Journal of Biopharmaceutical Statistics

ISSN: 1054-3406 (Print) 1520-5711 (Online) Journal homepage: https://www.tandfonline.com/loi/lbps20

TWO-SAMPLE CONTINUAL REASSESSMENT METHOD

John O'Quigley , Larry Z. Shen & Anthony Gamst

To cite this article: John O'Quigley , Larry Z. Shen & Anthony Gamst (1999) TWO-SAMPLE CONTINUAL REASSESSMENT METHOD, Journal of Biopharmaceutical Statistics, 9:1, 17-44, DOI: 10.1081/BIP-100100998

To link to this article: https://doi.org/10.1081/BIP-100100998

Published online: 02 Feb 2007.

Journal of Biopharmaceutical Statistics, 9(1), 17–44 (1999)

TWO-SAMPLE CONTINUAL REASSESSMENT METHOD

John O'Quigley,[1] *Larry Z. Shen,*[2] *and Anthony Gamst*[1]

[1]Department of Mathematics
University of California, San Diego
La Jolla, California 92093

[2]Biometrics and Statistical Sciences
Procter & Gamble Pharmaceuticals
Cincinnati, Ohio 45242

Key words. Dose-toxicity; Patient heterogeneity; Phase I clinical trials; Sequential design

Abstract

We discuss an extension of the continual reassessment method (CRM) for use in phase I dose-finding studies. The extension enables the method to be applied to two groups of patients to determine the appropriate dose levels for each group. The method takes the specification of a simple relationship between the dose-toxicity curves for the two groups and runs the CRM on the bivariate model using maximum likelihood. We prove consistency of the method under fairly weak conditions and provide several simulations to give an idea how the method works in practice. We also undertake an evaluation of its performance by considering three possible situations: The first is the two-sample CRM, which directly uses a working model for the relationship between the two groups, carrying out a single trial using this method; the second situation carries out single trials for each of the two groups separately using the original (one-sample) CRM. The third situation is the case where such heterogeneity is ignored and the two groups are pooled into a single group, again using the original (one-sample) CRM. Simulations are carried out under a large class of model misspecifications, both of the dose-toxicity relationships and of the functional form

linking the groups, and are backed up by asymptotic results. Our conclusions match intuition: The first scheme gives the most favorable results when the two groups are different but share some features. When the groups are very different, the second scheme performs similarly to the first for finite sample sizes while having some advantages in terms of asymptotic efficiency. The third, as expected, gives the best results in the absence of patient heterogeneity. The two-sample method appears particularly advantageous when there may not be enough subjects in one of the subgroups for it to be feasible to carry out two trials.

1. Introduction

Phase I cancer clinical trials are usually carried out on subjects for whom all currently available therapies have failed. There will always be hope in the therapeutic potential of the new experimental treatment, but such hope is invariably tempered by the almost inevitable life-threatening toxicity accompanying the treatment. Given that candidates for these trials have no other options concerning treatment, their inclusion appears contingent on maintaining some acceptable degree of control over the toxic side effects while simultaneously trying to maximize treatment efficacy (which translates as dose). The continual reassessment method (CRM), introduced by O'Quigley, Pepe, and Fisher (1), has gained acceptance as an attractive approach to dealing with such problems, and its use has steadily increased. In these phase I trials it is not uncommon that patients be categorized into two or more groups, although making distinctions among more than two groups may not be practically feasible given the typically small numbers of available subjects. An example of such a group distinction might be the degree of previous treatment known to be related to the patient's ability to tolerate a new regimen. Another example concerns patients with acute leukemia, where it has been observed that younger patients will better tolerate aggressive new treatments. We refer to the problem of determining the target dose level for two potentially heterogeneous groups of patients as the two-sample problem.

It is generally difficult to apply conventional design procedures [for example, the up-and-down scheme and its variants described by Storer (2)] to this situation, apart from carrying out two separate trials, one for each group. In fact, when it is believed that there are two distinct prognostic groups, the usual method is to carry out two separate trials. This method, however, has the disadvantage of failing to utilize efficiently information common to both groups and may not be feasible if one of the groups turns out to be too small. We develop a two-sample CRM that specifies a simple relationship between the dose-toxicity curves of the two groups and carries out a single trial based on information from both groups. The idea is that such a specification will make better use of the

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

information common to both groups. Of course, we also want the method to behave well if there turns out to be little shared information, or if the difference between the two groups turns out to be small. In particular, we are mainly interested in situations in which our working models are misspecified, with respect to both within-group toxic probabilities and the functional link between groups. We address the following questions:

1. For the situation of balanced groups in which it is possible to carry out two separate studies, how does the two-sample CRM compare to the original (one-sample) CRM applied separately to the two groups?
2. If the difference between the two groups turns out to be small, will the two-sample CRM give results comparable to the one-sample CRM carried out on the two groups combined?
3. Can the two-sample CRM be of use in the presence of such strong subgroup size imbalances that separate trials are not plausible?
4. Is it possible to test for the presence of patient heterogeneity using the two-sample CRM?

Our intuition tells us that if the two groups differ, then combining them could lead to the selection of a target dose level that is suboptimal for both groups, depending on just how different the two groups are. In this case we ought to be better off treating the two groups as distinct and using either the two-sample CRM or the one-sample CRM applied to the two groups separately. This intuition is confirmed by both theoretical arguments and Monte Carlo simulations. Our intuition also tells us that if the two groups are distinct but share some information, we ought to be able to exploit this information using an appropriate model, even if the model is only a working approximation to the true dose-toxicity relationship. This, too, is confirmed by theoretical arguments and Monte Carlo experiments. If the two groups are vastly different and sample sizes are small, our intuition suggests that the two-sample CRM may bias the estimate of target dose levels and we could be better off using the one-sample method applied to the two groups separately. Starting levels will also have an influence here. Should there be substantial imbalance in the two group sample sizes and should it not be realistic to carry out two separate studies, sharing information via the two-sample CRM may be an effective solution. These conjectures are confirmed by Monte Carlo simulation. For large samples, we show that the two-sample CRM is strongly consistent under mild conditions, and so the choice between the two-sample method and the "two-groups, two-trials" method becomes one of efficiency (and the two-sample method turns out to be more efficient).

In Section 2, we introduce the two-sample CRM and contrast its performance to that of other schemes for dealing with the two-sample problem: the

two-groups, two-trials scheme, and the one-sample scheme where the two groups are pooled together. In Section 3 we discuss a single simulation for the two-sample CRM with two models for the intergroup dose-toxicity relationship. In Section 4, we make a theoretical comparison among the two-sample CRM, the original CRM carried out on the two groups separately, and the one-sample scheme where the two groups are treated as homogeneous. In Section 5 we give the results from extensive Monte Carlo simulations on the three schemes. Proofs including the proof of strong consistency of the two-sample CRM are contained in the Appendixes.

2. The Two-Sample CRM and Other Schemes

Let I be the indicator variable for the two groups and take the values 1 or 2. Suppose that the target probability of toxicity for the ith groups is $\theta_i \in (0, 1)$, $i = 1, 2$. For simplicity, we only consider the case $\theta_1 = \theta_2 = \theta$ throughout, although this assumption is not essential to our conclusions. We test subjects at one of m dose levels, $d_1, \ldots, d_m$, and assume the two groups share these dose levels; that is, in principle, every patient, regardless of group, could be treated at any one of the m levels.

Let Y be a zero-one random variable representing the outcome of a treatment on a particular patient, where $Y = 1$ represents the observation of a toxicity and $Y = 0$ represents a nontoxicity. If a treatment is carried out at dose level x, the probability of observing toxicity is

$$P(Y = 1 \,|\, x,\ I = i) = R_i(x), \quad i = 1,\ 2.$$

For each i, $R_i(\cdot)$, the true dose-toxicity curve for group i, is assumed to be an increasing function of x. The target dose level for the ith group is defined to be D_i, where

$$R_i(D_i) = \theta, \quad i = 1,\ 2. \tag{2.1}$$

Usually, none of the $d_1, \ldots, d_m$ will exactly satisfy either of the equations (2.1). In that case, a level is treated as the target level if the corresponding probability of toxicity is closest to θ. If $R_1(\cdot)$ and $R_2(\cdot)$ were known, one would be able to solve for the values of D_i from (2.1). Of course, in practice the actual dose-toxicity curve is never known. Following O'Quigley et al. (1), we can introduce a parametric model for the unknown dose-toxicity relationships and carry out an experiment based on the model dose-toxicity curves. After a number of patients have been treated, we will be able to estimate the parameters and get a corresponding (local) estimate of the dose-toxicity curves. Then we can estimate the target dose level by finding the doses that come closest to solving (2.1) for the model. We describe three schemes based on this general method.

2.1 The Two-Sample CRM and Its Implementation

For scheme I, we model the dose-toxicity curves as

$$P(Y = 1 \mid x, I = 1) = \psi_1(x, a),$$
$$P(Y = 1 \mid x, I = 2) = \psi_2(x, a, b),$$

where (a, b) is a pair of one-dimensional parameters. This specification reflects the fact that the two groups have some common ground in terms of their reaction to the treatments. The common ground is reflected in the parameter a and that distinguishing the groups in the parameter b.

