Variability of Molecular Biomarker Measurements from Nonlinear Calibration Curves[1]

Stephen J. Gange, Alvaro Muñoz,[2] Jia-Sheng Wang, and John D. Groopman

Departments of Epidemiology [S. J. G., A. M.] and Environmental Health Sciences [A. M., J. S. W., J. D. G.], The Johns Hopkins School of Hygiene and Public Health, Baltimore, Maryland 21205

Abstract

In any immunoassay experiment for the detection of molecular biomarkers, a nonlinear calibration curve is constructed to relate fixed biomarker concentrations to observed tracer levels. The biomarker concentration in an experimental sample can then be estimated by projecting the experimental tracer measurements through the inverse of the calibration curve. Once an estimate of the biomarker level has been calculated, it is often of interest to determine its variability. Typically, methods for estimating this variability assume that the biomarker variability is due solely to the uncertainty in the estimation of the calibration curve. A more complete analysis would combine this uncertainty with the variability in the processing and measurement of the sample, including, e.g., measurement error of laboratory procedures or variation in enzymatic activity in enzyme-linked immunosorbent assays or radioactivity counts in RIAs. In this paper, we present a method of estimating the variability of inverse estimates assuming there is variation arising from both the determination of the calibration curve and from the preparation and measurement of the experimental sample. Our method uses a resampling algorithm that avoids requiring many distributional assumptions present in alternative procedures, can be easily implemented, and is generalizable to any immunoassay procedure. Methods for incorporating our results in the estimation of variability for planning and analyzing biomarker experiments are discussed. We provide an example using RIA data for aflatoxin B_1 detection. These biomarkers for aflatoxin exposure are used in the analysis of serum aflatoxin adduct levels in human and experimental samples, and they are important in hepatocellular carcinoma research.

Received 4/12/95; revised 8/14/95; accepted 8/16/95.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by NIH Grants P01 ES06052, P30 ES03819, and 5T32 ES07141.

[2] Reprint requests should be directed to Department of Epidemiology, The Johns Hopkins University, 797 Hampton House, 624 North Broadway, Baltimore, MD 21205.

Introduction

One of the advantages of using molecular biomarkers for measuring exposure to xenobiotics is the ability to accurately measure very small levels of exposure in individuals. Techniques such as ELISAs and RIAs have been used extensively for this purpose in molecular epidemiological investigations (1). In RIA of a toxin, for example, one first constructs a calibration curve [$P = f(X|\theta)$; where P is the percentage of radioactivity inhibition and X is $\log_{10}$ (AFB_1[3] level)] by estimating the unknown parameters (θ) of a model using experimentally determined radioactivity measurements ($P = p^*$) on samples with known concentrations of the toxin ($X = x^*$). Because the relationship of the concentration and the counts is generally nonlinear, statistical methods more sophisticated than simple linear regression are usually required for estimating θ (2). Once parameter estimates $\hat{\theta}$ for the calibration curve are determined and once the amount of radioactivity $P = p$ in an experimental sample has been measured from the RIA, the concentration of the toxin in the sample can be estimated by using an inverse calculation: $\hat{x} = f^{-1}(p/\hat{\theta})$. Because $\hat{x}$ will be a marker of the exposure to the toxin, we refer to it here as the biomarker of interest.

In addition to the biomarker concentration point estimate, it is of interest to understand the "technical" variability (3) of the inverse estimate $\hat{x}$. That is, given many independent repetitions of the experiment on a set of homogeneous samples, from determining the calibration curve to the processing and measuring of the experimental samples, what would be the dispersion in the inverse estimate $\hat{x}$? This variability provides useful information regarding the laboratory precision with which the biomarker can be measured and can be used in sample size/power calculations for planning biomarker experiments. Furthermore, experiments that are designed to determine the efficacy of interventions on biomarker levels should use analytical approaches that incorporate this variability of $\hat{x}$.

