

# Reliability of Pharmacodynamic Analysis by Logistic Regression

## A Computer Simulation Study

Wei Lu, Ph.D.,\* James M. Bailey, M.D., Ph.D.†

**Background:** Many pharmacologic studies record data as binary yes-or-no variables, and analysis is performed using logistic regression. This study investigates the accuracy of estimation of the drug concentration associated with a 50% probability of drug effect ( $C_{50}$ ) and the term describing the steepness of the concentration-effect relation ( $\gamma$ ).

**Methods:** The authors developed a technique for simulating pharmacodynamic studies with binary yes-or-no responses. Simulations were conducted assuming either that each data point was derived from the same patient or that data were pooled from multiple patients in a population with log-normal distributions of  $C_{50}$  and  $\gamma$ . Coefficients of variation were calculated. The authors also determined the percentage of simulations in which the 95% confidence intervals contained the true parameter value.

**Results:** The coefficient of variation of parameter estimates decreased with increasing  $n$  and  $\gamma$ . The 95% confidence intervals for  $C_{50}$  estimation contained the true parameter value in more than 90% of the simulations. However, the 95% confidence intervals of  $\gamma$  did not contain the true value in a substantial number of simulations of data from multiple patients.

**Conclusion:** The coefficient of variation of parameter estimates may be as large as 40–50% for small studies ( $n \leq 20$ ). The 95% confidence intervals of  $C_{50}$  almost always contain the true value, underscoring the need for always reporting confidence intervals. However, when data from multiple patients is naively pooled, the estimates of  $\gamma$  may be biased, and the 95% confidence intervals may not contain the true value. (Key words: Anesthetic effect; binary data; Monte Carlo Studies.)

PHARMACODYNAMIC data are often recorded as binary variables (e.g., whether the patient responds to command, whether the patient can maintain adequate spon-

taneous ventilation, whether the patient shows a hemodynamic or somatic response to surgical stimulus). In this situation, a common technique of data analysis is logistic regression, in which the probability ( $P$ ) of drug effect is evaluated as a function of the drug concentration in plasma or at the effect site ( $C$ ), using the following equation:

$$P = C^\gamma / (C_{50}^\gamma + C^\gamma) \quad (1)$$

where  $C_{50}$  is the concentration at which the probability of drug effect is 50% and  $\gamma$  is a measure of the steepness of the concentration-effect curve.‡ Values of  $C_{50}$  and  $\gamma$  are estimated by expressing the logarithm of the likelihood of the observed results (i.e., log likelihood) using equation 1 and then maximizing this with respect to  $C_{50}$  and  $\gamma$ . Logistic regression has been used for the analysis of the pharmacodynamics of inhaled and intravenous anesthetics,<sup>1-11</sup> usually with a primary focus of determining values of  $C_{50}$  and  $\gamma$ . However, any statistical technique should not only provide parameter estimates, but also measures of the accuracy of these estimates. Although typical statistical software programs provide confidence intervals for estimated parameters, because logistic regression is a nonlinear technique, these confidence intervals are only valid asymptotically as the number of data points ( $N$ ) goes to infinity.<sup>12</sup> In reality, most studies in the anesthesia literature involve relatively small numbers of data points. Therefore, it is often unclear how reliable parameter estimates are. The purpose of this study was to use simulations to analyze the relation between sample size and the accuracy of parameter estimates.

\* Research Fellow.

† Associate Professor.

Received from the Department of Anesthesiology, Emory University School of Medicine, Atlanta, Georgia. Submitted for publication December 8, 1998. Accepted for publication November 3, 1999. Support was provided solely from institutional and/or departmental sources.

Address reprint requests to Dr. Bailey: Department of Anesthesiology, Emory University School of Medicine, 1364 Clifton Road NE, Atlanta, Georgia 30322. Address electronic mail to: james\_bailey@emory.org

‡ Throughout this article,  $P$  refers to the probability of drug effect (e.g., the probability of ablating the response to some stimulus such as skin incision).

**Methods**

Although logistic regression commonly is used in anesthesiology research, the underlying statistical model is seldom described. A useful general reference for these types of statistical models is the textbook by Cox and Snell.<sup>12</sup> This model assumes that anesthetic effect is measured by an underlying continuous variable, denoted  $y$ , which is related to the drug concentration by the following equation:

$$y = \gamma \ln C - \gamma \ln C_{50} + x \tag{2}$$

where  $x$  is a random variable that is assumed to have a logistic distribution. The logistic distribution is described by

$$P(x) = \frac{\exp(-x)}{[1 + \exp(-x)]^2}$$

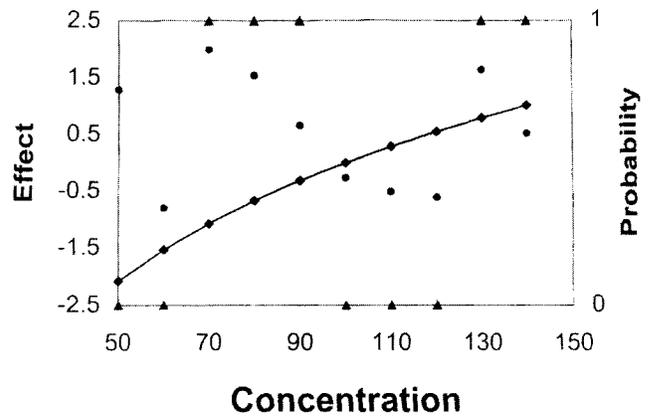
where  $P(x)$  is the probability of a given value of  $x$ . This distribution has a mean of zero and a variance of  $\pi^2/3$ . The model is applicable to binary yes-or-no data because we assume that a positive drug effect is observed only if  $y > 0$  or equivalently if  $x > -(\gamma \ln C - \gamma \ln C_{50})$ .

