Estimation and Confidence Regions for Multi-Dimensional Effective Dose

Jialiang Li[*,1], **Erik V. Nordheim**[2], **Chunming Zhang**[2], and **Charles E. Lehner**[3]

[1] Department of Statistics and Applied Probability, National University of Singapore, 117546, Singapore
[2] Department of Statistics, University of Wisconsin, Madison, WI 53706, USA
[3] Department of Surgical Sciences, University of Wisconsin, Madison, WI 53706, USA

Received 8 March 2007, revised 8 May 2007, accepted 13 June 2007

Summary

The problem of finding confidence regions for multiple predictor variables corresponding to given expected values of a response variable has not been adequately resolved. Motivated by an example from a study on hyperbaric exposure using a logistic regression model, we develop a conceptual framework for the estimation of the multi-dimensional effective dose for binary outcomes. The k-dimensional effective dose can be determined by conditioning on $k-1$ components and solving for the last component as a conditional univariate effective dose. We consider various approaches for calculating confidence regions for the multi-dimensional effective dose and compare them via a simulation study for a range of possible designs. We analyze data related to decompression sickness to illustrate our procedure. Our results provide a practical approach to finding confidence regions for predictor variables for a given response value.

Key words: Binary logistic regression; Decompression sickness; Effective dose; Inverse inference; Simultaneous confidence regions.

1 Introduction

Hyperbaric exposure, such as encountered in diving to depths and returning to the surface, can cause medical difficulties resulting in decompression sickness (DCS) and even death. Data collected on humans as well as on animals used as test organisms for humans are studied to relate the risk of dive factors to the onset of death or other (less severe) responses. The study motivating our work was a large study conducted on sheep that were dived and returned to surface pressure in chambers allowing for environmental control. 1108 dives were conducted at a range of depths and durations at depth. Data on a number of outcomes were recorded (e.g. bends, DCS, death) for each dive.

An important objective for diving researchers is to determine the range of depths and dive durations that correspond to a certain probability of a given outcome; for example, to find the range corresponding to a probability of 0.05 of death. The goal of our current work is to provide statistical methodology for finding such ranges and creating confidence regions for them.

This problem, when only one covariate is present, is usually called the estimation of effective dose (ED100p) in quantal bioassay research (Finney, 1978). We adopt this terminology and extend it to the DCS problem as the estimation of the "2-dimensional effective dose". This is straightforwardly generalized to multi-dimensional effective dose; our presentation considers this general case.

[*] Corresponding author: e-mail: stalj@nus.edu.sg, Phone: +65 6516 8932, Fax: +65 6872 3919

2 Statistical Setting

In a biological assay study, the response Y is a binary variable which represents absence or presence, often denoted "0" or "1", respectively of a certain clearly defined outcome (for example, death). The relationship between response and covariates is often described by the following logistic regression model:

$$\log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k} \beta_i X_i \,, \tag{1}$$

where $p = E(Y \mid X_1, \cdots, X_k)$ is the probability that the binary response equals "1". The number of predictors k is fixed for model (1) and the coefficients β_i $(i = 1, \cdots, k)$ are all assumed to be non-zero. When $k = 1$ and $X = X_1$ represents the dose level of a drug or treatment, the effective dose ED_{100p} is the value of $X = X_1$ that causes an outcome "1" with a given probability p $(0 < p < 1)$.

Most previous work on the estimation of the effective dose has focused on a single predictor variable, i.e., $k = 1$. Carter et al. (1986) describe a method to estimate an asymptotic confidence region about the ED_{100p} from the logistic curve with multiple explanatory variables, but the statistical properties of the method have not been carefully evaluated nor alternatives considered. In the following, we provide a general formulation for this problem.

For a given p in [0, 1], we define the multi-dimensional effective dose as the following set:

$$\Theta_p = \left\{ (x_1, x_2 \cdots, x_k) \in \mathbb{R}^k : \log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \right\}. \tag{2}$$

The elements of the set must satisfy $\log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i$. Thus, if we can determine the values of $\beta_i (i = 0, 1, \cdots, k)$, we can then obtain all the values contained in the set Θ_p. For example, when $k = 2$, $p = 0.5$ (expected value of Y) and $\beta_0 = \beta_1 = \beta_2 = 1$, $\Theta_{0.5}$ is the set of (x_1, x_2) such that $x_1 + x_2 = -1$.

Our first goal is to determine Θ_p. All methods with which we are familiar are logically equivalent to the procedure in which possible values for $k-1$ of the predictor variables are fixed and the solution obtained for the remaining variable. If a wide range of values for the $k-1$ predictor variables is used, this is (effectively) the unique approach to determine the points in the set Θ_p with β_i $(i = 0, 1, \cdots, k)$ known. Suppose $(x_2^*, \cdots, x_k^*)$ is one possible value for $(x_2, \cdots, x_k)$, then we consider the following $(x_2^*, \cdots, x_k^*)$-Conditioning Effective Dose:

$$\Theta_p^*(x_2^*, \cdots, x_k^*) = \left(\log \frac{p}{1-p} - \beta_0 - \sum_{i=2}^{k} \beta_i x_i^* \right) \Big/ \beta_1 \quad (\beta_1 \neq 0) \,. \tag{3}$$

We note that $\bigcup_* \{(x_1, x_2^*, \cdots, x_k^*) : x_1 = \Theta_p^*(x_2^*, \cdots, x_k^*)\} = \Theta_p$, where $\bigcup_*$ means a union across all possible values of $(x_2^*, \cdots, x_k^*)$. ${\Theta_p}^*(x_2^*, \cdots, x_k^*)$ is more tractable than Θ_p (particularly for finding confidence regions) and is employed to determine Θ_p indirectly. When conditioning variables are clear from the context, we shall suppress the conditioning values and write Θ_p^*.