The functions ψ_1 and ψ_2 are selected in such a way that for each $\theta \in (0, 1)$ and each dose level x there exists (a_0, b_0) satisfying $\psi_1(x, a_0) = \theta$ and $\psi_2(x, a_0, b_0) = \theta$. This condition is satisfied by many function-pairs.

Let $Z_j = (x_j, y_j, I_j)$, $i = 1, \ldots, k$, be the outcomes of the first k patients, where I_j indicates to which group the jth subject belongs, x_j is the dose level at which the jth subject is tested, and y_j indicates whether or not the jth subject suffered a toxic response. On the basis of the observations $Z_j = (x_j, y_j, I_j)$, $i = 1, \ldots, k$, on the first k_1 patients in group 1 and k_2 patients in group 2 ($k_1 + k_2 = k$), we can write down the likelihood as

$$\prod_{i=1}^{k_i} \psi_1(x_i, a)^{y_i}(1 - \psi_1(x_i, a))^{1-y_i} \times \prod_{i=k_1+1}^{k} \psi_2(x_i, a, b)^{y_i}(1 - \psi_2(x_i, a, b))^{1-y_i}$$

and denote by $(\hat{a}_k, \hat{b}_k)$ the values of a and b maximizing this equation. Two-dimensional Newton-Raphson, combined with simplex maximizing techniques, have worked well in our experience, although there may well be more efficient algorithms.

The estimated dose-toxicity curves are $\psi_1(x, \hat{a}_k)$ and $\psi_2(x, \hat{a}_k, \hat{b}_k)$, respectively. If the next patient belongs to group 1, the recommended dose for his or her treatment is the x_{k+1} that minimizes $|\psi_1(x, \hat{a}_k) - \theta|$ over $d_1, \ldots, d_m$. On the other hand, if the next subject is in group 2, the recommended dose for his or her treatment is the x_{k+1} that minimizes $|\psi_2(x, \hat{a}_k, \hat{b}_k) - \theta|$ over the available doses.

There are many ways to go about estimating the parameters (a, b). For example, one can use the Bayesian estimates proposed by O'Quigley et al. (1) or the maximum likelihood strategy studied by O'Quigley and Shen (3). Experience shows that for moderate sample sizes the Bayesian and maximum likelihood estimates give similar results unless there is some evidence in favor of a particular prior, in which case the Bayesian method often does better due to the difficulty in starting the maximum likelihood strategy. For the maximum likelihood approach it is necessary to wait until subjects in both groups have heterogeneous responses (both toxic and nontoxic); this can be troublesome in very small sam-

ples. Operating characteristics may be influenced by the choice of initial escalation scheme, and this is an area that warrants further study. The use of vague priors and various initial dose allocation schemes will be the subject of future work. For convenience, we focus here on the maximum likelihood strategy.

2.2 Two Groups, Two Separate Trials

For scheme II, we simply carry out a one-sample scheme on each of the two groups separately. For instance, one can carry out a trial for the first group using $\psi_1(x, \cdot)$ as the model dose-toxicity curve. The implementation of the procedure is then the same as that of Shen and O'Quigley (4). As for the second group, we could use the same model (ψ_1) for the dose-toxicity curve and the same implementation as well. Shen and O'Quigley (4) show that for a large class of incorrectly specified dose-toxicity models the recommended dose levels will converge to the correct dose levels under maximum likelihood. The actual conditions leading to nonconvergence, identified in (4), are similar to the conditions for the two-sample method (see the Appendixes). Note that the two trials are carried out independently of one another; that is, the dose allocation for one group does not rely on information from the other group.

2.3 Ignore the Group Difference and Pool the Two Groups

For scheme III, we ignore the group difference and pool the two groups of patients together. The original (one-sample) CRM is then applied to the combined group using ψ_1 as the model dose-toxicity curve. The details again are the same as in (4). Intuitively, this method should give the best results when the two dose-toxicity curves do not differ much near the target dose because it makes the most use of the available information. However, when the two groups have different dose-toxicity curves and therefore different target dose levels, we would expect the recommended dose to be some kind of average. This phenomenon will be seen in Section 5.

Examples

The following are a few candidates for the two-sample dose-toxicity curves; models 1 and 2 are variants of the model motivated by O'Quigley et al. (1), and model 3 is the more familiar logistic dose-response model.

(Model 1)
$$\begin{cases} \psi_1(x, a) = \left(\dfrac{\tanh(x) + 1}{2} \right)^a, \quad a > 0, \\ \psi_2(x, a, b) = \left(\dfrac{\tanh(x) + 1}{2} \right)^{a+b}, \quad a > 0, \ -a < b < \infty; \end{cases}$$

$$\text{(Model 2)} \qquad \begin{cases} \psi_1(x, a) = \left(\dfrac{\tanh(x) + 1}{2} \right)^a, \quad a > 0, \\ \psi_2(x, a, b) = \left(\dfrac{\tanh(x - b) + 1}{2} \right)^a, \quad a > 0, \ -\infty < b < \infty; \end{cases}$$

$$\text{(Model 3)} \qquad \begin{cases} \psi_1(x, a) = \dfrac{\exp(a + x)}{1 + \exp(a + x)}, \quad -\infty < a < \infty, \\ \psi_2(x, a, b) = \dfrac{b \ \exp(a + x)}{1 + b \ \exp(a + x)}, \quad -\infty < a < \infty, \ 0 < b < \infty. \end{cases}$$

All three models have the common feature that the second curve represents some kind of shift of the first. In model 1 the second curve shifts at the log scale—the second curve is a positive power of the first. In model 2 the second curve represents a shift at the original scale of doses—that is, the second curve is just a copy of the first "centered" at a (potentially) different dose. A feature of these models is that there exists a value of b, zero in the first two examples and one in the last, such that $\psi_2(x, a, b) \equiv \psi_1(x, a)$ for each a, signalling no difference between the two dose-toxicity curves and allowing one to test for the presence of patient heterogeneity.

3. A Simulated Example

Note that the techniques we present cannot be applied retrospectively; the design and analysis proceed sequentially and alternate as in stochastic approximation. In this section we discuss a Monte Carlo simulation of scheme I under models 1 and 2 from Section 2. We assume two distinct groups, and there are six dose levels, $d_1, \ldots, d_6$. For each group the probability of toxicity at each dose is given in Table 1. The probabilities have been chosen arbitrarily for illustrative purposes to represent the effect of misspecification. An inspection of Table 1 reveals that, for a chosen target of $\theta_0 = 0.20$, where θ_0 is the probability of toxicity at the targeted level, the corresponding target doses are d_2 and d_4 for group 1 and group 2, respectively.

Two groups of patients are simulated such that their dose-toxicity curves are those given in the table. Scheme I was applied using maximum likelihood

Table 1. The Probability of Toxicity at the Six Dose Levels

Group 1	0.02	0.19	0.31	0.45	0.51	0.63
Group 2	0.03	0.05	0.11	0.21	0.39	0.50

for both model 1 and model 2. Subjects entered the simulation randomly with probability 0.5 of being in group 1. This enabled us to simulate unequal sample sizes in each group as well as a random recruitment scheme. This may be important because, in practice, we cannot expect patients to present themselves in pairs, one from each group. Because maximum likelihood was used, it was necessary to have a method for dose allocation until the observation of heterogeneous responses in both groups (otherwise the likelihood is unbounded). There are many possibilities for achieving this, and we used the following scheme. Both groups start out at dose level 1, and at this level we wait to observe two successive nontoxicites before experimenting at one level higher. Thereafter, if three nontoxicities are observed in a row for subjects in the same group at the same dose level, then the recommended dose level is raised by one for the next cohort of subjects in that group. In other words, escalation prior to the implementation of the model takes place separately in the groups, and, after having seen three patients tolerating the treatment at a given dose, starting with the lowest. As soon as the first toxicity is observed, then, for the group in which the toxicity was observed, we can proceed using the usual one-sample CRM. For the other group, escalation continues independently. Once we observe a toxicity in the second group, then we have enough information to fit the model and dose allocation proceeds on the basis of two-sample CRM. This strategy is almost certainly suboptimal, but the optimal strategy for starting off the maximum likelihood method is unknown. A fairly typical realization of scheme I is given in Table 2.