Thus, the probability of drug effect is as follows:

$$P = \int_{-(\gamma \ln C - \gamma \ln C_{50})}^{\infty} \frac{\exp(-x)}{[1 + \exp(-x)]^2} dx = C^\gamma / (C_{50}^\gamma + C^\gamma)$$

We attempted to illustrate the underlying model in figure 1. The solid line through solid diamonds depicts the deterministic part of the  $\gamma \ln C - \gamma \ln C_{50}$ ; the adjective “deterministic” refers to the fact that this part of the effect is determined solely by the drug concentration and the pharmacodynamic parameters  $C_{50}$  and  $\gamma$ . This increases steadily as drug concentration ( $C$ ) increases. The solid circles are random numbers conforming to a normal distribution; we used a normally distributed variable for convenience because of the ready availability of normally distributed random-number generators in standard statistical software packages. When the sum of the deterministic component (solid diamonds) and the random variables (solid circles) is greater than or equal to zero, the observed drug effect (solid triangles) is one, and if the sum is less than zero, the observed drug effect is zero.

To investigate the standard deviations of parameter estimates using logistic regression, simulations of typical anesthesia studies were performed. Simulations were performed for varying values of  $n$  (the number of data



**Fig. 1.** An illustration of the underlying statistical model of logistic regression. The solid line (◆-◆) depicts the deterministic part of the effect,  $\gamma \ln C - \gamma \ln C_{50}$  for  $\gamma = 3$  and  $C_{50} = 100$ , where  $C$  is the drug concentration and  $C_{50}$  is the drug concentration associated with a 50% probability of drug effect. The filled circles (●) are the realizations of a random variable with a normal distribution. A normally distributed variable was used for convenience because of the ready availability of normally distributed random-number generators in standard statistical software packages. The magnitude of these two variables is shown on the left-hand effect axis. When the sum of the deterministic component (◆-◆) and the random variables (●) is greater than or equal to zero, the simulated response (▲) is one, and if the sum is less than zero, the simulated response is zero.

points) and  $\gamma$ , assuming that equation 2 was the underlying model and  $C_{50} = 100$  (a valid assumption because the units of concentration are arbitrary). To begin each simulation,  $n$  data points were generated by randomly selecting  $n$  drug concentrations distributed uniformly on a logarithmic scale from 25 to 400 units using the random number generator of an Excel (Microsoft, Redmond, WA) spreadsheet. In a human study, this corresponds to the investigator assigning a drug dose to each patient enrolled in the study. At this point, if the concentration effect is totally deterministic (*i.e.*, if  $\gamma$  is infinite), then a positive drug effect will be observed if the drug concentration ( $C$ ) assigned to the data point exceeded  $C_{50}$ . However, as noted previously, there is an element of randomness in the concentration-effect relation, embodied in the fact that  $\gamma$  is finite. To take into account this randomness, a uniformly distributed random variable from 0 to 1 was generated for each “patient,” again using the random-number generator of an Excel spreadsheet. If this number was less than  $C^\gamma / (C_{50}^\gamma + C^\gamma)$  the simulated patient was assumed to have a positive drug effect (the response variable [R] was given a value of 1). Otherwise, it was assumed that  $R = 0$ . In this manner, responses are obtained for a range of concentrations, and spreadsheets consisting of columns

of C (the drug concentration) and R (the response variable) are generated. For each simulation, the parameters ( $C_{50}$ ,  $\gamma$ ) were estimated by maximum likelihood estimation. The logarithm of the likelihood of observed results (LL) was maximized as a function of  $C_{50}$  and  $\gamma$

log likelihood

$$= \sum_i \{R_i \log (P_i) + (1 - R_i) \log (1 - P_i)\} \quad (3)$$

( $P_i$  is given by equation 1 for each value of C;  $R = 1$  if a drug effect is observed and  $R = 0$  otherwise; and  $i$  indexes data points). Equation 3 is the sum of the logarithms of the probabilities of independent outcomes, which is  $P_i$  when  $R_i = 1$  and  $1 - P_i$  when  $R_i = 0$ . The estimation procedure was implemented on the Excel spreadsheet, taking advantage of its Solver function. In some instances, the attempted maximization of the log likelihood failed with the Solver function returning a “#NUM” message, indicating it could not interpret the log likelihood as a number. In these cases, the simulated data set was discarded and a new set generated.