Since $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$ is usually unknown in model (1) with k fixed covariates, parametric estimators must be obtained through a model-fitting procedure based on a training sample of size n. An estimator of Θ_p can be obtained by substituting a consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in (2).

The next objective is to construct a confidence region such that we will have $100(1-\alpha)\%$ probability that the true set Θ_p is contained in the confidence region. We accomplish this by finding a $100(1-\alpha)\%$ confidence interval I^* for each $(x_2^*, \cdots, x_k^*)$-conditioning effective dose Θ_p^*. We explore several methods for calculating I^* and then join these confidence intervals to form a confidence region

$$G = \bigcup_* \{(x_1, x_2^*, \cdots, x_k^*) : x_1 \in I^*\} \,. \tag{4}$$

In dose-time-response models, Chen (2007) has used a method to construct confidence bands for ED50 by conditioning on the time variable (or for lethal time, LT50, by conditioning on the dose variable). Chen (2007) considered a specific solution similar to Method 3 in our paper to find the 2-dimensional ED_{100p}; However, our study provides a more general framework to deal with the multiple-covariate calibration problem.

3 Methods

3.1 Construction of confidence intervals

To illustrate the approach, we shall assume $k = 2$; the results generalize straightforwardly to higher dimensions. By definition in (3), the x_2^*-conditioning effective dose is

$$\Theta_p^*(x_2^*) = \left(\log \frac{p}{1-p} - \beta_0 - \beta_2 x_2^*\right) \Big/ \beta_1 \,. \tag{5}$$

This is the value of X_1 such that a subject randomly selected from the conditional population given $X_2 = x_2^*$ responds with probability p.

To estimate $\Theta_p^*(x_2^*)$, we can substitute the logistic regression coefficient estimates for $\beta_j (j = 0, 1, 2)$ and obtain the point estimate

$$\hat{\Theta}_p^*(x_2^*) = \left(\log \frac{p}{1-p} - \hat{\beta}_0 - \hat{\beta}_2 x_2^*\right) \Big/ \hat{\beta}_1 \,. \tag{6}$$

It is well known that under regularity conditions (Shao, 1999, p. 246) maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ are consistent and asymptotically normal,

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d N(\mathbf{0}, V) \,. \tag{7}$$

The asymptotic covariance matrix V/n is the inverse of the information matrix of the log-likelihood. The inverse of the sample information matrix is a consistent estimator $\hat{V}/n$. Most standard statistical packages for logistic regression produce $\hat{\boldsymbol{\beta}}$ and $\hat{V}/n$ in their outputs. In the following subsections we discuss four methods for constructing the confidence interval I^* for Θ_p^*.

3.1.1 Method 1

We first consider a method based on inverting Scheffé's simultaneous confidence interval (Scheffé, 1959, p. 30) for $\eta = \log \frac{p}{1-p} = \boldsymbol{x}^T\boldsymbol{\beta}$ for any $\boldsymbol{x} = (1, x_1, x_2)^T$. Based on the limit distribution of $\hat{\boldsymbol{\beta}}$ and the Cauchy–Schwarz inequality, such an asymptotic confidence interval is given by

$$\boldsymbol{x}^T\hat{\boldsymbol{\beta}} - \sqrt{\frac{\boldsymbol{x}^T\hat{V}\boldsymbol{x}\chi_3^2(\alpha)}{n}} \le \eta \le \boldsymbol{x}^T\hat{\boldsymbol{\beta}} + \sqrt{\frac{\boldsymbol{x}^T\hat{V}\boldsymbol{x}\chi_3^2(\alpha)}{n}}, \tag{8}$$

where $\chi_3^2(\alpha)$ is the upper α quantile of the χ^2 distribution with 3 degrees of freedom. The confidence region for the x_2^*-conditioning effective dose Θ_p^* at a given x_2^* is then

$$I_1^* = \left\{ x_1 : \boldsymbol{c}^T\hat{\boldsymbol{\beta}} - \sqrt{\frac{\boldsymbol{c}^T\hat{V}\boldsymbol{c}\chi_3^2(\alpha)}{n}} \le \log \frac{p}{1-p} \le \boldsymbol{c}^T\hat{\boldsymbol{\beta}} + \sqrt{\frac{\boldsymbol{c}^T\hat{V}\boldsymbol{c}\chi_3^2(\alpha)}{n}}, \text{ where } \boldsymbol{c} = (1, x_1, {x_2}^*)^T \right\} \tag{9}$$

This is an asymptotic $100(1-\alpha)\%$ simultaneous confidence interval for Θ_p^* at any x_2^* since

$$\begin{aligned} P(\Theta_p^* \in I_1^*, \text{ for any } x_2^*) &= P\Bigg(\boldsymbol{c}^T\hat{\boldsymbol{\beta}} - \sqrt{\frac{\boldsymbol{c}^T\hat{\boldsymbol{V}}\boldsymbol{c}\chi_3^2(\alpha)}{n}} \le \log\frac{p}{1-p} \\ &\qquad \le \boldsymbol{c}^T\hat{\boldsymbol{\beta}} + \sqrt{\frac{\boldsymbol{c}^T\hat{\boldsymbol{V}}\boldsymbol{c}\chi_3^2(\alpha)}{n}}, \text{ where } \boldsymbol{c} = (1, x_1, x_2^*)^T \text{ for any } x_2^*\Bigg) \\ &\ge P\Bigg(\boldsymbol{x}^T\hat{\boldsymbol{\beta}} - \sqrt{\frac{\boldsymbol{x}^T\hat{\boldsymbol{V}}\boldsymbol{x}\chi_3^2(\alpha)}{n}} \le \log\frac{p}{1-p} \\ &\qquad \le \boldsymbol{x}^T\hat{\boldsymbol{\beta}} + \sqrt{\frac{\boldsymbol{x}^T\hat{\boldsymbol{V}}\boldsymbol{x}\chi_3^2(\alpha)}{n}}, \text{ for any } \boldsymbol{x}\Bigg) \\ &= 1-\alpha\,. \end{aligned} \tag{10}$$