Table 2. A Simulated Experiment with 32 Subjects

i	I_i	x_i	y_i	x^1_{i+1}	x^2_{i+1}	i	I_i	x_i	y_i	x^1_{i+1}	x^2_{i+1}
1	1	d_1	0	d_1	d_1	17	1	d_4	1	d_4	d_3
2	1	d_1	0	d_2	d_1	18	2	d_3	0	d_4	d_3
3	1	d_2	0	d_2	d_1	19	2	d_3	0	d_3	d_3
4	1	d_2	0	d_2	d_1	20	1	d_3	0	d_3	d_3
5	1	d_2	0	d_3	d_1	21	1	d_3	1	d_3	d_3
6	2	d_1	0	d_3	d_1	22	2	d_3	0	d_3	d_3
7	2	d_1	0	d_3	d_2	23	1	d_2	0	d_2	d_4
8	1	d_3	0	d_3	d_2	24	1	d_2	0	d_2	d_4
9	2	d_2	0	d_3	d_2	25	2	d_4	0	d_2	d_4
10	2	d_2	0	d_3	d_3	26	2	d_4	0	d_2	d_4
11	1	d_3	0	d_3	d_2	27	1	d_2	0	d_2	d_4
12	2	d_2	0	d_3	d_3	28	1	d_2	0	d_2	d_4
13	2	d_3	0	d_3	d_3	29	1	d_2	1	d_2	d_4
14	2	d_3	1	d_3	d_3	30	1	d_2	0	d_2	d_4
15	1	d_3	0	d_4	d_3	31	1	d_2	1	d_2	d_4
16	2	d_3	0	d_4	d_3	32	1	d_2	0	d_2	d_4

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 3. Distributional Results for the Final Recommended Doses Based on 2000 Simulations from Models 1 and 2

			Group one					Group two				
Model	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
1	6–12	25	0.64	0.64	0.64	0.68	0.68	0.36	0.36	0.88	0.88	0.96
	13–19	151	0.64	0.64	0.64	0.85	0.85	0.42	0.42	0.75	0.75	0.85
	20–26	24	0.63	0.63	0.63	0.83	0.83	0.50	0.50	0.93	0.93	1.00
2	6–12	15	0.53	0.53	0.53	1.00	1.00	0.40	0.40	0.80	0.80	0.93
	13–19	161	0.55	0.55	0.55	0.80	0.80	0.35	0.35	0.70	0.70	0.77
	20–26	24	0.63	0.63	0.63	0.88	0.88	0.29	0.29	0.63	0.63	0.67

In this example there are 19 subjects from group 1 and 13 from group 2, and the groups alternate in a random pattern. The first column of Table 2 contains the patients' sequence numbers; the second column indicates to which group the subjects belong; the third tells us at which dose the subject was treated; the fourth indicates whether or not a toxicity was observed; and the fifth and sixth columns give the recommended dose levels for the next subject in group 1 and group 2 (respectively). The illustration of Table 2, although encouraging, represents only a single trial. Average performance over many trials is more informative; this is shown in Table 3. The table summarizes distributional information related to the recommended dose levels after the last subject was treated. The first column specifies the model being simulated; the second and third columns (labeled G1 and #, respectively) indicate the number of subjects that were treated in group 1 and their distribution; and the last five columns, under the headings $\alpha \in \{0.05, 0.08, 0.10, 0.13, 0.15\}$, give the proportion of simulations such that the recommended dose level was within α of 0.20. For example, column 5 (labeled 0.08) gives the proportion of simulations recommending a dose within 0.08 of 0.20. This kind of distributional information is useful both for the comparisons to be made in Section 5 and in practice, where the true probability of toxicity at dose d_j will be unknown.

4. A Study of Schemes

This section is devoted to a theoretical investigation of the two-sample CRM (scheme I), the original CRM carried out on the two groups separately (scheme II), and the one-sample method where the two groups are treated as homogenous (scheme III). The setup is the same as in Section 2. Scheme I models the dose-

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

toxicity relationship as

$$P(Y = 1 \mid x, I = 1) = \psi_1(x, a),$$
$$P(Y = 1 \mid x, I = 2) = \psi_2(x, a, b), \tag{4.1}$$

where Y is the indicator of toxic response, x is the dose level, and I is the group specifier. For the purposes of comparison, the model for scheme II is given by

$$P(Y = 1 \mid x, I = 1) = \psi_1(x, a),$$
$$P(Y = 1 \mid x, I = 2) = \psi_2(x, a_0, b) = \psi_2(x, b), \tag{4.2}$$

where a_0 is the target value for the parameter a; and the model for scheme III is

$$P(Y = 1 \mid x) = \psi_1(x, a). \tag{4.3}$$

All schemes use maximum likelihood for parameter estimation; the target parameters are denoted a_0 and b_0; and $\theta_0^i = R_i(x_0^i)$ is the underlying probability of toxicity $[R_i(\cdot)]$ at the target dose (x_0^i) for group $i = 1, 2$. Whichever scheme is adopted, correct dose recommendation relies in large part on precise estimation of the parameters in the model. The asymptotic variances for the corresponding estimates give some indication of their relative precision; thus, at least from a theoretical point of view, it is sufficient to base comparisons of the various schemes on comparisons of asymptotic variance.

We suppose the two groups are homogeneous. The optimal scheme when there are no differences between the two groups is scheme III because it makes the best use of the available information. Indeed, in this case, scheme III is consistent under mild regularity conditions (4), and the estimate $\hat{a}_n$ of the parameter a_0 is asymptotically normal with variance

$$\{\psi_1^{(a)}(x_0, a_0)\}^{-2}\, \theta_0(1 - \theta_0), \tag{4.4}$$

in which $\psi_i^{(a)} = \partial\psi_i(x, a)/\partial a (i = 1, 2)$. Scheme II, on the other hand, performs two experiments—one with a fraction p of the subjects and the other with $1 - p$. The parameter estimates are asymptotically normal with variances

$$V(\hat{a}_n) = \frac{\theta_0(1 - \theta_0)}{p\{\psi_1^{(a)}(x_0, a_0)\}^2},$$
$$V(\hat{b}_n) = \frac{\theta_0(1 - \theta_0)}{(1 - p)\{\psi_1^{(a)}(x_0, a_0)\}^2},$$

using ψ_1 as the model for both groups. So the asymptotic variance is expanded by at least $(1/\max(p,(1 - p)))$, but by properly averaging the two parameter estimates we can recover the "optimal" asymptotic variance (4.4). In a sense then, we lose nothing asymptotically. For finite samples, of course, we do lose some information. Because scheme I takes advantage of the information common to

both groups, it can do better than scheme II. Again, by averaging, we can recover the asymptotic variance (4.4), but the finite sample behavior (see Section 5) appears to be slightly better than scheme II.

When the two groups really are different, scheme III is not even consistent (the recommended dose converges to some kind of an average of the target doses for the two groups). So, under group differences, we compare only scheme I and scheme II. Assume that among the n patients already treated, n_1 of them belong to group 1 and the rest (n_2) to group 2 ($n_1 + n_2 = n$). For Scheme I, the maximum likelihood estimates ($\hat{a}_{1n}$, $\hat{b}_{1n}$) of (a, b) maximize

$$\prod_{i=1}^{n_1} \psi_1(x_i, a)^{y_i}(1 - \psi_1(x_i, a))^{1-y_i} \times \prod_{i=n_1+1}^{n} \psi_2(x_i, a, b)^{y_i}(1 - \psi_2(x_i, a, b))^{1-y_i}.$$

For scheme II, $\hat{a}_{2n}$ is chosen to maximize

$$\prod_{i=1}^{n_1} \psi_1(x_i, a)^{y_i}(1 - \psi_1(x_i, a))^{1-y_i}$$

and $\hat{b}_{2n}$ maximizes

$$\prod_{i=n_1+1}^{n} \psi_2(x_i, a_0, b)^{y_i}(1 - \psi_2(x_i, a_0, b))^{1-y_i}.$$

We introduce the functions

$$s_1(x, y, a) = y \log \psi_1(x, a) + (1 - y)\log(1 - \psi_1(x, a))$$

and

$$s_2(x, y, a, b) = y \log \psi_2(x, y, a, b) + (1 - y)\log(1 - \psi_2(x, y, a, b))$$

together with their first-order partial derivatives (the score functions)

$$s_{11} = \frac{\partial s_1}{\partial a}, \qquad s_{21} = \frac{\partial s_2}{\partial a}, \qquad s_{22} = \frac{\partial s_2}{\partial b}.$$

Clearly, ($\hat{a}_{1n}$, $\hat{b}_{1n}$) satisfy

$$\begin{cases} \sum_{i=1}^{n_1} s_{11}(x_i, y_i, \hat{a}_{1n}) + \sum_{i=n_1+1}^{n} s_{21}(x_i, y_i, \hat{a}_{1n}, \hat{b}_{1n}) = 0 \\ \sum_{i=n_1+1}^{n} s_{22}(x_i, y_i, \hat{a}_{1n}, \hat{b}_{1n}) = 0 \end{cases} \tag{4.5}$$

and ($\hat{a}_{2n}$, $\hat{b}_{2n}$) satisfy

$$\begin{cases} \sum_{i=1}^{n_1} s_{11}(x_i, y_i, \hat{a}_{2n}) = 0 \\ \sum_{i=n_1+1}^{n} s_{22}(x_i, y_i, a_{2n}, \hat{b}_{2n}) = 0. \end{cases} \tag{4.6}$$

Next, we derive approximate formulas for the asymptotic variances of $(\hat{a}_{1n}, \hat{b}_{1n})$ and $(\hat{a}_{2n}, \hat{b}_{2n})$. The variances of $\hat{a}_{1n}$ and $\hat{a}_{2n}$ serve as indicators of the relative performance of scheme I and scheme II. Under relatively weak conditions one can show that $(\hat{a}_{1n}, \hat{b}_{1n})$ tend to (a_0, b_0) almost surely (see Appendix 2). The technical details are similar to those in Huber (5) and Shen and O'Quigley (4). For ease of expression we drop the arguments (x_i, y_i, a_0) and (x_i, y_i, a_0, b_0) wherever possible. Approximate formulas for the asymptotic variances of $\hat{a}_{1n}$ and $\hat{b}_{1n}$ are given below.