Two separate simulation experiments were conducted. In the first, simulations were performed for  $\gamma = 1.0, 1.5, 3.0, 4.5, 6.0,$  and  $7.5$ , and it was assumed that  $C_{50} = 100$  (units are arbitrary) for every data point. Thus, in these simulations, interpatient variability was ignored by assuming that the data were derived from a single patient. Simulations were conducted for  $n = 10, 20, 30, 40, 50, 75,$  and  $100$  data points. Because any nonlinear regression technique necessitates the stipulation of initial parameter values (*i.e.*, an initial guess of what the true value is), we repeated each simulation for  $C_{50}$  initial = 40, 100 and 250, and  $\gamma$  initial = 0.6, 1.0, or 1.4 times the true value of  $\gamma$ . Each simulation was repeated 100 times, and the standard deviations of the estimates of  $C_{50}$  and  $\gamma$  were calculated. We did not calculate confidence intervals, but we determined, for each simulation, whether the confidence intervals would include the true values of  $C_{50}$  and  $\gamma$ . We did this by calculating the change in the logarithm of the likelihood, multiplied by a factor of 2 ( $2 \times \delta LL$ ), which occurred when the true value of either  $C_{50}$  or  $\gamma$  was substituted for the value determined by minimization of LL. This parameter ( $2 \times \delta LL$ ) has a chi-square distribution with one degree of freedom, which enables the researcher to determine whether any specific value of either  $C_{50}$  or  $\gamma$  is within the 95% confidence intervals.<sup>12</sup> It should be noted that the confidence intervals for this type of data often are markedly asymmetric, and the parameter estimates do not follow a t distribution.

In the second set of simulation experiments, we assumed that we pooled single data points from multiple patients and that the  $C_{50}$  and  $\gamma$  values had log-normal distributions (*i.e.*,  $P = P_{TV} \exp(\eta)$ , where  $P$  denotes the parameter ( $C_{50}, \gamma$ ),  $TV$  denotes the typical value, and  $\eta$  has a normal distribution with a mean value of zero and an SD of 0.3. We assumed that  $C_{50}TV = 100$ , and we considered  $\gamma_{TV}$  values of 1.0, 1.5, 3.0, 4.5, and 6.0. Simulations were again conducted for  $n = 10, 20, 30, 40, 50, 75,$  and  $100$  patients and for each value of simulations were repeated 100 times for starting values of  $C_{50}$  initial = 40, 100 and 250, and  $\gamma$  initial = 0.6, 1.0, or 1.4 times the typical value of  $\gamma$ . We also determined whether  $C_{50}TV$  and  $\gamma_{TV}$  were within the 95% confidence intervals, as described previously.

#### Statistical Analysis

Our results are presented as the coefficient of variation (*i.e.*, the SD of the parameter estimates normalized to the mean value of parameter estimates ( $C_{50}$  SD/mean or  $\gamma$  SD/mean) and estimate confidence, which we define as the percentage of parameter estimates for which the 95% confidence intervals include the “true” value (%CI<sub>95</sub>).

## Results

Bias (*i.e.*, the difference between the mean parameter estimate and the true value) is presented in table 1 for simulations in which the initial parameter values were the true values. The bias of  $C_{50}$  estimates generally was small ( $< 10\%$ ); unless  $n = 10$  and  $\gamma$  was small ( $\leq 2$ ), in which cases the bias was nearly 100%. The bias in the estimation of  $\gamma$  was less than 20% if the number of simulated data points ( $n$ ) was more than 10, with the notable exception of simulations in which the parameters had log-normal distributions (see Discussion).

Figures 2 and 3 present the coefficient of variation and estimate confidence of  $C_{50}$  and  $\gamma$  estimates as a function of sample size when we simulate data from a single patient (fixed  $C_{50}$  and  $\gamma$ ). In these figures, the initial “guesses” of the parameters are the true values. As expected, the coefficient of variation decreases as the number of patients ( $n$ ) increases. Additionally, the variability in the estimate of  $C_{50}$  increases as  $\gamma$  decreases (fig. 2). For  $\gamma = 1.0$ , the coefficient of variation of the  $C_{50}$  estimate is substantial ( $> 50\%$ ) unless the number of patients in the simulated study ( $n$ ) exceeds 30. However, if  $\gamma$  is larger ( $> 6$ ), the coefficient of variation of the  $C_{50}$

**Table 1. Mean  $C_{50}$  and  $\gamma$  Estimation**

n	Target $\gamma$				
	1.0	1.5	3.0	4.5	6.0
10					
$C_{50}$	119.5	119.0	114.4	107.3	102.9
$\gamma$	1.95	1.98	3.15	3.39	3.91
20					
$C_{50}$	116.0	112.7	103.8	98.0	102.5
$\gamma$	1.35	1.97	3.22	4.57	5.55
30					
$C_{50}$	107.8	115.0	105.2	99.0	102.4
$\gamma$	1.32	1.56	3.19	4.50	4.74
40					
$C_{50}$	110.0	105.9	104.3	98.9	102.2
$\gamma$	1.05	1.50	2.86	3.96	4.65
50					
$C_{50}$	110.2	111.6	102.4	96.7	102.8
$\gamma$	1.12	1.35	2.87	2.84	4.28
70					
$C_{50}$	95.3	107.2	103.9	99.6	102.4
$\gamma$	1.05	1.56	2.76	3.65	4.38
100					
$C_{50}$	105.7	102.1	99.8	100.5	101.2
$\gamma$	1.03	1.49	2.70	3.62	4.19