This method has been known to be conservative (e.g. Hsu, 1996) since Scheffé's result provides simultaneous confidence intervals for all possible linear contrasts of regression parameters. We also note that the usual F quantiles have been replaced by χ^2 quantiles because the normal distribution for the estimated regression coefficient is assymptotically correct (in which case the denominator degrees of freedom for the F-statistic go to infinity).

Our method is similar to Carter et al. (1986) in the sense that our constructions both follow from Scheffé's simultaneous confidence intervals. It is relatively easy to implement our method in practice since we work on one dimension after conditioning while Carter et al. (1986) only provided solutions in complicated multi-dimensional forms.

3.1.2 Method 2

We apply the delta method to obtain the limit distribution of the estimator $\hat{\Theta}_p^*$ and construct the confidence interval according to this distribution.

We define the function $g(t_1,t_2,t_3) = (\eta - t_1 - t_3x_2^*)/t_2$. Then, $\Theta_p^* = g(\beta_0,\beta_1,\beta_2)$ and $\hat{\Theta}_p^* = g(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2)$. Using the delta method and (7) gives

$$\sqrt{n}(\hat{\Theta}_p^* - \Theta_p^*) \to_d N(0, V_S)\,, \tag{11}$$

where $V_S = l^T V l$. The column vector $l = \nabla g(\beta_0,\beta_1,\beta_2)$, where

$$\nabla g(t_1,t_2,t_3) = \left(\frac{-1}{t_2}, \frac{\eta - t_1 - t_3x_2^*}{-t_2^2}, \frac{-x_2^*}{t_2}\right)^T, \tag{12}$$

is the gradient of $g(t_1,t_2,t_3)$. A consistent estimator for V_S/n is $\hat{l}^T\hat{V}\hat{l}/n$, where $\hat{l} = \nabla g(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2)$.

The asymptotic $100(1-\alpha)\%$ confidence interval can be constructed as

$$I_2^* = \left\{x_1 : \hat{\Theta}_p^* - z_{\alpha/2}\sqrt{\hat{V}_S/n} \le x_1 \le \hat{\Theta}_p^* + z_{\alpha/2}\sqrt{\hat{V}_S/n}\right\}, \tag{13}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

3.1.3 Method 3

Fieller's Theorem (Fieller, 1954) has been employed to compute confidence intervals for the ratios of linear combinations of the regression parameters. We extend this idea to the construction of confi-

dence intervals for Θ_p^*. Let

$$T = \frac{\eta - \hat{\beta}_0 - \hat{\beta}_2 x_2^* - \hat{\beta}_1 x_1}{\sqrt{\boldsymbol{c}^T \hat{\boldsymbol{V}} \boldsymbol{c}/n}}\,, \tag{14}$$

where $\boldsymbol{c} = (1, x_1, x_2^*)^T$. Given η, x_2^*, and letting x_1 be the corresponding Θ_p^* of x_2^*, T is asymptotically distributed as a standard normal variable. Thus,

$$1 - \alpha = P(T^2 \le z_{\alpha/2}^2)\,. \tag{15}$$

The inequality $T^2 \le z_{\alpha/2}^2$ can be rearranged and expressed as a quadratic inequality in x_1

$$Ax_1^2 + Bx_1 + C \le 0\,, \tag{16}$$

where

$$\begin{aligned} A &= \hat{\beta}_1^2 - z_{\alpha/2}^2 v_{22} \\ B &= 2[\hat{\beta}_1(\hat{\beta}_0 + \hat{\beta}_2 x_2^* - \eta) - z_{\alpha/2}^2(v_{21} + v_{23} x_2^*)] \\ C &= (\hat{\beta}_0 + \hat{\beta}_2 x_2^* - \eta)^2 - z_{\alpha/2}^2(v_{11} + 2v_{13} x_2^* + v_{33} x_2^{*2}) \end{aligned}$$

and v_{ij} is the (i, j)-th element of the matrix $\hat{V}/n$.

A $100(1-\alpha)\%$ asymptotic confidence interval for Θ_p^* is then

$$I_3^* = \left\{ x_1 : \frac{-B - \sqrt{B^2 - 4AC}}{2A} \le x_1 \le \frac{-B + \sqrt{B^2 - 4AC}}{2A} \right\}, \tag{17}$$

provided $A > 0$ and $B^2 - 4AC > 0$. Method 3 is equivalent to the inversion of the pointwise confidence interval for $\eta(\boldsymbol{x})$.

3.1.4 Method 4

Another method for the construction of the confidence interval is the bootstrap re-sampling method. We generate B bootstrap samples $\{\Omega_b, b = 1, 2, \cdots, B\}$, each consisting of n observations $\Omega_b = \{(X_{1i,b}, X_{2i,b}, Y_{i,b}), i = 1, 2, \cdots, n\}$ which are randomly drawn with replacement from the original sample $\{(X_{1i}, X_{2i}, Y_i), i = 1, 2, \cdots, n\}$. We refit the logistic model (1) for each of the bootstrap samples and evaluate the x_2^*-conditioning effective dose

$$\hat{x}_{1,b} = \left(\log \frac{p}{1-p} - \hat{\beta}_{0,b} - \hat{\beta}_{2,b} x_2^* \right) \Big/ \hat{\beta}_{1,b}, \quad b = 1, 2, \cdots, B \tag{18}$$

where $\hat{\beta}_{i,b}, (i = 1, 2, 3)$ are the coefficient estimates for the b-th bootstrap sample.