Several methods have been developed that attempt to estimate the variability of inverse estimates. A traditional method is based on "fiducial intervals" (4, 5), which depend on the distribution of ratios of Gaussian random variables to define intervals for $\hat{x}$. A more recent proposal (6) relies on asymptotic δ method transformations of the variability of parameter estimates. A limitation of both of these methods, however, is that the source of the variability in $\hat{x}$ is considered to be solely due to the variability in the parameter estimates $\hat{\theta}$. That is, using the RIA example, the observed radioactivity in a particular sample is assumed to be determined without error, and any dispersion associated with $\hat{x}$ occurs because of variability of the parameter estimates $\hat{\theta}$, which determine the calibration curve.

Because the measured level of p in an experimental sample would be expected to possess variability due to uncertainty in

[3] The abbreviations used are: AFB_1, aflatoxin B_1; Ab, antibody.

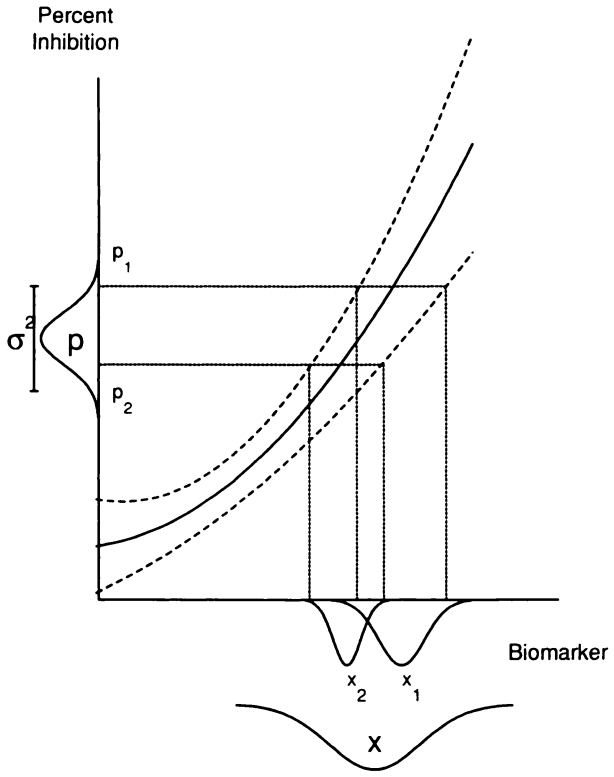

Fig. 1. Schematic illustrating the effect of variability in the radioactivity response (p) and the variability associated on the calibration curve (*dashed lines*) with the variability of the inverse estimate $x = f^{-1}(p)$.

radioactivity particle counts or enzymatic activity or slight fluctuations in laboratory preparation techniques, the variance estimate solely due to the variability of the parameter estimate of the calibration curve may be too small. Fig. 1 depicts the qualitative effect of variation in p in combination with variability from the parameter estimates. The increasing solid line represents a nonlinear calibration curve estimated from fixed biomarker samples. The dashed lines enveloping the calibration curve represent confidence bands for the curve that incorporate the variability of the parameter estimates $\hat{\theta}$. Standard methods assume that a fixed radioactivity count, say p_1, is obtained from a sample and is used to estimate the biomarker level $\hat{x}_1$. The variability of this estimate is determined by the uncertainty in the calibration curve, which is depicted with the distribution around $\hat{x}_1$.

We hypothesize that p_1 may be but one realization from an underlying true radioactivity count p that has a distribution depicted with variance σ^2. If we had repeated the experiment with a sample identical at collection but processed and measured independently, we might have obtained a different radioactivity count p_2, which would result in a different inverse estimate $\hat{x}_2$. Hence, we can see that the variability in the true biomarker estimate $\hat{x}$ will be dependent on both the variation in the radioactivity counts (p), and the uncertainty of the parameters ($\hat{\theta}$) that determine the calibration curve.

In this study, we describe a general method for estimating the variability of an inverse estimate when both the response and the calibration curve are subject to variability. Our method is based on a resampling algorithm that can be programmed by using statistical software such as SAS or S, has the advantage of not requiring the strict mathematical structure or large sample approximations found in other methods, and is applicable to any immunoassay method that uses a calibration curve. We demonstrate the use of the method with the analysis of a RIA calibration curve for experimental samples of AFB_1 for application to a serum aflatoxin-albumin biomarker. This biomarker has been suggested for use in both monitoring human exposure and has application to preventative interventions against hepatocellular carcinoma (7, 8). Lastly, we describe how to use the results from our method in the analysis and planning of biomarker experiments.