Target  $C_{50} = 100$ . For these simulations, the initial estimates of  $C_{50}$  and  $\gamma$  were equal to the target value.

estimate is approximately 10% if  $n \geq 20$ . Equally important, the true value of  $C_{50}$  lies within the 95% confidence intervals of the estimates in more than 90% of simulations. In contrast to  $C_{50}$  estimates, the magnitude of the true value of  $\gamma$  has less influence on the variability of the estimates of  $\gamma$ , as is shown in figure 3. The coefficient of variation of  $\gamma$  estimates is somewhat greater than that of  $C_{50}$  estimates and is approximately 60% for  $n = 20$ , decreasing only to approximately 40% for  $n = 100$ . However, the confidence of  $\gamma$  estimates is similar to that of  $C_{50}$  estimates, with %CI<sub>95</sub> values well in excess of 90% in these simulations with no interpatient variability.

Figures 4 and 5 present the coefficient of variation and estimate confidence for simulations in which  $C_{50}$  and  $\gamma$  are different for different simulated patients. We assumed that the parameters have log-normal distributions. These simulations mirror the common situation in which data from multiple patients are pooled for analysis. Again, we present simulations in which the initial guesses of  $C_{50}$  and  $\gamma$  are equal to  $C_{50TV}$  and  $\gamma_{TV}$ . As in the case with no interpatient variability, the coefficient of variation of estimates of either  $C_{50}$  or  $\gamma$  decrease as  $n$  increases. The magnitude of this variability is similar to the previous simulations, in which  $C_{50}$  and  $\gamma$  were fixed. As before, the coefficient of variation of  $C_{50}$  estimates decreases as  $\gamma$  increases. For  $C_{50}$ , the confidence of the

estimates are excellent, with the true value (100) within the 95% confidence intervals of the estimates in more than 90% of the simulations. However, it can be seen in figure 5 that the fraction of simulations in which the true value of  $\gamma$  ( $\gamma_{TV}$ ) was within the 95% confidence intervals of the estimate decreased to as low as 60% as  $n$  increased for values of  $\gamma$  more than or equal to 4.5.

Simulations were also conducted using starting values for  $C_{50}$  and  $\gamma$  other than the true values, which were used for the simulations illustrated in figures 2-5. The results were similar to those described previously, with the exception that, for  $n = 10$  or  $n = 20$ , use of an overestimate as the starting value of  $C_{50}$  resulted in a markedly larger coefficient of variation in the  $C_{50}$  estimate for small values of  $\gamma$  ( $\leq 3$ ). Despite this increase in variability, %CI<sub>95</sub> remained more than 90%.

## Discussion

Evaluation of drug concentration-response curves is an important component of pharmacologic research. It has become customary to describe concentration-response curves with sigmoid functions characterized by two parameters:  $C_{50}$ , the concentration of drug which results in 50% of maximal effect, and  $\gamma$ , a parameter that reflects the steepness of the curve. For many applica-

## STATISTICS OF LOGISTIC REGRESSION

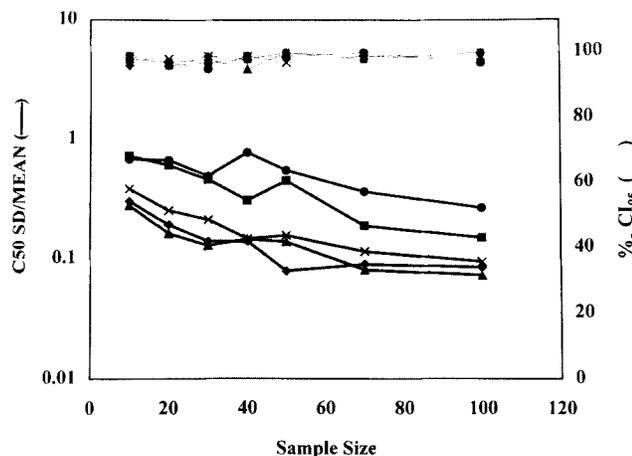


Fig. 2. (Lower) The coefficient of variation of  $C_{50}$  estimates. (Upper) The percentage of the simulations (as a function of the sample size  $n$ ) in which the 95% confidence intervals contained the true value of  $C_{50}$ . For these simulations, the initial values of  $C_{50}$  and steepness of the concentration-effect relation ( $\gamma$ ) were 100 and 1.0 times the true value, respectively, and it was assumed that data came from a single patient. The different curves present results for  $\gamma = 1.0$  (●-●), 1.5 (■-■), 3.0 (x-x), 4.5 (◆-◆), and 6.0 (▲-▲).

tions in anesthesiology research, the response is a binary all-or-none variable, and the maximal drug response or effect is unity. In this situation, estimation of  $C_{50}$  and  $\gamma$  is usually referred to as logistic regression. Techniques of logistic regression are based on the principle of maximum likelihood and are inherently nonlinear. Any