We then choose the $\alpha/2$ quantile $\hat{x}_{1,([B\alpha/2])}$ and $1 - \alpha/2$ quantile $\hat{x}_{1,([B(1-\alpha/2)])}$ from the ordered set $\{\hat{x}_{1,(b)} : b = 1, 2, \cdots, B\}$, where $[u]$ is the largest integer smaller than u, and use these two values as the lower and upper limits of the $1 - \alpha$ confidence interval for Θ_p^*, respectively. The estimated interval can be expressed as

$$I_4^* = \{x_1 : \hat{x}_{1,([B\alpha/2])} \le x_1 \le \hat{x}_{1,([B(1-\alpha/2)])}\}\,. \tag{19}$$

The closeness of the coverage rate of this interval to the desired confidence level depends on the sample size n as well as the number of bootstrap samples, B. Efron and Tibshirani (1993, p. 52) suggest $B > 250$ be considered for estimating the standard error in practice; Davis and Hinkley (1997, p. 21) indicates that at least 1000 bootstrap samples should be generated for density estimation. We note that an alternative to the method we described here is the parametric bootstrap (Efron and Tibshirani, 1993, p. 53).

3.2 Properties of confidence regions

3.2.1 Simultaneous vs. pointwise confidence regions

In the general case we form the confidence region G in (4) by combining the confidence intervals I^* of Θ_p^* for all possible $(x_2^*, \cdots, x_k^*)$. The resulting confidence coefficient of the region G for capturing the true Θ_p is

$$\begin{aligned} P(\Theta_p \in G) &= P\Big[\bigcup_{*}\{(x_1, x_2^*, \cdots, x_k^*) : x_1 \in \Theta_p^*\} \\ &\qquad \subset \bigcup_{*}\{(x_1, x_2^*, \cdots, x_k^*) : x_1 \in I^*\}\Big] \\ &\geq P\{\Theta_p^*(x_2^*) \in I^*(x_2^*), \text{ for all } {x_2}^*\}. \end{aligned}$$

This probability argument indicates that in order to make the coverage probability for Θ_p no less than $1-\alpha$, it is necessary that the simultaneous coverage probability for all Θ_p^* be $1-\alpha$. Only Method 1 can ensure this simultaneous coverage probability for Θ_p^* and therefore achieve the correct coverage probability for Θ_p as is demonstrated in (10).

The choice of simultaneous or pointwise methods will depend on the application. For some applications there may be primary interest in one variable conditional on others. In other applications the focus may be placed on a small subregion. In such circumstances a pointwise confidence interval I^* may be desired; Methods 2, 3 and 4 are all possibilities.

We note that the resulting confidence regions can be different for Methods 2 and 4 depending on the choice of the conditioning variable. Such a conditioning dependent difference diminishes as the sample size n becomes large. Asymptotically, the regions obtained by Methods 2 or 4 approach the region obtained by Method 3 (Cox, 1990), which is not affected by the choice of the conditioning variable. Like Method 3, Method 1 does not depend on the conditioning variable choice.

3.2.2 Lengths of confidence intervals

For the four methods, we can easily show that Method 1 should always result in the widest interval. The other three methods should have similar lengths asymptotically. The exact differences among methods depend on the distribution of the covariates as well as on the pre-specified probability of occurrence of response.

4 Simulation Study

4.1 Models with two covariates

We conducted a numerical study with 1000 simulations. For each simulation, we generated $n = 36$ or 360 samples of X_1 and X_2 from the 6 designs depicted in Figure 1. These designs were selected to represent a balanced situation and various patterns of imbalance. The number shown in the plot is the number of replicates at that point for simulations with sample size 36. For $n = 360$, the number of replicates at each point is increased by a factor of 10. The true parameters were fixed to be $\beta_0 = -6$, $\beta_1 = 6$, $\beta_2 = 6$. This choice of betas was made so that the "probability of success" would be 0.5 for the line $X_1 + X_2 = 1.0$ and $\beta_0 + \beta_1 X_1 + \beta_2 X_2$ would be -6 and $+6$, respectively, at the lower left and upper right corners of the unit square.

The binary response was generated from the Bernoulli distribution with success probability $p = \frac{1}{1+e^{-(-6+6X_1+6X_2)}}$. The contours projected from the likelihood surface (at different p's) are straight lines with identical slopes but different intercepts. Each straight line represents the true 2-dimensional effective dose corresponding to p. For each of the 1000 simulations, a logistic regression model was fitted to the generated data.