Materials and Methods

Estimation of the Calibration Curve. The RIA methods used to create the AFB_1 calibration curve were modifications of established procedures (9). Briefly, seven serial dilutions of unlabeled AFB_1 were made in triplicate to obtain samples with known AFB_1 concentrations (x^*) of 0.15, 0.45, 1.37, 4.12, 12.35, 37.04, and 111.11 pmol. Each of these 21 samples was processed independently by adding equal amounts of [^{3}H]AFB_1 tracer and monoclonal antibody 2B11 (Ab). The specific amount of both antibody and tracer used in this procedure had been determined experimentally such that the amount of antibody used could bind only 50% of the tracer added to each assay tube. Thus, the addition of the graded amounts of unlabeled AFB_1 to each assay tube established a competitive reaction for the fixed amount of antibody between the limiting amount of tracer ([^{3}H]AFB_1) and the unlabeled AFB_1 present in the assay tube. At equilibrium, the amount of Ab bound to the tracer (*i.e.*, the [^{3}H]AFB_1-Ab complex) will be dependent on the level of formation of Ab bound to the unlabeled AFB_1 (*i.e.*, the AFB_1-Ab complex). Therefore, with greater amounts of unlabeled AFB_1 present in an assay tube, there will be less Ab bound to the tracer.

AFB_1-antibody complex (unlabeled AFB_1-Ab and [^{3}H]AFB_1-Ab) were precipitated by saturated ammonium sulfate via centrifugation, and the radioactivity levels of the remaining supernatant were measured independently in a scintillation counter for 10 min. The p^* for each sample was determined from the ratio of the difference between the radioactivity counts for the experimental sample and a control sample with Ab and [^{3}H]AFB_1 divided by the difference between the radioactivity counts for a control sample with only [^{3}H]AFB_1 (no Ab) and a control sample with Ab and [^{3}H]AFB_1.

We used a two-parameter logistic model to describe the relationship between the aflatoxin level and inhibition. The form of the model was:

$$P = \frac{\exp(\alpha + \exp(\beta) \cdot X)}{1 + \exp(\alpha + \exp(\beta) \cdot X)}$$

where $\theta = (\alpha, \beta)$ are model parameters. The use of $\exp(\beta)$, instead of a traditional parameter β, is consistent with an expected positive association between the aflatoxin level and the radioactivity inhibition. It is important to note that the methods which follow are not dependent of this specific choice of the nonlinear model. If the data had not been in the range of 0 to 1, we could have chosen a more general 3- or 4-parameter logistic model (2) to allow for estimation of minimal or maximal responses. We might have also chosen another model form, such as one that relates the probit-transformed inhibition response as a function of AFB_1 concentration.

Using the calibration data (*i.e.*, experimentally determined

percentage inhibition $P = p^*$ for fixed aflatoxin levels $X = x^*$), the model in Eq. A was fit using a standard nonlinear regression algorithm (10) to obtain parameter estimates ($\hat{\alpha}$, $\hat{\beta}$).

Variability of Inverse Calibration Curve Estimates. As described in the "Introduction," we are interested in estimating the variability of inverse estimates $\hat{x} = ([\log\{p/(1 - p)\} - \hat{\alpha}])/\exp(\hat{\beta})$ given variation in both the nonlinear regression model parameter estimates ($\hat{\alpha}$, $\hat{\beta}$) and the radioactivity inhibition response p from a particular experimental sample. This will be denoted by $\mathrm{Var}(\hat{x})$, although it should be noted that this is conditional on three components: (*a*) the (x^*, p^*) data for the estimation of the calibration curve; (*b*) the dispersion estimate of p; and (*c*) a fixed inhibition chosen *a priori*.

Estimates of the parameters in the logistic model are obtained by using a nonlinear least-squares algorithm. Thus, the dispersion of the parameter estimates can be estimated using the variance-covariance matrix $\hat{\Sigma}$ of the asymptotic Gaussian distribution as described by Jennrich (11).