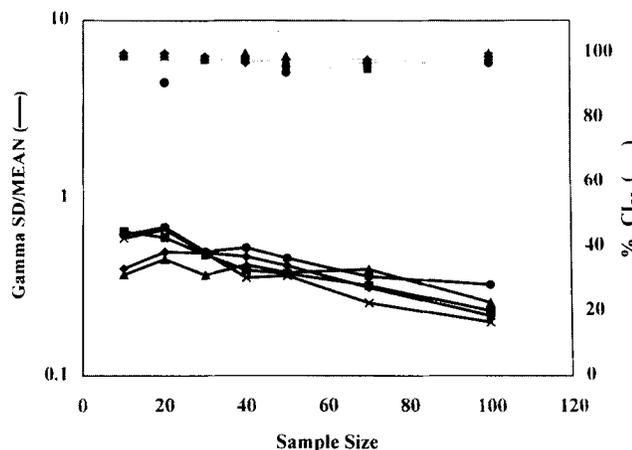


Fig. 3. (Lower) The coefficient of variation of  $\gamma$  estimates. (Upper) The percentage of simulations in which the 95% confidence intervals contained the true value of  $\gamma$ . For these simulations, the initial values of drug concentration associated with a 50% probability of drug effect ( $C_{50}$ ) and  $\gamma$  were 100 and 1.0 times the true value, respectively, and it was assumed that data came from a single patient. The different curves present results for  $\gamma = 1.0$  (●-●), 1.5 (■-■), 3.0 (x-x), 4.5 (◆-◆), and 6.0 (▲-▲).

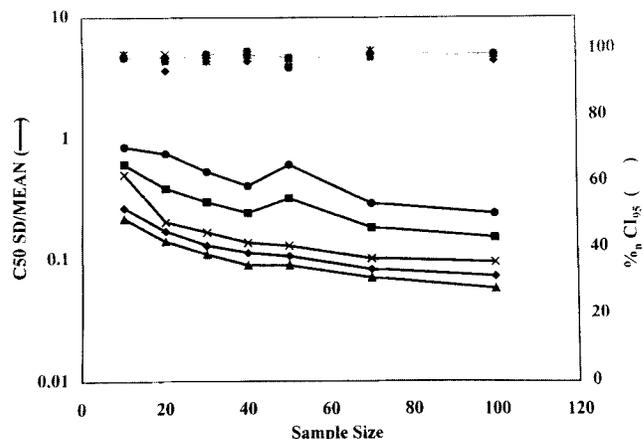


Fig. 4. (Lower) The coefficient of variation of  $C_{50}$  estimates. (Upper) The percentage of simulations in which the 95% confidence intervals contained the typical value of  $C_{50}$ . For these simulations, the initial values of  $C_{50}$  and steepness of the concentration-effect relation ( $\gamma$ ) were 100 and 1.0 times the typical value, respectively, and it was assumed that data from multiple patients was pooled and that  $C_{50}$  and  $\gamma$  had log-normal distributions, with SDs of  $\log C_{50}$  and  $\log \gamma$  equal to 30% of the typical values. The typical value of  $C_{50}$  was 100. The different curves present results for  $\gamma_{TV} = 1.0$  (●-●), 1.5 (■-■), 3.0 (x-x), 4.5 (◆-◆), and 6.0 (▲-▲).

method of statistical analysis should provide parameter estimates and confidence intervals. In contrast to linear regression, the confidence intervals provided by most software packages for nonlinear regression are approxi-

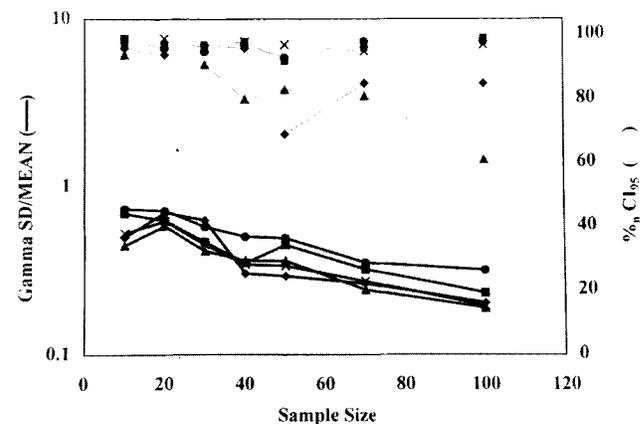


Fig. 5. (Lower) The coefficient of variation of  $\gamma$  estimates. (Upper) The percentage of simulations in which the 95% confidence intervals contained the typical value of  $\gamma$ . For these simulations, the initial values of drug concentration associated with a 50% probability of drug effect ( $C_{50}$ ) and  $\gamma$  were 100 and 1.0 times the typical value, respectively, and it was assumed that data from multiple patients was pooled and that  $C_{50}$  and  $\gamma$  had log-normal distributions with SDs of  $\log C_{50}$  and  $\log \gamma$  equal to 30% of the typical values. The typical value of  $C_{50}$  was 100. The different curves present results for  $\gamma_{TV} = 1.0$  (●-●), 1.5 (■-■), 3.0 (x-x), 4.5 (◆-◆), and 6.0 (▲-▲).