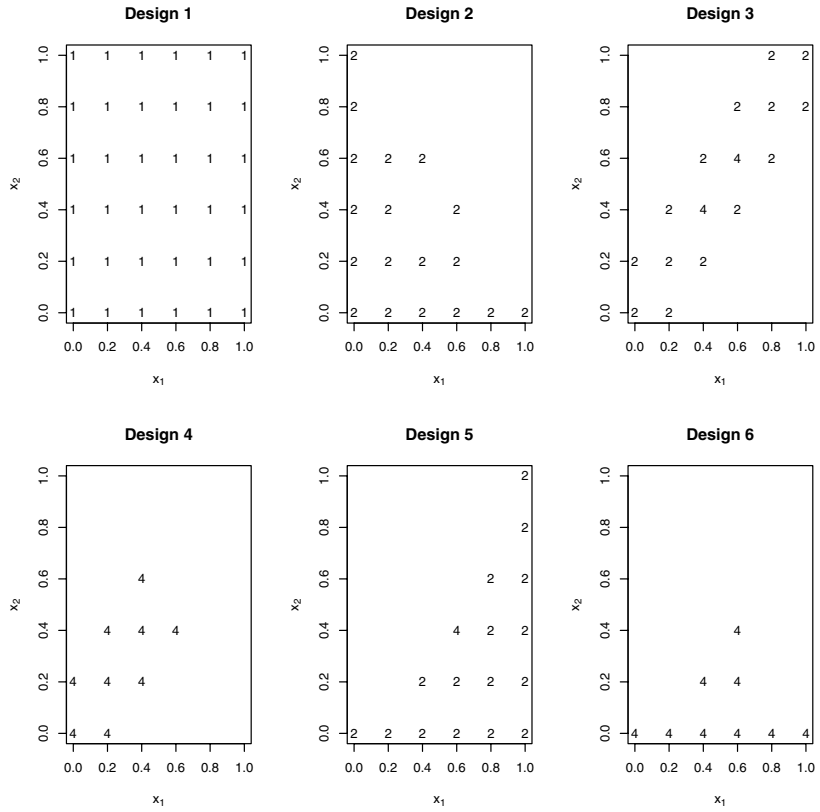

Figure 1 The Distribution of X_1 and X_2 used in the Simulations.

We chose 100 equally spaced values of X_2 between [0, 1] and for each, calculated the x_2^*-conditioning effective dose $\hat{\Theta}_p^*$ for $p = 0.01, 0.05, 0.1, 0.5, 0.9, 0.95,$ and 0.99 and the associated 95% confidence intervals I^* by using the four methods described in Section 3.1.

For Method 1, we report the simultaneous coverage rate "R1" and average length "L". Only if all 100 intervals covered all 100 true $\Theta_p^*(x_2^*)$ in one simulation do we report a correct *simultaneous* coverage. For the other three methods, we report the pointwise coverage rate "R2" and average length "L".

We note that the logistic regression model can result in computational instability when the data are in a state of "complete separation" (Albert and Anderson, 1984). This phenomenon can occur more frequently for some types of designs when sample sizes are small. Consequently confidence intervals are impossible to obtain for some cases with sample size 36 (only Designs 1 and 5 are completely reported in this paper).

In Table 1, the realized simultaneous coverage rates for Method 1 are all above 95%, often substantially so. This wide coverage might be partially due to the fact that the asymptotic distribution has 3 degrees of freedom but there are only 2 degrees of freedom in the linear contrast since the intercept is constant. In addition, our simulations only covered the unit square; the coverage rates almost certainly would be smaller if we examined a larger region. In summary, Method 1 is the method of choice for simultaneous inference although the results are often overly conservative.

With $n = 360$ Methods 2, 3, and 4 all result in coverage rates close to 95% although those for Method 2 are sometimes a bit low. Overall, Method 3 appears the most stable. By looking at the proportions of intervals above and below the true value (not reported here), we note that Method 2 demonstrates a bias.

Table 1 Coverage rates for 95% confidence regions based on four methods from 1000 simulations.

			Method 1[a]		Method 2[b]		Method 3[b]		Method 4[b]	
n	p	Design	R1	L	R2	L	R2	L	R2	L
36	.01	1	.996	8.6	.882	1.25	.983	28.59	.945	78.3
	.05	1	.996	76.9	.888	.88	.976	2.48	.948	10.5
	.1	1	.988	71.8	.902	.73	.980	13.96	.932	6.25
	.5	1	1.00	11.5	.948	.53	.980	14.19	.951	1.42
	.9	1	.992	9.9	.904	.73	.984	1.67	.955	2.56
	.95	1	1.00	96.4	.881	.88	.985	2.37	.929	6.44
	.99	1	.991	75.4	.875	1.46	.976	3.98	.910	4.50
	.01	5	1	11.29	.899	1.51	.997	19.39	.952	2.18
	.05	5	1	148.61	.902	1.17	.996	12.41	.948	10.56
	.1	5	1	65.26	.906	.99	.993	14.61	.954	13.75
	.5	5	1	14.62	.935	.69	.982	6.97	.950	1.48
	.9	5	.978	38.43	.932	.76	.976	8.59	.955	2.58
	.95	5	.964	52.54	.926	.87	.981	6.33	.964	2.37
	.99	5	.956	97.38	.923	1.14	.991	1.16	.933	4.04
360	.01	1	.994	.586	.951	.37	.957	.39	.945	.46
	.05	1	.993	.408	.947	.26	.951	.27	.940	.29
	.1	1	.985	.337	.942	.21	.955	.23	.954	.25
	.5	1	.983	.214	.950	.14	.954	.14	.949	.15
	.9	1	.990	.335	.954	.22	.960	.22	.950	.25
	.95	1	.993	.410	.947	.26	.947	.27	.946	.28
	.99	1	.988	.586	.945	.37	.953	.38	.952	.36
	.01	2	.990	.68	.945	.39	.949	.43	.947	.46
	.05	2	.987	.42	.945	.25	.956	.27	.952	.25
	.1	2	.986	.31	.938	.19	.946	.20	.935	.18
	.5	2	.989	.27	.949	.16	.943	.17	.955	.14
	.9	2	.996	.58	.949	.34	.947	.37	.944	.30
	.95	2	.988	.69	.939	.41	.942	.44	.947	.52
	.99	2	.988	.98	.934	.56	.950	.62	.946	.65
	.01	3	.991	4.35	.933	.85	.956	1.22	.941	1.33
	.05	3	.985	2.65	.934	.66	.962	.93	.946	.99
	.1	3	.987	2.04	.946	.60	.959	.85	.945	.80
	.5	3	.991	2.47	.941	.49	.953	1.05	.958	.65
	.9	3	.983	2.21	.935	.59	.953	.79	.955	.89
	.95	3	.982	2.51	.932	.66	.953	.85	.949	.96
	.99	3	.989	3.39	.928	.86	.951	1.34	.940	1.09
	.01	4	.983	2.14	.933	.71	.946	.93	.942	1.05
	.05	4	.985	1.44	.946	.52	.953	.65	.943	.68
	.1	4	.983	1.27	.941	.46	.946	.61	.935	.57
	.5	4	.984	1.15	.949	.41	.949	.61	.957	.43
	.9	4	.984	1.48	.945	.54	.940	.71	.952	.66
	.95	4	.99	1.80	.938	.63	.945	.96	.939	.83
	.99	4	.992	2.27	.927	.84	.951	1.14	.932	1.00