The choice of variability in the response requires more consideration. Conceptually, we are interested in the variation of immunoassay responses that would occur among experimental samples that were initially indistinguishable but subjected to independent processing and measurement. This is not the same as the variability that would occur if a single sample was divided into many subsamples after processing but before the measurement of radioactivity levels. Nor is this the same as the variability that would occur among heterogeneous samples processed and measured independently. In the former case, the variability of interest would be underestimated, because the dispersion of the outcomes would be due solely to variability of radioactivity counts (*e.g.*, as in multiple readings from a liquid scintillation counter) and not from both radioactivity measurement and laboratory preparation. In the latter case, the variability of interest would be overestimated because the heterogeneity in the samples before preparation would contribute an undesired additional variability component.

An additional consideration in the specification of the response variability is the potential limitation of the response range. For our example, the response of interest is the sample radioactivity inhibition relative to a standard, which is limited to the range 0 to 1. Assuming an additive Gaussian random variation to this response would not be technically consistent because the response with error would be allowed to have an unlimited range. For our model, we can make an easy transformation of the response so that we assume an additive Gaussian variation is applicable to the log-odds (logit) of the inhibition: $\log[P/(1 - P)]$.

Hence, to estimate the variability of the logit responses, denoted by σ^2, we computed the sample variance among the three observed replicates in the logit scale for each of the seven fixed AFB_1 concentrations. Because each of these three samples were originally identical and because each was processed and measured independently the variation observed among these samples was due solely to the differences in the processing and measurement of the samples. Because of the small number of replicates at each fixed AFB_1 dose, we used the average variance over all seven concentrations.

The algorithm that we propose to calculate the variability of an inverse estimate $\hat{x}$ for some fixed value p_0 can be described as repeated iterations of four steps: (*a*) from a normal distribution with mean $\mathrm{logit}(p_0)$ and variance σ^2, generate a random $\mathrm{logit}(\hat{p})$. Transform to obtain $\hat{p}$; (*b*) generate 500 independent bivariate parameter vectors $(\alpha_1, \beta_1), \ldots, (\alpha_{500}, \beta_{500})$ from a bivariate normal distribution with mean ($\hat{\alpha}$, $\hat{\beta}$) and variance-covariance matrix $\hat{\Sigma}$; (*c*) for each set (α_i, β_i), generate the inverse estimates of $\hat{p}$ as $\hat{x}_i = (\log(\hat{p}/(1 - \hat{p})) - \alpha_i)/\exp(\beta_i)$. Note that the choice of $\exp(\beta_i)$ prevents any problems with division by zero, and hence no restrictions on the range of (α_i, β_i) are required; and (*d*) calculate the sample mean $\hat{m}_1$ and sample variance $\hat{s}_1^2$ of the set $\hat{x}_1, \ldots, \hat{x}_{500}$ and save these values for the remainder of the algorithm. These four steps are repeated 400 times, *i.e.*, we choose another random $\mathrm{logit}(\hat{p})$, generated another 500 parameter values, etc. This results in 400 sample mean estimates ($\hat{m}_1, \ldots, \hat{m}_{400}$) and variance estimates ($\hat{s}_1^2, \ldots, \hat{s}_{400}^2$) from step (*d*). To obtain the overall estimate of the variance of the inverse estimate, we sum the mean of the 400 sample variances and the variance of the 400 sample means around the overall mean $\bar{m}$ of the $\hat{m}_1, \ldots, \hat{m}_{400}$:

$$\mathrm{Var}(\hat{x}) = \left\{400^{-1} \sum_{i=1}^{400} \hat{s}_i^2\right\} + \left\{399^{-1} \sum_{i=1}^{400} (\hat{m}_i - \bar{m})^2\right\}$$

We chose to use 400 samples of 500 observations because after replicating the simulations ten times (*i.e.*, obtaining ten estimates of $\mathrm{Var}(\hat{x})$), the coefficient of variation of these ten replicates was <33% with this sample size. For more precise estimates, larger sample sizes could be used.

