyte concentration, should be estimated.

Appendix II: SAS Computer Program for Generating Random Sample Sequences

The enclosed PC SAS program was used to generate the random sample sequences that are given in Table II for the hypothetical data example.

INPUTS:

seed = randomly chosen seed number that should be changed with each execution of the computer program;

run = number of runs requiring random sample sequences;

n = number of samples per run (an even number).

OUTPUT (written to the output file TEMP.DAT):

jj = index for run number;

ii = sequence number (running from 1 to n/2);

num = randomly generated number between 1 and n.

PROGRAM:

```
filename temp 'temp.dat';
data a;
file temp;
seed = 126578;
run = 3;
n = 30;
m = n/2;
large = n + 500;
do j = 1 to run;
jj = j;
array nn{100} n1-n100;
do k = 1 to m;
nn{k} = large;
end;
do i = 1 to m;
ii = i;
start: num = ranuni(seed);
div = 1/n;
result = num/div;
mult = int(result);
if result gt mult then num = mult + 1;
flag = 0;
do k = 1 to m;
if num eq nn{k} then flag = 1;
end;
if flag eq 1 then go to start;
if flag eq 0 then nn{i} = num;
put jj ii num;
end;
end;
```

References

1. Broughton PMG, Eldjarn L. Methods of assigning accurate values to reference serum. Part 1. The use of reference laboratories and consensus values, with an evaluation of a procedure for transferring values from one reference serum to another. Ann Clin Biochem 1985;22:625–34.
2. Broughton PMG, Eldjarn L. Methods of assigning accurate values to reference serum. Part 2. The use of definitive methods, reference laboratories, transferred values and consensus values. Ann Clin Biochem 1985;22:635–49.
3. Robbins H, Monro S. A stochastic approximation method. Ann Math Statist 1951;22:400–7.

4. Wei CZ. Asymptotic properties of least squares estimates in stochastic regression models. Ann Statist 1985;13(4):1498–1508.

5. Cochran WG, Davis M. The Robbins-Monro method for estimating the median lethal dose. J Roy Statist Soc B 1965;27:28–44.

6. Draper N, Smith H. Applied regression analysis, 2nd ed. New York: John Wiley, 1981:16–7.

7. Brownlee KA. Statistical theory and methodology in science and engineering, 2nd ed. New York: John Wiley, 1965:62.

8. Daniel C. Calibration designs for machines with carry-over and drift. J Qual Technol 1975;7(3):103–8.

9. Cox DR. Some systematic experimental designs. Biometrika 1951;38:312–23.

10. Scheffe H. Chapter 8. The analysis of variance. New York: John Wiley, 1959:261–90.

11. Schlain B. Chapter 14. Design, data and analysis by some friends of cuthbert daniel. Mallows CL, ed. New York: John Wiley, 1987: 291–304.

12. Schlain B. A 2-stage method for detecting reagent-to-sample carry-over on random access blood analyzers. Clin Chem 1996;42(5):725–31.

13. Johnson NL, Welch BL. Application of the non-central *t*-distribution. Biometrika 1939;31:362–89.