Table 1 Continued.

			Method 1[a]		Method 2[b]		Method 3[b]		Method 4[b]	
n	p	Design	R1	L	R2	L	R2	L	R2	L
	.01	5	.992	1.03	.939	.60	.958	.66	.939	.60
	.05	5	.987	.79	.947	.46	.945	.49	.935	.57
	.1	5	.986	.68	.929	.40	.949	.43	.941	.44
	.5	5	.983	.43	.966	.27	.947	.29	.952	.25
	.9	5	.980	.47	.938	.28	.944	.30	.935	.33
	.95	5	.989	.52	.936	.32	.955	.34	.950	.40
	.99	5	.989	.70	.938	.42	.959	.45	.929	.52
	.01	6	.989	.92	.949	.54	.963	.59	.938	.51
	.05	6	.985	.67	.938	.39	.954	.43	.963	.49
	.1	6	.985	.57	.946	.34	.951	.37	.949	.28
	.5	6	.987	.46	.956	.28	.957	.30	.930	.25
	.9	6	.992	.66	.935	.39	.957	.42	.956	.49
	.95	6	.990	.75	.942	.45	.955	.47	.936	.50
	.99	6	.996	1.01	.937	.58	.947	.64	.937	.54

[a] Simultaneous confidence interval.
[b] Pointwise confidence interval.

The lengths of confidence intervals are computed in the same scale as the covariates in the simulation. The average length of confidence intervals from Method 1 is much greater than those from the other methods. The other three methods are quite similar to each other. Method 3 always yields slightly wider confidence intervals than Method 2 with Method 4 usually between the other two.

4.2 Models with four covariates

We also conducted a simulation for a logistic regression model with four covariates. The distribution of the four covariates used is the discrete uniform distribution on $4^4 = 256$ equally spaced grid points in the hypercube $[0, 1]^4$ with one replicate per point. The binary response was generated from the Bernoulli distribution with success probability $p = \frac{1}{1+e^{-(-6+3X_1+3X_2+3X_3+3X_4)}}$. Logistic regression models were fitted to the generated data.

The values for (X_2, X_3, X_4) were fixed at the designed grid points and the (x_2^*, x_3^*, x_4^*)-conditioning effective dose $\hat{\Theta}_p^*$ for $p = .01, .05, .1, .3, .5, .7, .9, .95$, and .99 and the associated 95% confidence intervals I^* were calculated. Methods 2 and 4 compare with Method 3 in a manner similar to that found for the two-covariate case, and hence are not reported here. Table 2 provides the results from the 1000 simulations for the two methods. Again, Method 1 produces confidence regions with simultaneous coverage probabilities of at least 95% and Method 3 appears to perform well on a point-wise basis.

5 Example

It is well known that divers returning to the water surface from depths can encounter medical difficulties. After hyperbaric exposures, it is best for divers to return to normal pressure using a carefully calculated protocol. This process, called decompression, allows gradual, safe elimination of inert gases

Table 2 Coverages rates for 95% confidence regions based on Methods 1 and 3 from 1000 simulations for four covariates.

p	Computing Methods			
	Method 1[a]		Method 3[b]	
	R1	L	R2	L
.01	.992	3.33	.954	1.35
.05	.994	2.23	.948	.96
.1	.985	1.86	.949	.82
.3	.989	1.41	.957	.62
.5	.987	1.29	.952	.59
.7	.994	1.42	.951	.63
.9	.987	1.94	.948	.82
.95	.994	2.53	.951	.95
.99	.994	3.67	.947	1.35

from the tissues. When decompression happens too quickly, the liberated gas forms bubbles that can block blood vessels and damage tissue, producing the medical condition called decompression sickness (DCS) which can potentially lead to a lethal consequence (Atkins et al., 1988, Dromsky et al., 2000).

The body mass of adult sheep approximates that of adult humans. Thus, sheep offer a large animal model for DCS with susceptibility quite similar to the human. Sheep decompression trials are used to test decompression profiles viewed as too risky for humans (Lehner et al., 1997, Lehner et al., 2000). In a recent study at the University of Wisconsin, Madison, sheep were used as a model for submariners and divers to determine what might occur in humans undergoing similar dive profiles. Data on 1108 observations were collected. (Many sheep were used for multiple dives; however, the time interval between dives was considered large enough so that each dive can be viewed as independent) The experimental care and use of animals had been approved by the University of Wisconsin–Madison.