Table 1 Experimental inhibition measurements (p^*) for three replicate samples at seven fixed aflatoxin levels (x^*) used to construct the calibration curve.

Aflatoxin level (pmol)	Log_{10} aflatoxin	Inhibition		
0.15	−0.8239	0.0596	0.0615	0.0438
0.45	−0.3468	0.1143	0.1112	0.1026
1.37	0.1461	0.2619	0.2540	0.2774
4.12	0.6128	0.5188	0.5290	0.5389
12.35	1.0899	0.8199	0.8077	0.8209
37.04	1.5682	0.9542	0.9470	0.9347
111.11	2.0453	0.9803	0.9795	0.9653

Results

The data from the 21 samples used to estimate the calibration curve are shown in Table 1. The estimated parameters for the model in Eq. (A) were $\hat{\alpha} = -1.3724$ and $\hat{\beta} = 0.9323$. The sample covariance matrix $\hat{\Sigma}$ for the parameters had diagonal elements 0.0431^2 and 0.0243^2 and off diagonal elements (−0.7769) (0.0431) (0.0243). The fitted regression line is displayed with the raw data in Fig. 2. By using these parameter estimates, Table 2 shows the point estimates $\hat{x}$ for values of p_0 from 0.10 to 0.50, which were calculated via:

$$\hat{x} = \frac{[\log\{p_0/(1 - p_0)\} - (-1.3724)]}{\exp(0.9323)}$$

For example, the $\log_{10}$ AFB_1 dose, which evoked an inhibition of 0.50, was estimated to be (0 + 1.3724)/exp(0.9323) = 0.5402.

The four-step algorithm in the previous section was run with the variance σ^2 of the logit responses equal to $0.0275 = 0.1658^2$, the average of the seven variances computed for each replicate. For comparison, we also calculated $\mathrm{Var}(x)$ under the assumption that $\sigma^2 = 0$, which would be similar to the variance obtained using alternative methods based on the δ method, which ignores the variability in p.

The results using the two methods are displayed in Table 2. Under the assumption of no measurement error ($\sigma^2 = 0$), but

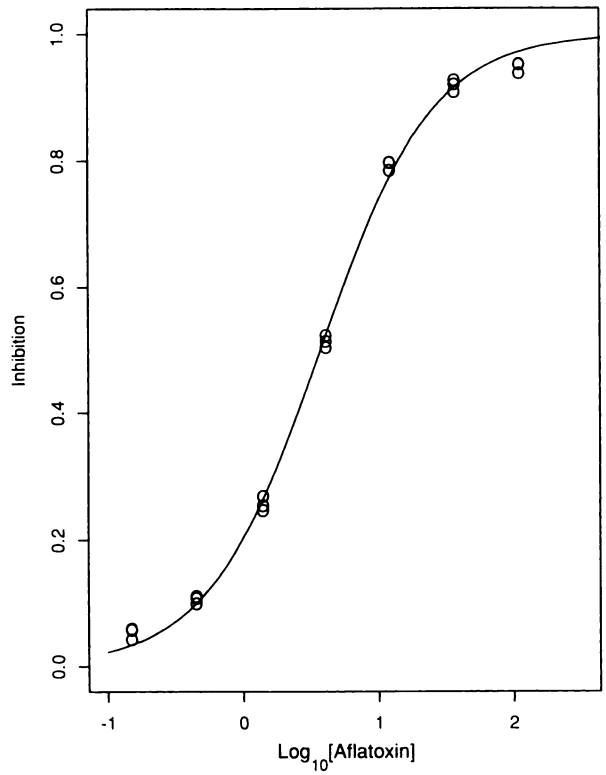

Fig. 2. Fitted nonlinear calibration curve with experimental data points.