Theorem 1: Assume $n_1/n \to p$ for some $0 < p < 1$ as $n \to \infty$. Then we can make the following asymptotic approximations:

$$V(\hat{a}_{1n}) = \frac{\Sigma_{i=n_1+1}^{n} s_{22}^2}{\Sigma_{i=n_1+1}^{n} s_{22}^2 \left(\Sigma_{i=1}^{n_1} s_{11}^2 + \Sigma_{i=n_1+1}^{n} s_{21}^2\right) - \left(\Sigma_{i=n_1+1}^{n} s_{21} s_{22}\right)^2} \tag{4.7}$$

and

$$V(\hat{b}_{1n}) = \frac{\Sigma_{i=1}^{n_1} s_{11}^2 + \Sigma_{i=n_1+1}^{n} s_{21}^2}{\Sigma_{i=n_1+1}^{n} s_{22}^2 \left(\Sigma_{i=1}^{n_1} s_{11}^2 + \Sigma_{i=n_1+1}^{n} s_{21}^2\right) - \left(\Sigma_{i=n_1+1}^{n} s_{21} s_{22}\right)^2}, \tag{4.8}$$

where the above formulas are evaluated at (a_0, b_0).

A proof of the theorem is given in Appendix 1.

Remark: Replacing (a_0, b_0) in (4.7 and 4.8) with the estimates $(\hat{a}_{1n}, \hat{b}_{1n})$ enables us to estimate standard errors for $\hat{a}_{1n}$ and $\hat{b}_{1n}$ and thus get approximate confidence intervals.

For scheme II, we have the following corollary.

Corollary 1: Under the conditions of Theorem (1), the asymptotic variances of $\hat{a}_{2n}$ *and* $\hat{b}_{2n}$ can be approximated by

$$V(\hat{a}_{2n}) = \left(\sum_{i=1}^{n_1} s_{11}^2\right)^{-1} \tag{4.9}$$

and

$$V(\hat{b}_{2n}) = \left(\sum_{i=n_1+1}^{n} s_{22}^2\right)^{-1}, \tag{4.10}$$

where the formulas are evaluated at (a_0, b_0).

To compare the performance of $\hat{a}_{1n}$ and $\hat{a}_{2n}$ we apply the Cauchy-Schwarz inequality

$$\left(\sum_{i=n_1+1}^{n} s_{21} s_{22}\right)^2 \le \sum_{i=n_1+1}^{n} s_{21}^2 \sum_{i=n_1+1}^{n} s_{22}^2,$$

from which it follows that $V(\hat{a}_{1n}) \leq V(\hat{a}_{2n})$. Moreover, a sufficient condition for their equality is that there exist a scalar κ such that

$$s_{21}(x, y, a_0, b_0) = \kappa s_{22}(x, y, a_0, b_0), \quad \text{for all } x, y. \tag{4.11}$$

Looking at the examples from Section 2, we see that only model 1 satisfies (4.11). Therefore, we expect $V(\hat{a}_{1n}) \leq V(\hat{a}_{2n})$ for both model 2 and model 3. In this sense, we expect scheme I to do better than scheme II for models 2 and 3 and just as well for model 1.

Another way of comparing schemes I and II is to look at the asymptotic variance-covariance matrices. If we use this as a measure of reliability, scheme II does better than scheme I unless

$$\frac{\Sigma_{i=n_1+1}^{n} s_{21}^2}{(\Sigma_{i=n_1+1}^{n} s_{22}^2)\,(\Sigma_{i=n_1+1}^{n} s_{11}^2)}$$

is small (see Appendix 1). Looking at this quantity as a measure of how different the groups are under the model, we see that scheme II does better the further apart the two groups. This behavior can be seen in the simulations in Section 5.

5. Monte Carlo Studies

A Monte Carlo study was carried out to investigate the three schemes. The total number of subjects in each trial was 32. For schemes I and III, the 32 subjects were assigned randomly (with equal probability) to either group 1 or group 2. For scheme II, 16 subjects were run in each of the two groups. A total of eight different situations were explored (see Tables 4–19), each selected to address a different aspect of the two-sample CRM and each testing at six different dose levels. The simulations were run a total of $N = 2000$ times on each scheme, including both models 1 and 2 from scheme I (see Section 2). The target toxicity level for each trial was $\theta_0 = 0.20$. The start-up method, the same for all the schemes, was the "two-up" method discussed in Section 3.

Because model misspecification was of primary concern, all but one of the simulated situations (situation C, Tables 8 and 9) did not fit the model, in terms of both the working models for the within-group toxic probabilities and the func-

Table 4. Situation A: The Probability of Toxicity at the Six Dose Levels

Group 1	0.02	0.11	0.24	0.39	0.56	0.73
Group 2	0.02	0.11	0.24	0.39	0.56	0.73

Table 5. Distributional Results for Situation A

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	192	0.24	0.24	0.81	0.81	0.81	0.45	0.45	0.83	0.83	0.83
	13–19	1593	0.29	0.29	0.82	0.82	0.82	0.44	0.44	0.78	0.78	0.78
	20–32	215	0.31	0.31	0.79	0.79	0.79	0.39	0.39	0.71	0.71	0.71
S I, model 2	0–12	206	0.38	0.38	0.85	0.85	0.85	0.55	0.55	0.82	0.82	0.82
	13–19	1578	0.43	0.43	0.89	0.89	0.89	0.46	0.46	0.75	0.75	0.75
	20–32	216	0.35	0.35	0.85	0.85	0.85	0.38	0.38	0.71	0.71	0.71
Scheme II	16	2000	0.31	0.31	0.82	0.82	0.82	0.31	0.31	0.80	0.80	0.80
Scheme III	0–12	177	0.42	0.42	0.88	0.88	0.88	0.42	0.42	0.88	0.88	0.88
	13–19	1573	0.42	0.42	0.91	0.91	0.91	0.42	0.42	0.91	0.91	0.91
	20–32	250	0.47	0.47	0.91	0.91	0.91	0.47	0.47	0.91	0.91	0.91

Distributional results for the final recommended doses based on 2000 simulations of each scheme and each model for the case where no group differences exist.

Table 6. Situation B: The Probability of Toxicity at the Six Dose Levels

Group 1	0.01	0.03	0.09	0.11	0.20	0.35
Group 2	0.07	0.23	0.31	0.35	0.45	0.57

Table 7. Distributional Results for Situation B

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	187	0.27	0.27	0.53	0.77	0.77	0.44	0.44	0.44	0.92	0.99
	13–19	1567	0.08	0.08	0.62	0.77	0.85	0.42	0.42	0.42	0.90	0.99
	20–32	246	0.10	0.10	0.73	0.89	0.91	0.28	0.28	0.28	0.79	0.96
S I, model 2	0–12	193	0.34	0.34	0.51	0.84	0.84	0.46	0.46	0.46	0.91	0.98
	13–19	1562	0.05	0.05	0.33	0.76	0.85	0.32	0.32	0.32	0.91	0.99
	20–32	245	0.07	0.07	0.48	0.84	0.86	0.27	0.27	0.27	0.88	0.99
Scheme II	16	2000	0.39	0.39	0.68	0.82	0.92	0.57	0.57	0.57	0.91	0.98
Scheme III	0–12	202	0.02	0.02	0.20	0.44	0.44	0.54	0.54	0.54	0.79	0.97
	13–19	1577	0.06	0.06	0.30	0.57	0.57	0.41	0.41	0.41	0.69	0.93
	20–32	221	0.20	0.20	0.50	0.78	0.78	0.22	0.22	0.22	0.50	0.80

Distributional results for the final recommended doses based on 2000 simulations of each scheme and each model for the case where the two groups are very different.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 8. Situation C: The Probability of Toxicity at the Six Dose Levels

Group 1	0.07	0.20	0.35	0.55	0.70	0.75
Group 2	0.01	0.05	0.10	0.20	0.32	0.38

Table 9. Distributional Results for Situation C

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	201	0.47	0.47	0.47	0.67	0.94	0.49	0.49	0.80	0.91	0.98
	13–19	1585	0.57	0.57	0.57	0.79	0.96	0.48	0.48	0.79	0.89	0.98
	20–32	214	0.65	0.65	0.65	0.79	0.95	0.36	0.36	0.66	0.87	1.00
S I, model 2	0–12	189	0.55	0.55	0.55	0.67	0.92	0.41	0.41	0.66	0.81	0.88
	13–19	1560	0.59	0.59	0.59	0.75	0.97	0.34	0.34	0.63	0.79	0.86
	20–32	251	0.57	0.57	0.57	0.72	0.96	0.34	0.34	0.58	0.80	0.95
Scheme II	16	2000	0.63	0.63	0.63	0.82	0.97	0.36	0.36	0.61	0.82	0.95
Scheme III	0–12	169	0.32	0.32	0.32	0.33	0.82	0.17	0.17	0.66	0.68	1.00
	13–19	1601	0.53	0.53	0.53	0.54	0.92	0.08	0.08	0.46	0.46	0.99
	20–32	230	0.69	0.69	0.69	0.73	0.97	0.03	0.03	0.27	0.27	0.96

Distributional results for the final recommended doses based on 2000 simulations of each scheme and each model for the case where scheme I, model 2 is the data-generating process.

tional link between the two groups. Another important concern was imbalance in the group sample sizes. To address this question, the results for schemes I and III are split into three categories based on the number of subjects in group 1 (0–12, 13–19, or 20–32; see column G1 in the tables).