The major risk factors investigated in this study are exposure pressure (depth) and exposure duration. Each sheep underwent simulated dives (in a pressure chamber) with a designed pressure and duration and its outcome for CNS-DCS, limb bends, respiratory DCS and mortality was determined. The pressure was measured in absolute atmospheres and duration at depth was measured in minutes. All observed outcomes are dichotomous variables. Here we report on the mortality response.

A major goal of this study was to determine the ranges of dive depths and durations that correspond to certain risks of incurring a response, e.g. mortality. In particular, the researchers wanted to determine the range of depths and durations corresponding to a relatively low mortality rate – say .05. The empirical distribution of the two predictor variables (depth and duration) in the study is shown in Figure 2.

We fit a logistic regression of mortality (Y) on log base 10 exposure duration (X_1) and log base 10 exposure pressure (X_2). The fitted model is given by

$$\log \frac{p}{1-p} = -19.253 + 3.758X_1 + 14.196X_2\,, \tag{20}$$

where $p = E(Y \mid X_1, X_2)$ is the probability of mortality. The regression coefficients are all statistically significant at the .05 level. Hence, the two predictor variables are both important for predicting mor-

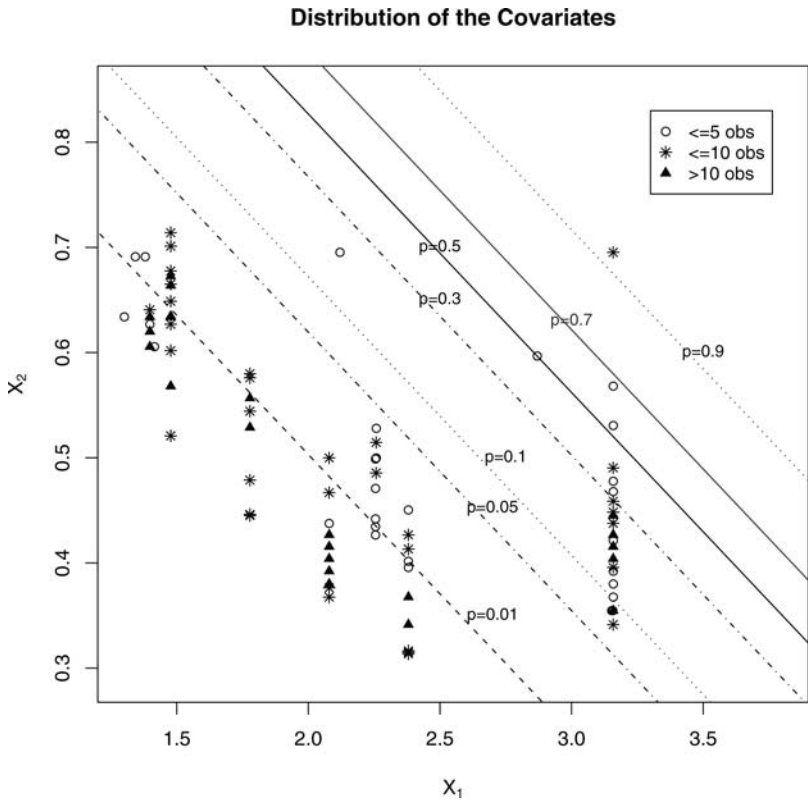

Figure 2 Design Plots for X_1 (log 10 duration) vs. X_2 (log 10 pressure) and Contour of the Fitted Surface from Logistic Regression for Sheep Example.

tality. Both a longer exposure duration and a larger exposure pressure (dive depth) are associated with a larger chance of death.

We determined the values of X_1 and X_2 that correspond to a fixed probability of mortality. For $p = .05, .1, .3, .5, .7, .9$, we display the contours of the fitted surface in Figure 2. Each line represents the estimated effective dose for a given p. We note that these straight lines have the same slopes (the same regression coefficients) but different intercepts.

For $p = .05$, we estimated the multi-dimensional effective doses for X_1 and X_2 by a straight line in the two-dimensional plane (Figure 3). We also produced the four types of confidence bands for these two-dimensional effective doses and depict them with different types of lines in the figures. Since Methods 2 and 4 depend on the order of conditioning, we chose to fix X_2 at a series of grid points x_2^* within the experimental range and then evaluate the x_2^*-conditioning effective dose of X_1. In Method 4, the bootstrap sample size was $B = 500$.

The results shown in these figures are consistent with our numerical study. Method 1 is the widest but is the only one satisfying the simultaneous coverage requirement. The other three methods are similar to each other. This is primarily due to the large sample size of our data set. In part because of the invariance to order, and also due to the results of our simulation studies, we focus on the Method 3 regions as the most useful in practice. Figures of this type will be useful to the diving community in assessing the dive profiles that can meet certain risk probabilities.

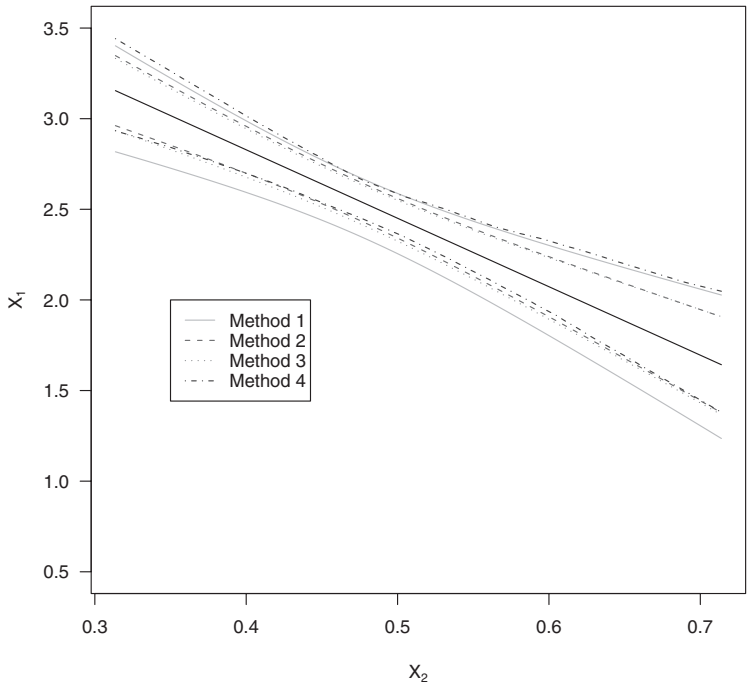

Figure 3 Plot of X_1 (log 10 duration) vs. X_2 (log 10 pressure) for mortality probability .05 and the associated confidence regions.