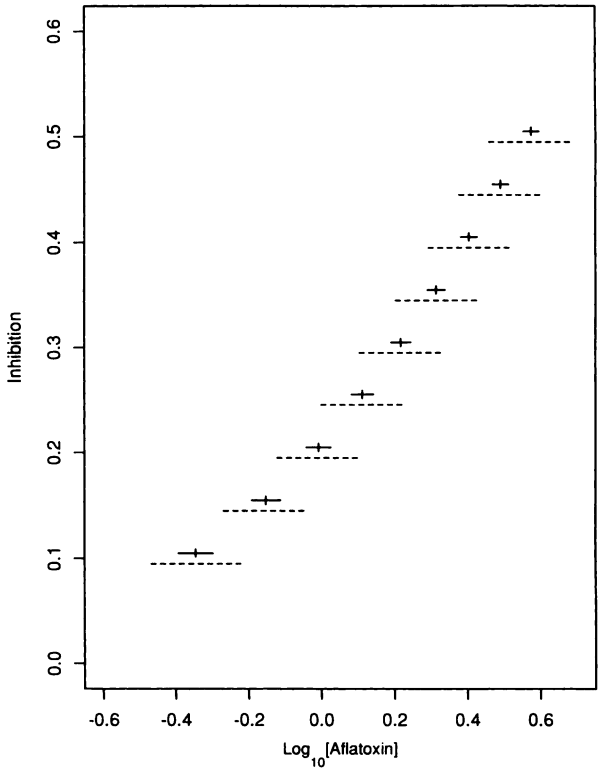

Fig. 3. A comparison of confidence intervals for inverse estimates for inhibition in the range 0.10–0.50. *Solid lines*, intervals assuming no measurement error; *dashed lines*, intervals assuming measurement error with $\sigma = 0.1658$. *Vertical ticks*, inverse point estimate $\hat{x}$.

Table 2 Inverse point estimates ($\hat{x}$) for fixed inhibition levels (p) with resampling SE estimates from the measurement error model under the assumption that p is measured without error ($\sigma = 0$) or with additive Gaussian error on the logit scale with SD $\sigma = 0.1658$.

Inhibition (p)	Log_{10} aflatoxin estimate ($\hat{x}$)	SE estimates	
		$\sigma = 0$	$\sigma = 0.1658$
0.10	−0.3247	0.02375	0.06813
0.15	−0.1426	0.01984	0.06850
0.20	−0.0055	0.01716	0.06624
0.25	0.1078	0.01508	0.06837
0.30	0.2067	0.01355	0.06433
0.35	0.2966	0.01236	0.06452
0.40	0.3806	0.01149	0.06615
0.45	0.4612	0.01098	0.06577
0.50	0.5402	0.01081	0.06447

incorporating uncertainty in the calibration curve parameter estimates, the estimates of the SE of $\hat{x}$ decrease from a high of 0.02375 at an inhibition of 0.10 to a low of 0.01081 at an inhibition of 0.50. This difference occurs because, as with standard linear regression, the calibration curve has smaller variability around the mean and higher variability around the extremes. Thus, if we had included values increasing from 0.50 to 0.95, we would have seen an symmetric increase in the variability.

The variance estimates from the model incorporating measurement error with $\sigma = 0.1658$ are larger by an approximately constant amount of 0.004. Thus, because the SEs under $\sigma = 0$ are decreasing for p between 0.10 and 0.50, the relative magnitude of the SEs under $\sigma = 0.1658$ increases by a factor of 2.9–6.0. The estimate of $\text{Var}(\hat{x})$ under $\sigma = 0.1658$ decreases only slightly as p varies from 0.10 to 0.50. This indicates that the major contribution to the total variance is due to the variance of p and not to the estimation of the calibration curve.

The inverse point estimates and approximate confidence intervals, calculated as $\hat{x} \pm 1.96$ times the SEs in Table 2, are displayed in Fig. 3. It is shown that the measurement error substantially contributes to the uncertainty of the estimation of the biomarker level for a given observed inhibition.