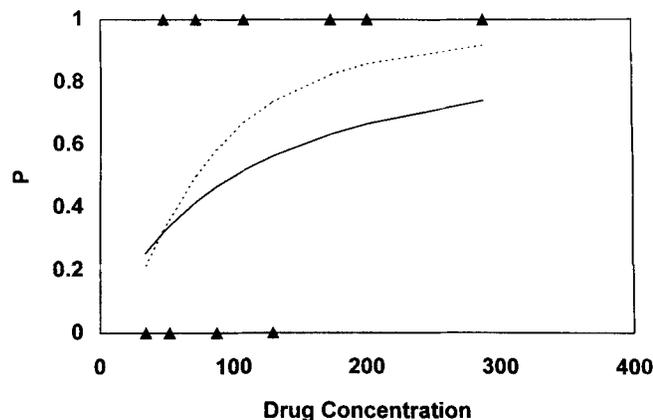


Fig. 6. A typical simulation for  $n = 10$  data points and steepness of the concentration–effect relation ( $\gamma = 1$ ). For this simulation, it was assumed that each data point came from the same patient and that drug concentration associated with a 50% probability of drug effect ( $C_{50} = 100$ ). Filled triangles indicate the simulated drug effect (*i.e.*, whether there was a positive drug effect or not) as a function of concentration. The solid line indicates the true concentration–effect relation. The dashed line shows the concentration–effect predicted by the maximum likelihood estimates of  $C_{50}$  and  $\gamma$ .

mations that are only accurate asymptotically as the number of data points increases toward infinity. Because many applications of logistic regression in anesthesiology research involve 20–50 data points, the accuracy of  $C_{50}$  and  $\gamma$  estimates are unclear. In this study, we used simulations (a well-known technique for analyzing statistical methodology, also known as Monte Carlo simulation) to investigate the relation between study population size and the accuracy of parameter estimates.

The primary measures we used to assess accuracy are the coefficient of variation and the fraction of simulations for which the 95% confidence intervals contained the true parameter value ( $\%CI_{95}$ ). The coefficient of variation is the SD of the parameter estimate expressed as a percentage of the mean estimate. We did not graphically present bias (*i.e.*, the difference between the mean estimate and the true value) because, in general, bias was small with the mean  $C_{50}$  estimate within less than 10% of the true value and the mean estimate of  $\gamma$  within less than 20% of the true value. However, for simulations of small ( $n = 10$ ) studies, we found a much larger bias for both parameters, although  $\%CI_{95}$  was still more than 90% when we simulated data from single patients. When we simulated pooled data from multiple patients (with log-normal distributions for  $C_{50}$  and  $\gamma$ ), there was a larger bias in  $\gamma$  estimates (up to 30%), even when  $n$  was large and  $\%CI_{95}$  was significantly smaller.

As expected, the variability of parameter estimates

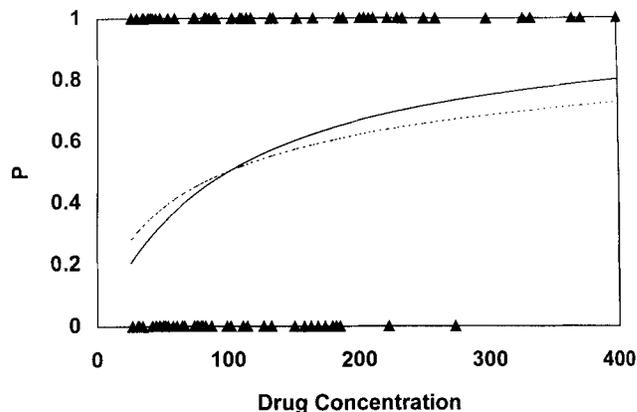


Fig. 7. A typical simulation for  $n = 100$  data points and steepness of the concentration–effect relation ( $\gamma = 1$ ). For this simulation, it was assumed that each data point came from the same patient and that drug concentration associated with a 50% probability of drug effect ( $C_{50} = 100$ ). The filled triangles indicate the simulated drug effect (*i.e.*, whether there was a positive drug effect) as a function of concentration. The solid line indicates the true concentration–effect relation. The dashed line shows the concentration–effect predicted by the maximum likelihood estimates of  $C_{50}$  and  $\gamma$ .

decreases (*i.e.*, the coefficient of variation decreases) as the number of patients in the simulated study ( $n$ ) increases. This is evident in figures 2–5. We did not anticipate that the variability of  $C_{50}$  estimation would depend on the value of  $\gamma$ , improving as  $\gamma$ , the measure of the steepness of the concentration–response curve, in-

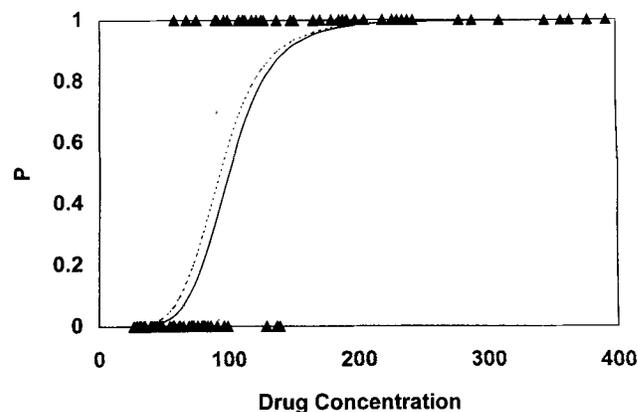


Fig. 8. A typical simulation for  $n = 100$  data points and steepness of the concentration–effect relation ( $\gamma = 6$ ). For this simulation, it was assumed that each data point came from the same patient and that drug concentration associated with a 50% probability of drug effect ( $C_{50} = 100$ ). The filled triangles indicate the simulated drug effect (*i.e.*, whether there was a positive drug effect) as a function of concentration. The solid line indicates the true concentration–effect relation. The dashed line shows the concentration–effect predicted by the maximum likelihood estimates of  $C_{50}$  and  $\gamma$ .