Tables 4–19 delineate the results of the Monte Carlo experiments. Tables 4, 6, 8, 10, 12, 14, 16, and 18 list the probability of toxicity at each of the six dose levels for both groups 1 and 2. Tables 9, 11, 13, 15, 17, and 19 give distributional results. The first three columns of these tables give the scheme (and model), the number of subjects in group 1 (G1), and the number of simulations fitting this profile (#). The next several columns give the proportion of the simu-

Table 10. Situation D: The Probability of Toxicity at the Six Dose Levels

Group 1	0.02	0.19	0.31	0.45	0.51	0.63
Group 2	0.03	0.05	0.11	0.21	0.39	0.50

Table 11. Distributional Results for Situation D

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	234	0.53	0.53	0.53	0.76	0.76	0.51	0.51	0.80	0.80	0.89
	13–19	1512	0.59	0.59	0.59	0.80	0.80	0.47	0.47	0.75	0.75	0.86
	20–32	254	0.64	0.64	0.64	0.88	0.88	0.36	0.36	0.68	0.68	0.81
S I, model 2	0–12	204	0.48	0.48	0.48	0.81	0.81	0.49	0.49	0.76	0.76	0.81
	13–19	1546	0.58	0.58	0.58	0.85	0.85	0.34	0.34	0.64	0.64	0.73
	20–32	250	0.65	0.65	0.65	0.88	0.88	0.29	0.29	0.55	0.55	0.69
Scheme II	16	2000	0.65	0.65	0.65	0.84	0.84	0.34	0.34	0.59	0.59	0.77
Scheme III	0–12	207	0.29	0.29	0.29	0.68	0.68	0.29	0.29	0.68	0.68	0.97
	13–19	1579	0.44	0.44	0.44	0.83	0.83	0.15	0.15	0.54	0.54	0.98
	20–32	214	0.55	0.55	0.55	0.88	0.88	0.12	0.12	0.45	0.45	1.00

Distributional results for the final recommended doses based on 2000 simulations from each scheme and each model under "normal" circumstances.

Table 12. Situation E: The Probability of Toxicity at the Six Dose Levels

Group 1	0.03	0.19	0.39	0.61	0.83	0.95
Group 2	0.01	0.07	0.19	0.37	0.55	0.79

Table 13. Distributional Results for Situation E

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	204	0.66	0.66	0.66	0.66	0.66	0.55	0.55	0.55	0.77	0.77
	13–19	1577	0.66	0.66	0.66	0.66	0.66	0.51	0.51	0.51	0.72	0.72
	20–32	219	0.69	0.69	0.69	0.69	0.69	0.45	0.45	0.45	0.66	0.66
S I, model 2	0–12	189	0.62	0.62	0.62	0.62	0.62	0.49	0.49	0.49	0.75	0.75
	13–19	1591	0.67	0.67	0.67	0.67	0.67	0.49	0.49	0.49	0.72	0.72
	20–32	220	0.65	0.65	0.65	0.65	0.65	0.38	0.38	0.38	0.64	0.64
Scheme II	16	2000	0.74	0.74	0.74	0.74	0.74	0.41	0.41	0.41	0.76	0.76
Scheme III	0–12	178	0.58	0.58	0.58	0.58	0.58	0.39	0.39	0.39	0.97	0.97
	13–19	1584	0.69	0.69	0.69	0.69	0.69	0.29	0.29	0.29	0.98	0.98
	20–32	238	0.82	0.82	0.82	0.82	0.82	0.17	0.17	0.17	0.99	0.99

Distributional results for the final recommended doses based on 2000 simulations of each scheme and each model in the case where the underlying dose-toxicity curves are sharply sloping near the target dose. For comparison with situation F (Tables 14 and 15), where the underlying curves are less sloped near the target dose. Increased slope improves the concentration of scheme I's dose allocation, but lower slope implies that nearby doses are closer together in terms of toxicity.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 14. Situation F: The Probability of Toxicity at the Six Dose Levels

Group 1	0.08	0.19	0.29	0.37	0.51	0.63
Group 2	0.03	0.12	0.21	0.43	0.55	0.71

lations fitting the profile with final recommended dose within α of θ_0, where the value of α is the column label. For instance, column 5 is labeled 0.08. This means column 5 lists the proportion of the simulations having a final recommended dose with a probability of toxicity within 0.08 of 0.20. It is worth noting that in each situation only one dose level comes within 0.05 of 0.20, and so the 0.05 column just gives the proportion of the simulations having a final recommended dose equal to the target dose level. It is also worth noting that in all the

Table 15. Distributional Results for Situation F

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	191	0.51	0.51	0.65	0.91	0.91	0.51	0.51	0.83	0.83	0.83
	13–19	1576	0.54	0.54	0.71	0.94	0.94	0.42	0.42	0.77	0.77	0.77
	20–32	233	0.58	0.58	0.76	0.90	0.90	0.38	0.38	0.71	0.71	0.71
S I, model 2	0–12	176	0.53	0.53	0.82	0.94	0.94	0.51	0.51	0.87	0.87	0.87
	13–19	1592	0.51	0.51	0.76	0.95	0.95	0.42	0.42	0.74	0.74	0.74
	20–32	232	0.55	0.55	0.77	0.94	0.94	0.36	0.36	0.66	0.66	0.66
Scheme II	16	2000	0.56	0.56	0.73	0.90	0.90	0.33	0.33	0.82	0.82	0.82
Scheme III	0–12	190	0.65	0.65	0.91	0.94	0.94	0.26	0.26	0.91	0.91	0.91
	13–19	1577	0.60	0.60	0.89	0.93	0.93	0.29	0.29	0.89	0.89	0.89
	20–32	233	0.63	0.63	0.88	0.94	0.94	0.25	0.25	0.88	0.88	0.88

Distributional results for the final recommended doses based on 2000 simulations of each scheme and model in the situation where the underlying dose-toxicity curves are not sharply sloping near the target dose. For comparison with situation E (Tables 12 and 13), where the underlying curves are more sharply sloped near the target dose. Increased slope improves the concentration of scheme I's dose allocation, but lower slope implies that nearby doses are closer together in terms of toxicity.

Table 16. Situation G: The Probability of Toxicity at the Six Dose Levels

Group 1	0.21	0.33	0.45	0.57	0.71	0.89
Group 2	0.07	0.20	0.32	0.45	0.57	0.76

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 17. Distributional Results for Situation G

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	214	0.70	0.70	0.70	0.70	0.97	0.59	0.59	0.59	0.97	0.97
	13–19	1525	0.71	0.71	0.71	0.71	0.97	0.53	0.53	0.53	0.95	0.95
	20–32	261	0.75	0.75	0.75	0.75	0.99	0.46	0.46	0.46	0.91	0.91
S I, model 2	0–12	182	0.63	0.63	0.63	0.63	0.95	0.46	0.46	0.46	0.98	0.98
	13–19	1559	0.70	0.70	0.70	0.70	0.97	0.42	0.42	0.42	0.96	0.96
	20–32	259	0.77	0.77	0.77	0.77	0.99	0.39	0.39	0.39	0.92	0.92
Scheme II	16	2000	0.66	0.66	0.66	0.66	0.96	0.61	0.61	0.61	0.95	0.95
Scheme III	0–12	181	0.35	0.35	0.35	0.35	0.97	0.62	0.62	0.62	1.00	1.00
	13–19	1559	0.42	0.42	0.42	0.42	0.97	0.55	0.55	0.55	1.00	1.00
	20–32	260	0.53	0.53	0.53	0.53	0.97	0.44	0.44	0.44	1.00	1.00

Distributional results for the final recommended doses based on 2000 simulations of each scheme and model when the target dose levels are near the starting dose. For comparison with situation H (Tables 18 and 19), where the target levels are further away from the starting dose. The closer the initial dose to the target, the better all schemes will run.

simulated situations, scheme I puts almost all its mass within 0.10 of the target dose level in spite of the misspecification.

Situation A (Tables 4 and 5) covered the case where there were no group differences. As one might expect, scheme III performed best in this situation but only very slightly better than scheme I, model 2. We conclude from this that, apart from the extra programming and calculation, we stand to lose little by using the more complicated model even when not really necessary. Situation B (Tables 6 and 7) covered the case where the two groups are very different. As mentioned in Section 4, for smaller samples this can cause some difficulty for scheme I, and scheme II performed best. Situation B is interesting in that, for cases like this where the two groups are very different, sharing information can slow down convergence for one of the groups. We could anticipate improvements in performance by varying the starting scheme so that the encountered toxicities of, say, group 1, leading to caution, do not lead to overcaution in group 2, at least before we have encountered any toxicity also in group 2. For such situations, scheme II

Table 18. Situation H: The Probability of Toxicity at the Six Dose Levels

Group 1	0.01	0.03	0.09	0.11	0.20	0.35
Group 2	0.01	0.01	0.05	0.09	0.13	0.19

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 19. Distributional Results for Situation H