Discussion

Our method uses a resampling algorithm that does not require specific distributional assumptions present in other procedures. The key relationship that forms the basis for the resampling algorithm is from a general formula that expresses an unconditional variance in terms of conditional expectations and variances:

$$\text{Var}(\hat{x}) = \text{Exp}(\text{Var}(\hat{x}|\hat{p}_i)) + \text{Var}(\text{Exp}(\hat{x}|\hat{p}_i)) \qquad \text{(B)}$$

The left side of this equation is the "unconditional" variance of the inverse estimate, which is what we are interested in calculating (as discussed previously, this estimate is still conditional on the calibration curve data (x^*, p^*), σ^2, and the value of p that is of interest). The right side of the equation is the sum of two components. The first component requires the computation of the conditional variance of the inverse estimates, denoted in the algorithm as $\hat{s}_i^2$, where we condition on a single estimate of $\hat{p}_i$. We then take the mean, over different values of $\hat{p}_i$, of these

conditional variance estimates. The second component requires the computation of the conditional expectation of the inverse estimates, denoted in the algorithm as $\hat{m}_i$ in which we again condition on a single estimate of $\hat{p}_i$. We then take the variance of these conditional mean estimates, where the variance is over different values of $\hat{p}_i$.

The methods described above are appropriate for the calculation of Var($\hat{x}$) whether the distribution of $\hat{x}$ is Gaussian. By assuming $\hat{x}$ follows a normal distribution, familiar techniques can be used to compute the 95% confidence intervals depicted in Fig. 3 (*i.e.*, $\hat{x} \pm 1.96(\text{Var}(\hat{x}))^{1/2}$). If this assumption is not valid, we can still use the proposed algorithm to generate inverse estimates $\hat{x}$ that incorporate variability in the calibration curve and in p. Instead of computing summary statistics $\hat{m}_i$ and $\hat{s}_i^2$ for use in Eq. (B), we can save all 500 estimates of $\hat{x}$ generated at each step of the algorithm. A 95% confidence interval for $\hat{x}$ can then be calculated by computing the 2.5 and 97.5 percentiles of all the 400*500 estimates of $\hat{x}$. For example, using this method to compute a 95% confidence interval for the inverse estimate at $p = 0.10$, we obtained an interval of -0.451 to -0.197, which was only slightly shorter and slightly shifted to the right than the interval using a normal approximation: (-0.458, -0.191).

There is another instance in the algorithm where normality is assumed: (α_i, β_i) is a random draw from a bivariate normal distribution with mean $(\hat{\alpha}, \hat{\beta})$ and covariance $\hat{\Sigma}$. This can be justified by using the asymptotic distribution of nonlinear least squares estimates (11). However, it is important that the parameters be expressed in such a way that there is no restriction in the range as required by the normal distribution. In our case, this is accomplished by parameterizing the coefficient of X in Eq. (A) as $\exp(\beta)$ to allow β to be unrestricted. The algorithm can still be used, however, in cases where the parameters of the nonlinear model are not Gaussian; the only change is that the parameter generation in step (2) of the algorithm is done from an appropriate non-Gaussian distribution.

This algorithm can also be used when additional sources of variability or measurement error are present (3). That is, in the above derivation, we assumed that σ^2 represented the variability of the response that would occur from independent processing and measurement of a set of homogeneous samples. However, suppose that the original samples were from a mixture of heterogeneous populations that induces an additional component of variation (*e.g.*, within-individual biological variation from repeated biomarker measurements on a single person or between-individual variation such as seasonal variation). Then, the choice of σ^2 could be modified to account for different sources of variability.

The variability obtained using our method should be used when estimating the sample size of an experiment in which laboratory variability and measurement error is large relative to the measurement of interest. Because the sample size is directly proportional to the variance, any increase in the variance because of measurement error will result in an increase of the sample size for detecting a fixed treatment difference. Strictly speaking, we have provided a method for estimating Var($x|p_0$) for a fixed value p_0. Investigators planning a study should, therefore, provide a range of expected values p_0 from which Var($x|p_0$) and the corresponding sample sizes are calculated.