## STATISTICS OF LOGISTIC REGRESSION

creases. However, the basis for this observation is more evident with consideration of specific simulated data sets. Figure 6 illustrates a simulation of 10 data points (from the same patient, so that interpatient variability is not an issue) with  $\gamma = 1$ . One could surmise that there is little information about either  $C_{50}$  or  $\gamma$  in this data. Figure 7 presents a larger data set ( $n = 100$ ) for  $\gamma = 1$ . Comparing the two figures, one can see how  $C_{50}$  may be estimated more reliably when  $n$  is larger. Figure 8 presents simulated data for  $n = 100$  and  $\gamma = 6$ . It is now clear that  $C_{50}$  and  $\gamma$  both are estimated more reliably in this situation.

Statistical techniques should not only provide parameter estimates, but also measures of the accuracy of these estimates. The confidence intervals provided by typical statistical software programs for nonlinear techniques, such as logistic regression, are only valid asymptotically as the number of data points ( $n$ ) approaches infinity.<sup>12</sup> We investigated the accuracy of these asymptotic confidence intervals by determining the frequency with which the 95% confidence intervals contained the true value of the parameter. We used the change in log likelihood that occurred when the true parameter value was substituted for the maximum likelihood estimate as a means of assessing this frequency. When we simulated data from a single patient, we found that the 95% confidence intervals included the true parameter value in more than 90% of the simulations, even for small  $n$  ( $n = 10$ ) and small  $\gamma$ , when the coefficient of variation of parameter estimates was large. This suggests that, even when there is large estimate variability (*i.e.*, even when the point estimate of a parameter may be significantly in error), the 95% confidence intervals will be large enough to include the true value, if interpatient variation is not an issue. When we considered interpatient variation and simulated pooled data from multiple patients, we again found that the 95% confidence intervals included the true value of  $C_{50}$  in more than 90% of the simulations. Thus, analyzing data from a population of patients in a naive fashion does not appear to compromise the accuracy of  $C_{50}$  estimates. However, the 95% confidence intervals of  $\gamma$  estimates were not as reliable. For larger  $\gamma$  ( $\gamma \geq 4.5$ ), the fraction of simulations in which the 95% confidence intervals included the true value actually decreased to less than 90% as  $n$  increased. We believe this stems from the inability to distinguish between inpatient and interpatient variability when data from multiple patients are pooled for analysis in a naive fashion (*i.e.*, without explicitly accounting for interpatient variability). In one sense, the parameter  $\gamma$  is a measure of

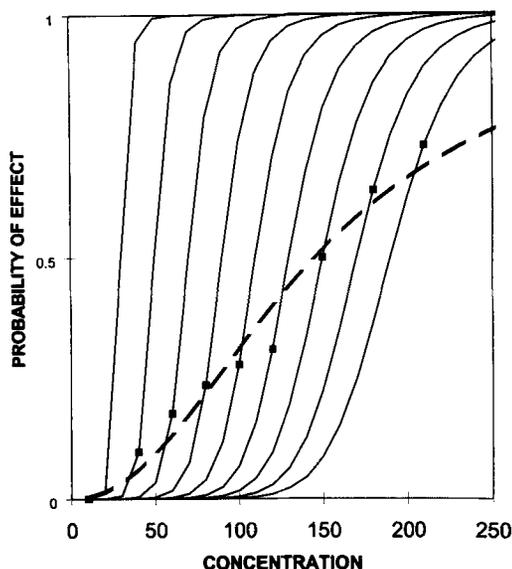


Fig. 9. A possible explanation of how the estimate of steepness of the concentration–effect relation ( $\gamma$ ) may be biased when data from multiple patients is pooled for analysis. In this example, single data points are taken from each of nine different patients, each of whom have a steep concentration–effect relation (—) but different values of drug concentration associated with a 50% probability of drug effect ( $C_{50}$ ). The resultant pooled concentration–effect curve (—■—) appears flat (*i.e.*, the apparent value of  $\gamma$  is lower than the true value).

inpatient variability. If  $\gamma$  is large, the concentration–response curve is steep. If the concentration of drug ( $C$ ) is slightly larger than  $C_{50}$ , the probability of drug response is close to one, and if  $C$  is slightly less than  $C_{50}$ , the probability of drug response is close to zero. The concentration range over which the probability of drug effect is intermediate, where there is significant inpatient variability, is narrow. In contrast, if  $\gamma$  is small, the concentration–response curve is relatively flat, and the probability of drug response takes on intermediate values over a wider concentration range (*i.e.*, the region of significant inpatient variability is larger). In figure 9, we illustrate how interpatient  $C_{50}$  variability may be indistinguishable from inpatient variability. This figure presents concentration–response curves for nine hypothetical patients, each with  $\gamma = 10$ , but with varying  $C_{50}$  values. If one data point is collected from each patient and pooled for analysis, the resulting curve (dashed line) appears to have a much lower value of  $\gamma$ . We believe this is the basis for the failure of the 95% confidence intervals to include the true value of  $\gamma$  when we simulated data from multiple patients with log-normal distributions for  $C_{50}$  and  $\gamma$ . This is consistent with our observation that the bias in  $\gamma$  estimation was increased in this case and