			Group 1					Group 2				
	G1	#	0.05	0.08	0.10	0.13	0.15	0.05	0.08	0.10	0.13	0.15
S I, model 1	0–12	212	0.32	0.32	0.73	0.92	0.92	0.24	0.62	0.62	0.87	0.96
	13–19	1535	0.24	0.24	0.62	0.78	0.93	0.26	0.65	0.65	0.89	0.98
	20–32	253	0.25	0.25	0.55	0.76	0.92	0.00	0.55	0.55	0.91	0.99
S I, model 2	0–12	202	0.29	0.29	0.59	0.93	0.93	0.37	0.63	0.63	0.88	0.97
	13–19	1573	0.10	0.10	0.54	0.80	0.94	0.34	0.55	0.55	0.85	0.96
	20–32	225	0.13	0.13	0.51	0.74	0.91	0.01	0.44	0.44	0.80	0.94
Scheme II	16	2000	0.37	0.37	0.67	0.83	0.93	0.26	0.68	0.68	0.91	0.97
Scheme III	0–12	183	0.74	0.74	0.96	1.00	1.00	0.00	0.74	0.74	0.96	1.00
	13–19	1585	0.70	0.70	0.94	0.99	0.99	0.00	0.70	0.70	0.94	0.99
	20–32	232	0.64	0.64	0.91	0.99	0.99	0.00	0.64	0.64	0.91	0.99

Distributional results for the final recommended doses based on 2000 simulations of each scheme and each model when the target doses are far from the initial dose levels. For comparison with situation G (Tables 16 and 17), where the opposite is true.

does better than scheme I when we have available approximately the same number of patients per group and this number itself is not very large. Exactly how to construct procedures, for the data in hand, that may sequentially indicate which of the schemes may be the most accurate may be an area worthy of investigation. Scheme I, model 2 is simulated in situation C (Tables 8 and 9). Situation D (Tables 10 and 11) was intended to be simply an average situation. The effect of "flatness" in the dose-response curves was covered by situations E and F. Here, situation E (Tables 12 and 13) is steeper near the target dose than situation F (Tables 14 and 15), and scheme I performs correspondingly better. Of course, if the curve is really flat near the target dose, then any nearby dose could be considered acceptable. Finally, situation G (Tables 16 and 17) and situation H (Tables 18 and 19) cover the effect of the starting scheme. Because we start at the lowest level and step up only after two nontoxicities, it can take a while to reach the target dose for testing. This effect, which can be seen throughout the simulations, raises the question of optimal starting schemes. This is another area that may be worthy of additional research.

6. Further Points

The heterogeneous CRM model presented here appears to behave as do other regression models with which we may be more familiar—the linear, logistic, and proportional hazards models, for example. For perfectly balanced covariates,

we may not gain very much by modeling group effect as opposed to fitting separate and distinct models to the various subgroups. The useful gains are to be made when there are group imbalances and there may not be enough subjects in one group to carry out effective fitting. Further light could be thrown on this by research analogous to that done on the comparison of the stratified Cox model as opposed to a Cox model directly modeling the subgroup effects.

The particularity of the phase I situation is that there will be relatively few subjects and that we work with underparameterized models. The impact of these has been studied to some extent in this article, in which we can show convergence to the right dose level under misspecification of the dose-toxicity relationships as well as the relationships linking the two groups. As heterogeneity goes beyond two to several groups and even continuous covariates, large-sample theoretical results become of lesser interest and it becomes important to have tools enabling us to determine when in practice it is realistic as well as advantageous to use more complex models.

More complete inference, under misspecification of both types, also would be helpful. Obtaining ultimate convergence to the correct level for both groups is nice, but we would like more, such as meaningful confidence intervals for the probability of toxicity at the recommended dose for both groups. It should be possible to adapt some of the approaches of O'Quigley (6) to this problem, and they would already be applicable to scheme II.

Some recent work on stopping rules for the CRM exploit the convergence properties of the method to decide when enough information is available to determine that experimentation can be brought to a close (7). In the context of heterogeneity, additional problems arise. We could bring experimentation to a close when it is indicated that the probability of having converged to both levels is high. We might also want to use the information provided on a single level to close experimentation on this level and concentrate future experimentation on the other level. Such a setup would not work in the usual phase I clinical trial, where no control can be exercised on group membership. However, in other situations, for example, controlled toxicology experimentation, such methods may lead to improved efficiency in design.

Acknowledgments

The authors thank the reviewers and editors for helpful suggestions. We also thank Xavier Paoletti and Laurent Vabre for discussion and comments. The work was supported in part by grant number 1817 from the Association Pour la Recherche Contre le Cancer and by NIH grant R01CA77404 from the National Institute of Health.

References

1. O'Quigley J, Pepe M, Fisher L: Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics* 46:33–48, 1990.
2. Storer BE: Design and analysis of phase I clinical trials. *Biometrics* 45:925–937, 1989.
3. O'Quigley J, Shen LZ: Continual reassessment method: A likelihood approach. *Biometrics* 52: 163–174, 1996.
4. Shen LZ, O'Quigley J: Consistency of the continual reassessment method under model misspecification. *Biometrika* 83:395–406, 1996.
5. Huber PJ: The behavior of maximum likelihood estimates under nonstandard conditions. *Proc 6th Berkeley Symp* 1:221–233, 1967.
6. O'Quigley J: Estimating the probability of toxicity at the recommended dose following a phase I clinical trial in cancer. *Biometrics* 48:853–862, 1992.
7. O'Quigley J, Reiner E: Stopping rules for the continual reassessment method. *Biometrika* 85: 741–748, 1998.

Appendix 1

Proof of Theorem 1 (Sketch)

Denote the second-order partial derivatives of s_1 and s_2 by

$$s_{111}(x, y, a) = \frac{\partial^2 s_1}{\partial a^2}, \qquad s_{211}(x, y, a, b) = \frac{\partial^2 s_2}{\partial a^2}, \qquad s_{222}(x, y, a, b) = \frac{\partial^2 s_2}{\partial b^2},$$

$$s_{221}(x, y, a, b) = s_{212}(x, y, a, b) = \frac{\partial^2 s_2}{\partial b \partial a}.$$

A Taylor expansion of (4.5) around (a_0, b_0) leads to

$$-\begin{pmatrix} \sum_{i=1}^{n_1} s_{11}(x_i, y_i, a_0) + \sum_{i=n_1+1}^{n} s_{21}(x_i, y_i, a_0, b_0) \\ \sum_{i=n_1+1}^{n} s_{22}(x_i, y_i, a_0, b_0) \end{pmatrix} \approx A \begin{pmatrix} \hat{a}_{1n} - a_0 \\ \hat{b}_{1n} - b_0 \end{pmatrix}, \tag{A1.1}$$

where

$$A = \begin{pmatrix} \sum_{i=1}^{n_1} s_{111}(x_i, y_i, a_0) + \sum_{i=n_1+1}^{n} s_{211}(x_i, y_i, a_0, b_0) & \sum_{i=n_1+1}^{n} s_{212}(x_i, y_i, a_0, b_0) \\ \sum_{i=n_1+1}^{n} s_{221}(x_i, y_i, a_0, b_0) & \sum_{i=n_1+1}^{n} s_{222}(x_i, y_i, a_0, b_0) \end{pmatrix}.$$

Let $\hat{p}_{11}, \ldots, \hat{p}_{1m}$ be the frequency that the experiments are carried out at dose levels $d_1, \ldots, d_m$ for the first n_1 patients and $\hat{r}_{11}, \ldots, \hat{r}_{1m}$ be the corresponding frequencies of observing toxicities among those patients. Let $\hat{p}_{21}, \ldots, \hat{p}_{2m}$ and

$\hat{r}_{21}, \ldots, \hat{r}_{2m}$ be defined analogously for the last n_2 patients. Keep in mind that the first n_1 patients are from the first group and the last n_2 patients are from the second. When both n_1 and n_2 are large, we expect that $\hat{r}_{1i} \approx \psi_1(d_i, a_0)$ and $\hat{r}_{2i} \approx \psi_2(d_i, a_0, b_0)$ for $i = 1, \ldots, m$. Thus

$$\begin{aligned}\sum_{i=1}^{n_1} s_{111}(x_i, y_i, a_0) &= n_1 \sum_{i=1}^{m} \hat{p}_{1i} \times s_{111}(d_i, \hat{r}_{1i}, a_0) \\ &\approx n_1 \sum_{i=1}^{m} \hat{p}_{1i} \times s_{111}(d_i, \psi_1(d_i, a_0), a_0) \\ &= -n_1 \sum_{i=1}^{m} \hat{p}_{1i} \frac{\psi_1'^2}{\psi_1(1-\psi_1)}(d_i, a_0),\end{aligned}$$

where ψ_1' is the partial derivative of ψ_1 with respect to a. On the other hand, it can be shown that

$$\begin{aligned}\sum_{i=1}^{n_1} s_{11}^2(x_i, y_i, a_0) &= n_1 \sum_{i=1}^{m} \hat{p}_{1i} \times \left\{ \hat{r}_{1i} \left(\frac{\psi_1'}{\psi_1} \right)^2 + (1 - \hat{r}_{1i}) \left(\frac{-\psi_1'}{1-\psi_1} \right)^2 \right\} \\ &\approx n_1 \sum_{i=1}^{m} \hat{p}_{1i} \times \left\{ \psi_1(d_i, a_0) \left(\frac{\psi_1'}{\psi_1} \right)^2 + [1 - \psi_1(d_i, a_0)] \left(\frac{-\psi_1'}{1-\psi_1} \right)^2 \right\} \\ &= n_1 \sum_{i=1}^{m} \hat{p}_{1i} \frac{\psi_1'^2}{\psi_1(1-\psi_1)}(d_i, a_0) = -\sum_{i=1}^{n_1} s_{111}(x_i, y_i, a_0).\end{aligned}$$