The results of our methods should also be incorporated in the statistical analysis of biomarker data. In the context of comparing two groups, suppose we have obtained experimental radioactivity measurements from n_T individuals in a treatment group (p_i, $i = 1, \ldots, n_T$) where T indicates a measurement from individuals in a treatment group and n_C individuals in a control group (q_j, $j = 1, \ldots, n_C$) where C indicates a measurement from individuals in the control group. A calibration curve can be used to obtain estimated biomarker concentrations ($x_1, \ldots, x_{n_T}$), and ($z_1, \ldots, z_{n_C}$), via inverse estimation for each group, respectively, and let Var(x_i) and Var(z_j) be the variances calculated using the methods proposed in this paper. The correct method for computing a confidence interval for the true difference of the two group means (Δ) uses a resampling algorithm with K iterations that can be described as follows. For every iteration $k = 1, \ldots, K$, generate a random $x_i^{(k)}$ from a normal distribution with mean x_i and variance Var(x_i) for $i = 1, \ldots, n_T$. Similarly, generate a random $z_j^{(k)}$ from a normal distribution with mean z_j and variance Var(z_j) for $j = 1, \ldots, n_C$. Compute $\hat{\Delta}^{(k)} = \bar{x}^{(k)} - \bar{z}^{(k)}$ and $V^{(k)} = (n_T + n_C - 2)^{-1} (\Sigma_i(x_i^{(k)} - \bar{x}^{(k)})^2 + \Sigma_j(z_j^{(k)} - \bar{z}^{(k)})^2)$. From the K samples, the estimated difference of the group means is $\hat{\Delta} = K^{-1} \Sigma_k \hat{\Delta}^{(k)}$, which has variance

$$\text{Var}(\hat{\Delta}) = \left(\frac{1}{n_T} + \frac{1}{n_C}\right)\left\{\frac{1}{K}\sum_k V^{(k)} + \frac{K+1}{K}\frac{1}{K-1}\sum_k(\hat{\Delta}^{(k)} - \hat{\Delta})^2\right\}$$

A 95% confidence interval for Δ is constructed as $\hat{\Delta} \pm 1.96 \cdot (\text{Var}(\hat{\Delta}))^{1/2}$.

In summary, this paper proposes an algorithm that is useful for estimating the variance of inverse estimates from nonlinear calibration curves when both the calibration curve and sample response are assumed to be measured with error. As we have shown in an example with aflatoxin data, the variability estimates obtained from this algorithm can be much larger than those obtained with traditional methods which ignore the sample response variability. As described above, the results from the proposed algorithm can be used in the planning and analysis of biomarker experiments.

References

1. Schulte, P. A., and Perera, F. P. (eds.), Molecular Epidemiology: Principles and Practices. San Diego, CA: Academic Press, 1993.
2. Robard, D., and Frazier, G. R. Statistical analysis of radioligand assay data. Methods Enzymol. *37:* 3–22, 1975.
3. Vineis, P., Schulte, P. A., and Vogt, R. F. Technical variability in laboratory data. *In:* P. A. Schulte and F. P. Perera (eds.), Molecular Epidemiology: Principles and Practices. San Diego, CA: Academic Press, 1993.
4. Finney, D. J. Radioligand assay. Biometrics, *32:* 721–740, 1976.
5. Finney, D. J. Statistical Method in Biological Assay (Ed. 3). London, United Kingdom: Charles Griffin & Co., 1978.
6. Schwenke, J. R., and Milliken, G. A. On the calibration problem extended to nonlinear models. Biometrics, *47:* 563–574, 1991.
7. Groopman, J. D., Wogan, G. N., Roebuck, B. D., and Kensler, T. W. Molecular biomarkers for aflatoxins and their application to human cancer prevention. Cancer Res., *54:* 1907s–1911s, 1994.
8. Helzlsouer, K. J. Serological markers of cancer and their application in clinical trials. Cancer Res., *54:* 2011s–2014s, 1994.
9. Groopman, J. D., Trudel, L. J., Donahue, P. R., Rothstein, A., and Wogan, G. N. High affinity monoclonal antibodies for aflatoxins and their application to solid phase immunoassay. Proc. Natl. Acad. Sci. USA, *81:* 7728–7731, 1984.
10. Bates, D. M., and Chambers, J. M. Nonlinear models. *In:* J. M. Chambers and T. J. Hastie (eds.), Statistical models in S, Chap. 10. Pacific Grove, CA: Wadsworth & Brooks, 1992.
11. Jennrich, R. I. Asymptotic properties of nonlinear least squares estimators. Ann. Math. Stat. *40:* 633–643, 1969.