that the mean  $\gamma$  estimate was less than the true value. It is somewhat surprising that the confidence of the estimate of  $\gamma$  decreased with increasing  $n$ ; however, it should be noted that the confidence intervals are derived from the change in log likelihood, which results from substitution of other parameter values for the maximum likelihood estimates. The width of the confidence intervals decrease as  $n$  increases. Conversely, the bias in estimation of  $\gamma$  will not improve as  $n$  increases, because the steepness of the apparent concentration-effect curve will reflect the range of  $C_{50}$  values; this is evident from figure 9. Consequently, as the confidence intervals become more narrow, the chance that they will include the true value of  $\gamma$  decreases.

We used simulations to investigate the accuracy of analyses of concentration-binary response relations using logistic regression. The implications of these simulations for pharmacodynamic investigations are as follows:

1. The accuracy of parameter estimates increases as  $n$  and  $\gamma$  increase. The coefficient of variation of  $C_{50}$  in small studies ( $\leq 20$ ) may be as high as 30–40%. It should be emphasized that estimates of either  $C_{50}$  or  $\gamma$  are statistics. The distributions of these statistics are unknown; thus, the exact meaning of the coefficient of variation is unclear. Nevertheless, it seems reasonable to assume that point estimates of  $C_{50}$  may be in error by 30–40% in a large number of studies of this size. The coefficient of variation of  $\gamma$  is larger, indicating even greater error in point estimates of  $\gamma$ .
2. The 95% confidence intervals of  $C_{50}$  estimates contain the true value of  $C_{50}$  in most of the simulations ( $> 90\%$ ), even in simulations with larger coefficients of variation and even when we simulated naively pooling data from multiple patients. Confidence intervals for  $C_{50}$ , which are based on asymptotic statistical theory, appear to be reliable. Given the significant coefficient of variation that may occur in some studies, indicating the possibility of error in point estimates of  $C_{50}$ , it is clear that confidence intervals should always be reported.
3. When we simulated pooling of data from multiple patients, we found that the 95% confidence intervals

of the estimate of  $\gamma$  frequently did not include the true value of  $\gamma$ . The reason for this bias was discussed previously. Because of this, we believe that estimates of  $\gamma$  from studies in which data from multiple patients was naively pooled must be viewed with suspicion. In this type of analysis, inpatient variability (embodied in the parameter  $\gamma$ ) cannot be distinguished from interpatient variability. Accurate estimates of  $\gamma$  necessitate methods of analysis that take interpatient variability into account.

## References

1. Waud DR: On biological assays involving quantal responses. *J Pharmacol Exp Ther* 1972; 183:577–607
2. De Jong RE, Eger EI II: MAC expanded:  $AD_{50}$  and  $AD_{90}$  values of common inhalation anesthetics in man. *ANESTHESIOLOGY* 1975; 42:384–9
3. Fisher DM, Zwass MS: MAC desflurane in 60% nitrous oxide in infants and children. *ANESTHESIOLOGY* 1992; 76:354–6
4. Chortkoff BS, Bennett HL, Eger EI II: Subanesthetic concentrations of isoflurane suppress learning as defined by the category-example task. *ANESTHESIOLOGY* 1993; 79:16–22
5. Ausems ME, Hug CC Jr, DR, Stanski DR, Burm AGL: Plasma concentrations of alfentanil required to supplement nitrous oxide anesthesia for general surgery. *ANESTHESIOLOGY* 1986; 65:362–73
6. Wessen A, Persson PM, Nilsson A, Hartvig P: Concentration-effect relationship of propofol after total intravenous anesthesia. *Anesth Analg* 1993; 77:1000–7
7. Vuyk J, Lim T, Engbers FHM, Burm AGL, Vletter AA, Bovill JG: Pharmacodynamics of alfentanil as a supplement to propofol or nitrous oxide for lower abdominal surgery in female patients. *ANESTHESIOLOGY* 1993; 78:1036–45
8. Glass PSA, Doherty M, Jacobs JR, Goodman D, Smith LR: Plasma concentration of fentanyl, with 70% nitrous oxide, to prevent movement at skin incision. *ANESTHESIOLOGY* 1993; 78:842–7
9. Vuyk J, Lim T, Engbers FHM, Burm AGL, Vletter AA, Bovill JG: The pharmacodynamic interaction of propofol and alfentanil during lower abdominal surgery in women. *ANESTHESIOLOGY* 1995; 83:8–22
10. Jacobs JR, Reves JG, Marty J, White WD, Bai SA, Smith LR: Aging increases pharmacodynamic sensitivity to the hypnotic effects of midazolam. *Anesth Analg* 1995; 80:143–8
11. Bailey JM, Schwiager IM, Hug CC Jr: Evaluation of sufentanil anesthesia obtained by a computer-controlled infusion for cardiac surgery. *Anesth Analg* 1993; 76:247–52
12. Cox DR, Snell EJ: *Analysis of Binary Data*, 2nd edition. London, Chapman and Hall, 1988