In the same way one can establish

$$\begin{aligned}&\sum_{i=n_1+1}^{n} s_{211}(x_i, y_i, a_0, b_0) \approx -\sum_{i=n_1+1}^{n} s_{21}^2(x_i, y_i, a_0, b_0), \\ &\sum_{i=n_1+1}^{n} s_{221}(x_i, y_i, a_0, b_0) = \sum_{i=n_1+1}^{n} s_{212}(x_i, y_i, a_0, b_0) \approx -\sum_{i=n_1+1}^{n} (s_{21} s_{22})(x_i, y_i, a_0, b_0), \\ &\sum_{i=n_1+1}^{n} s_{222}(x_i, y_i, a_0, b_0) \approx -\sum_{i=n_1+1}^{n} s_{22}^2(x_i, y_i, a_0, b_0).\end{aligned}$$

From the above derivation, the matrix A is approximately

$$-\begin{pmatrix} \sum_{i=1}^{n_1} s_{11}^2 + \sum_{i=n_1+1}^{n} s_{21}^2 & \sum_{i=n_1+1}^{n} s_{21}\, s_{22} \\ \sum_{i=n_1+1}^{n} s_{21}\, s_{22} & \sum_{i=n_1+1}^{n} s_{22}^2 \end{pmatrix}.$$

Now the central limit theorem for martingales implies that the asymptotic distribution of the left-hand side of (A1.1) is normal with zero mean vector and covariance matrix B, which can be estimated by

$$\begin{pmatrix} \sum_{i=1}^{n_1} s_{11}^2 + \sum_{i=n_1+1}^{n} s_{21}^2 & \sum_{i=n_1+1}^{n} s_{21}s_{22} \\ \sum_{i=n_1+1}^{n} s_{21}s_{22} & \sum_{i=n_1+1}^{n} s_{22}^2 \end{pmatrix} \approx -A. \tag{A1.2}$$

Solving (A1.1) yields

$$\begin{pmatrix} \hat{a}_{1n} - a_0 \\ \hat{b}_{1n} - b_0 \end{pmatrix} = -A^{-1} \begin{pmatrix} \sum_{i=1}^{n_1} s_{11} + \sum_{i=n_1+1}^{n} s_{21} \\ \sum_{i=n_1+1}^{n} s_{22} \end{pmatrix}. \tag{A1.3}$$

Hence the asymptotic distribution of $(\hat{a}_{1n}, \hat{b}_{1n})$ is normal with mean vector (a_0, b_0) and with covariance matrix ABA'^{-1}, which is approximately A^{-1}. Simple calculation of A^{-1} reveals that its diagonal elements are

$$\frac{\Sigma_{i=n_1+1}^{n} s_{22}^2}{\Sigma_{i=n_1+1}^{n} s_{22}^2 (\Sigma_{i=1}^{n_1} s_{11}^2 + \Sigma_{i=n_1+1}^{n} s_{21}^2) - (\Sigma_{i=n_1+1}^{n} s_{21}s_{22})^2} \tag{A1.4}$$

and

$$\frac{\Sigma_{i=1}^{n_1} s_{11}^2 + \Sigma_{i=n_1+1}^{n} s_{21}^2}{\Sigma_{i=n_1+1}^{n} s_{22}^2 (\Sigma_{i=1}^{n_1} s_{11}^2 + \Sigma_{i=n_1+1}^{n} s_{21}^2) - (\Sigma_{i=n_1+1}^{n} s_{21}s_{22})^2}. \tag{A1.5}$$

Therefore the asymptotic variances for $\hat{a}_{1n}$ and $\hat{b}_{1n}$ can be approximated by (A1.4) and (A1.5), respectively.

Appendix 2

Consistency and Asymptotic Normality for the Two-Sample CRM Under Model Misspecification

Let R_1 and R_2 be the dose-response curves for the two groups. Let $D = \{d_1, \ldots, d_k\}$ be the collection of doses at which the two groups are to be tested. We assume there exists an $\varepsilon > 0$ such that

$$\varepsilon \le R_j(\mathrm{d}_i) \le 1 - \varepsilon, \qquad \mathrm{j} = 1, 2, \quad \mathrm{i} = 1, \ldots, \mathrm{k}. \tag{A2.1}$$

In order to fit the model, we require that the initial phase be carried out using some other method (up-and-down, isotonic Robbins-Monro, etc.) until we have observed heterogeneity in the responses for both groups. Clearly, we would like this initial phase of the trial to end in a finite (we hope, small) number of steps. With the aim of ensuring this will happen, we make the following definitions. Take g_j to be the indicator that subject j is in group 1 (i.e., $g_j = I_j - 1$), and let y_j be the indicator of toxic response for subject j. We will need the following minor assumption.

Assumption: For some $0 < p < 1$ we have

$$\lim_{k\to\infty} k^{-1} \sum_{j=1}^{k} g_j = p \quad \text{almost surely.} \tag{A2.2}$$

Assumption (A2.2) ensures that there is an $n < \infty$ such that for all $j > n$ both groups are represented in the sample and hence that no group completely dominates the sample. Now, if we define

$$n_0 = \inf\left\{k : 0 < \sum_{j=1}^{k} g_j y_j < \sum_{j=1}^{k} g_j,\ 0 < \sum_{j=1}^{k} (1-g_j) y_j < \sum_{j=1}^{k} (1-g_j)\right\}$$

to be the first time we observe heterogeneity in the responses for both groups, then assumptions (A2.1) and (A2.2) allow us to make the following claim.

Proposition 1: Supposing (A2.1) and (A2.2) hold; then $P(n_0 < \infty) = 1$.

Proof: Suppose up to time k we have observed $(x_1, y_1, g_1), \ldots, (x_k, y_k, g_k)$, where x_j is the dose level, y_j the indicator of toxicity, and g_j the group indicator for subject j. We define

$$T_k^1 = \sum_{j=1}^{k} g_j [y_j - R_1(x_j)]$$

and

$$T_k^2 = \sum_{j=1}^{k} (1-g_j)[y_j - R_2(x_j)].$$

Clearly, T_k^1 and T_k^2 are bounded martingales and, by the martingale convergence theorem, $k^{-1}T_k^1$ and $k^{-1}T_k^2$ tend to zero, almost surely. In fact, by assumption (A2.2), $(\Sigma\, g_j)^{-1}T_k^1$ and $(\Sigma(1-g_j))^{-1}T_k^2$ must tend to zero, almost surely. Furthermore, condition (A2.1) implies

$$0 < \varepsilon \sum_{j=1}^{k} g_j \le \sum_{j=1}^{k} g_j R_1(x_j) \le (1-\varepsilon) \sum_{j=1}^{k} g_j$$

and

$$0 < \varepsilon \sum_{j=1}^{k} (1-g_j) \le \sum_{j=1}^{k} (1-g_j) R_2(x_j) \le (1-\varepsilon) \sum_{j=1}^{k} (1-g_j),$$

which together with (A2.2) says that, for large enough k,

$$0 < (\varepsilon/2) \sum_{j=1}^{k} g_j \le \sum_{j=1}^{k} g_j y_j \le (1-\varepsilon/2) \sum_{j=1}^{k} g_j < \sum_{j=1}^{k} g_j$$

and

$$0 < (\varepsilon/2) \sum_{j=1}^{k} (1-g_j) \le \sum_{j=1}^{k} (1-g_j) y_j \le (1-\varepsilon/2) \sum_{j=1}^{k} (1-g_j) < \sum_{j=1}^{k} (1-g_j),$$

as desired. Hence, $P(n_0 < \infty) = 1$.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Now we have to worry about the model. Suppose we specify a two-parameter model as follows. We model $R_1(\cdot)$ as $\psi_1(\cdot, a)$ and $R_2(\cdot)$ as $\psi_2(\cdot, a, b) = \mu(\psi_1(\cdot, a), b)$, where μ is a one-parameter link function; different values of μ allow for different tests of heterogeneity between the two groups. We make the following assumptions about the model.

Model 1

1. For each (a, b), the functions $\psi_1(\cdot, a)$ and $\psi_2(\cdot, a, b)$ are smooth and strictly increasing in x, the dose level.
2. The function $\psi_1(x, \cdot)$ is continuous and strictly increasing in a in the same direction for all x, and the function $\psi_2(x, a, \cdot)$ is continuous and strictly increasing in b in the same direction for each a and each x.

Model 2

1. The partial derivatives of ψ_1 and ψ_2 with respect to a (written $\psi_1^{(a)}$ and $\psi_2^{(a)}$, respectively) and the partial derivative of ψ_2 with respect to b ($\psi_2^{(b)}$) are continuous in a, b, and x.
2. For all $n > n_0$, the function $L^{(n)}(a, b)$ given by

$$n^{-1} \sum_{i=1}^{n} g_i L_1(y_i, x_i, a) + (1 - g_i) L_2(y_i, x_i, a, b)$$

has a unique root, $(\hat{a}_n, \hat{b}_n)$, where the functions L_1 and L_2 are given as follows

$$L_1(y, x, a) = \begin{bmatrix} y \dfrac{\psi_1^{(a)}}{\psi_1}(x, a) + (1 - y) \dfrac{-\psi_1^{(a)}}{1 - \psi_1}(x, a) \\ 0 \end{bmatrix}$$

and

$$L_2(y, x, a) = \begin{bmatrix} y \dfrac{\psi_2^{(a)}}{\psi_2}(x, a, b) + (1 - y) \dfrac{-\psi_2^{(a)}}{1 - \psi_2}(x, a, b) \\ y \dfrac{\psi_2^{(b)}}{\psi_2}(x, a, b) + (1 - y) \dfrac{-\psi_2^{(b)}}{1 - \psi_2}(x, a, b) \end{bmatrix}.$$

The assumptions under model 1 and (A.1.1) enable us to restrict the parameters (a, b) to lie in a compact convex set $K \subset \mathrm{IR}^2$ by taking K to be the smallest closed rectangle containing all the points satisfying these three assumptions. The assumption of model 2 forces the model to have continuous partial derivatives and ensures the existence of a unique root of the likelihood equation for large enough n.