6 Discussion

In this paper we have generalized the concept of univariate effective dose to a multivariate setting. By conditioning on fixed values for all variables except one, we utilize an asymptotic inference procedure similar to that for the univariate effective dose. With our diving example, our development has proven useful in a practical application.

Our approach can be easily extended to include interaction or polynomial terms. However, the inclusion of these terms can result in confidence regions with complex shapes. Also, our methodology can be generalized to a broader range of statistical calibration problems involving multiple covariates. The distribution of the response variable can be any member of the exponential family. Our research thus may benefit a wide variety of scientific disciplines for which it is important to determine a confidence region of the predictor variables corresponding to a given response.

The choice of region for the conditioning variables should be based on practical needs. As demonstrated in our simulation studies and real examples, simultaneous confidence region tends to be wider and provides adequate coverage probability across the whole range of dose values. However, if we only need to concentrate on particular dose levels of some covariates, the pointwise confidence region for the conditioning effective dose might be sufficiently useful. Among the three pointwise methods, Fieller's method seems to have the best small sample performance.

In this paper we mainly focus on confidence bands for two predictors. When $k > 2$, the region is a high-dimensional band (or tube) and usually hard to visualize. The procedures introduced in Section 2 can be slightly modified to construct the confidence region. The results, as seen in a small simulation example in Section 4.2, are quite similar to what we demonstrate for two predictors.

The logistic regression function can suffer from computing instability for small sample sizes. Moreover, the parametric form of the dose-response relationship can be overly restrictive and might mis-

specify the real biological mechanisms. We believe that a flexible non-parametric or semi-parametric model might capture the underlying relationship more accurately. We intend to explore such an approach.

The procedure in this paper has been implemented in R. The computing code is available at

http://www.stat.nus.edu.sg/ ∼ stalj.

Acknowledgements We thank Kjell Doksum, Mari Palta, and Jun Shao for providing constructive comments. This work was partly funded by the University of Wisconsin Sea Grant Institute under grants from the National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, and from the State of Wisconsin. Federal grant number NA16RG2257, project number R/NI-31. We thank the U.S. Navy's Deep Submergence Biomedical Development program (NAVSEA) for supporting this research with funding through the OPNAV N773B Submarine Rescue and Diving Programs. This work was also supported by National Science Foundation grant DMS-03-53941 and DMS-07-05209. We thank Tass Dueland, Glen Moen, Jim Schwarz, Ralph Stauffacher, and Janet Welter for their roles in data acquisition. Generous comments from the Associated Editor and two referees have helped improve the manuscript.

References

Albert A. and Anderson J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10.

Atkins, C. E., Lehner, C. E., Beck, K. A., Dubietzig, R. R., Nordheim, E. V., and Lanphier, E. H. (1988). Experimental respiratory decompression sickness in sheep. *Journal of Applied Physiology* **65**, 1163–1171.

Cox, C. (1990). Fieller's theorem, the likelihood and the delta method. *Biometrics* **46**, 709–718.

Carter W. H., Chinchilli V. M., Wilson J. D., Campbell E. D., Kessler, F. K., and Carchman R. A. (1986). An asymptotic confidence region for the ED_{100p} from the logistic response surface for a combination of agents. *The American Statistician* **40**, 124–128.

Chen, D. G. (2007). Dose-Time-Response cumulative multinomial generalized linear model. *Journal of Biopharmaceutical Statistics* **17**, 173–185.

Davis, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods And Their Applications*. Cambridge Uni. press.

Dromsky, D. M., Toner, C. B., Survanshi, S., Fahlman, A., Parker, E., and Weathersby, P. (2000). Natural history of severe decompression sickness after rapid ascent from air saturation in a porcine model. *Journal of Applied Physiology* **89**, 791–798.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B* **16**, 175–185.

Finney, D. J. (1978). *Statistical method in biological assay.* New York: Macmillan.

Hsu, J. C. (1996). *Multiple Comparisons*. Chapman Hall/CRC, Boca Raton.

Lehner, C. E., Ball, R., Gummin, D. D., Lanphier, E. H., Nordheim, E. V., and Crump, P. M. (1997). Large-animal model of human decompression sickness: sheep database and preliminary analysis. Technical Report, NMRI 97-02. Bethesda, MD. Naval Med. Res. Institute.

Lehner, C. E., Dueland, R. T., Wilson, M. A., Nordheim, E. V., Vaicekavicius, L., Gendron-Fitzpatrick, A. P., and Crump, P. M. (2000). Decompression risk in submarine escape with "drop-outs" after hyperbaric saturation: UW sheep model findings. *Undersea Hyperbaric Medicine* **27** (Suppl.), 41.

Scheffe, H. (1959). *The Analysis of Variance*. Wiley: New York.

Shao, J. (1999). *Mathematical Statistics*. Springer-Verlag, New York.