Suppose we would like to know the doses d_0^1 and d_0^2 such that $R_1(d_0^1) = \theta_0^1$ and $R_2(d_0^2) = \theta_0^2$. Because it is unlikely that d_0^1 or d_0^2 are members of D, we are interested in finding the "nearest" members x_0^1 and x_0^2 of D in the sense that x_0^j minimizes $\|\theta_0^j - R_j(d)\|$ over D (or the subset of D over which the group is being tested). Here, $\|\cdot\|$ can be taken to be an arbitrary metric. We need the following assumptions about the model.

Model 3

For every pair of doses (d_i, d_j) in $D \times D$ (or the corresponding cross product of subsets), there is a unique (a_{ij}, b_{ij}) in K such that $\psi_1(d_i, a_{ij}) = R_1(d_i)$ and $\psi_2(d_j, a_{ij}, b_{ij}) = R_2(d_j)$.

This assumption ensures that the model can be fit regardless of which dose levels are "nearest" in the sense above. Of course, the model is unlikely to hold exactly, and so it is of primary importance to deal with the possibility of model misspecification. For the purpose of characterizing the difference between the model and the underlying dose-response curves, we introduce the following set. Suppose we have a selection procedure ρ such that, given a pair (a, b) in K, ρ selects a value (x_{ab}^1, x_{ab}^2) for use as the next dose pair.* Let S be the set of (a, b) such that ρ selects (x_0^1, x_0^2) for use in the next stage of the trial. (The set S is very similar to the "controlling" set defined in (4).

In order to ensure consistency in spite of likely misspecification, the following assumption about S is needed.

Assumption: For every $0 < p < 1$, the function $\tilde{L}(a, b)$ given by

$$\sum_{d \in D_1} pL_1(R_1(d), d, a) + \sum_{d \in D_2} (1-p)L_2(R_2(d), d, a, b)$$

has a unique root $(\tilde{a}, \tilde{b})$ in S, where D_j is the subset of D at which members of group j are tested.†

Now that all the assumptions have been delineated, we are in a position to prove the following result.

Proposition 2: Assume conditions (A2.1), (A2.2), models 1–3 and the assumption above are satisfied. Let $(\hat{a}_n, \hat{b}_n)$ be the maximum likelihood estimate of the parameters (a_0, b_0), and (x_{n+1}^1, x_{n+1}^2) be the recommended dose levels for the next stage of the trial. Then almost surely, $(\hat{a}_n, \hat{b}_n)$ tends to (a_0, b_0) and (x_{n+1}^1, x_{n+1}^2) tends to (x_0^1, x_0^2). Furthermore,

$$\sqrt{n}\begin{bmatrix} \hat{a}_n - a_0 \\ \hat{b}_n - b_0 \end{bmatrix} \to Z$$

*For example, ρ could choose (x_{ab}^1, x_{ab}^2) so that $|\theta_0^1 - \psi_1(x_{ab}^1, a)| + |\theta_0^2 - \psi_2(x_{ab}^2, a, b)|$ is minimized over $D \times D$.

†This is much the same as (D.3) from Shen and O'Quigley (4).

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

in distribution, where Z is a $N(0, \Gamma)$ random variable, and

$$\Gamma = \begin{bmatrix} \dfrac{R_1(1-R_1)}{p(\psi_1^{(a)})^2} & 0 \\ 0 & \dfrac{R_2(1-R_2)}{(1-p)(\psi_2^{(b)})^2} \end{bmatrix} + \dfrac{R_1(1-R_1)}{p(\psi_1^{(a)})^2} \begin{bmatrix} 0 & -\left(\dfrac{\psi_2^{(a)}}{\psi_2^{(b)}}\right) \\ -\left(\dfrac{\psi_2^{(a)}}{\psi_2^{(b)}}\right) & \left(\dfrac{\psi_2^{(a)}}{\psi_2^{(b)}}\right)^2 \end{bmatrix}$$

evaluated at (x_0^1, x_0^2, a_0, b_0).

Proof: Note that for each dose level x_j and each fixed y in (0, 1) the functions $L_1(y, x_j, a)$ and $L_2(y, x_j, a, b)$ are uniformly continuous on the compact set K. Thus, for any $\varepsilon > 0$, for each x_j in D, and each y in (0, 1), there exists a finite set of points $K_\varepsilon \subset K$ such that for any (a, b) in K, there exists a point $(a_\varepsilon, b_\varepsilon)$ in K_ε such that

$$|L_1(y, x_j, a) - L_1(y, x_j, a_\varepsilon)| < \varepsilon$$

and

$$|L_2(y, x_j, a, b) - L_2(y, x_j, a_\varepsilon, b_\varepsilon)| < \varepsilon.$$

Now because there are only m possible dose levels, the mesh points K_ε may be chosen in such a way that these inequalities hold simultaneously for all the x_j in D. This shows we can approximate the likelihood uniformly over K by its values on a finite set of points.

The next task is to show that L and $\tilde{L}$ are close, where

$$\tilde{L}(a, b) = n^{-1} \sum_{i=1}^{n} g_i L_1(R_1(x_i), x_i, a) + (1 - g_i) L_2(R_2(x_i), x_i, a, b).$$

First, we approximate L by its values on the mesh set K_ε:

$$\begin{aligned} \sup_{(a,b)\in K} |L(a, b) - L(a_\varepsilon, b_\varepsilon)| &\le n^{-1} \sum_{i=1}^{n} g_i |L_1(y_i, x_i, a) - L_1(y_i, x_i, a_\varepsilon)| \\ &\quad + (1 - g_i) |L_2(y_i, x_i, a, b) - L_2(y_i, x_i, a_\varepsilon, b_\varepsilon)| \\ &< n^{-1} \sum_{i=1}^{n} g_i \varepsilon + (1 - g_i)\varepsilon = \varepsilon. \end{aligned}$$

Then we approximate L by $\tilde{L}$ on the mesh set:

$$\begin{aligned} L(a_\varepsilon, b_\varepsilon) - \tilde{L}(a_\varepsilon, b_\varepsilon) = n^{-1} \sum_{i=1}^{n} & g_i L_1(y_i - R_1(x_i), x_i, a_\varepsilon) \\ & + (1 - g_i) L_2(y_i - R_2(x_i), x_i, a_\varepsilon, b_\varepsilon). \end{aligned}$$

Now for any fixed $(a_\varepsilon, b_\varepsilon)$ this difference forms a bounded martingale and converges to zero as the sample size n increases.

Finally, we just show that $\tilde{L}$ does not vary much off the mesh set:

$$\sup_{(a,b)\in K} |\tilde{L}(a, b) - \tilde{L}(a_\varepsilon, b_\varepsilon)| < n^{-1} \sum_{i=1}^{n} g_i\varepsilon + (1 - g_i)\varepsilon = \varepsilon,$$

(again) by the uniform continuity of L_1 and L_2.

So we have shown that $L - \tilde{L}$ converges pointwise to zero, and because for each n (large enough) both L and $\tilde{L}$ have unique roots, our parameter estimates (the roots of L) must converge to the roots of $\tilde{L}$. As a result, we can replace individual responses y_i with population averages $R_j(x_i)$ for asymptotic computations.

Rewriting $\tilde{L}$ as a sum over doses (instead of subjects) we see that the assumption above implies that our estimated parameters must eventually be in the set S. Therefore, almost surely, we will select the correct dose and continue to test there. This proves strong consistency for the recommended dose level. That is, we have shown $(x_n^1, x_n^2) \to (x_0^1, x_0^2)$ almost surely.

Now, because there are only finitely many dose levels, the almost sure consistency of the recommended dose implies that for all but a finite number of subjects we will be testing at (x_0^1, x_0^2), and thus, for the purpose of asymptotic computations, we may as well assume we are always testing at this dose. If we are always testing at (x_0^1, x_0^2), the asymptotic distribution of $(\hat{a}_n, \hat{b}_n)$ will be the same as that of the solution to

$$n^{-1} \sum_{i=1}^{n} pL_1(y_i, x_0^1, a) + (1 - p)L_2(y_i, x_0^2, a, b).$$

Because $(x_0^1, y_1), \ldots, (x_0^1, y_n)$ and $(x_0^2, y_1), \ldots, (x_0^2, y_n)$ are (allowing for the abuse of notation with respect to the y_j) independent and identically distributed samples, the asymptotic normality of the parameter estimates follows from the standard approach for maximum likelihood estimates [see Huber (5)]. Computation of the asymptotic variance matrix is also straightforward and based on (5).

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016