Information Recovery in a Study With Surrogate Endpoints

Song Xi CHEN, Denis H. Y. LEUNG, and Jing QIN

Recently, there has been a lot of interest in statistical methods for analyzing data with surrogate endpoints. In this article, we consider parameter estimation from a model that relates a variable Y to a set of covariates, X, in the presence of a surrogate, S. We assume that the data are made up of two random samples from the population, a validation set where (Y, X, S) are observed on every subject and a nonvalidation set where only (X, S) are measured. We show how information from the nonvalidation set can be incorporated to improve upon estimation of a parameter β using the validation data only. The method we suggest does not require knowledge on the joint distribution between (Y, S), given X. It is based on a two-sample empirical likelihood that simultaneously combines the estimating equations from the validation set and the nonvalidation set. The proposed nonparametric likelihood formulation brings a few attractive features to the inference in β. First, the maximum empirical likelihood estimate is more efficient than that using only the validation sample. Second, confidence regions can be readily constructed without the need to estimate the variance-covariance matrix. Finally, the coverage of the confidence regions can be further improved by an empirical Bartlett correction based on the bootstrap. We show that the method gives favorable results in simulation studies.

KEY WORDS: Auxiliary outcome; Bartlett correction; Bootstrap; Confidence regions; Empirical likelihood; Estimating equations; Surrogate endpoint.

1. INTRODUCTION

The past decade has witnessed a surge of interest in statistical methods for studying surrogate endpoints or auxiliary outcome data. This is partly due to the increasing interest of the medical community, in particular, in conducting studies using surrogate endpoints. A landmark example is the use of CD4 counts in acquired immunodeficiency syndrome (AIDS) research where it is postulated that the CD4 count has a positive correlation with subsequent survival in human immunodeficiency virus (HIV) patients. Other examples of surrogate endpoint studies can be found in Ellenberg and Hamilton (1989), Wittes, Lakatos, and Probstfield (1989), Fleming and Demets (1996), and the references therein.

A surrogate endpoint study can be thought of as follows. Data are made up of two parts. The first part, called the validation sample (V), consists of observations with information on the endpoint of interest (Y) along with information on a surrogate (S) and also some covariate information (X). The second part, called the nonvalidation sample ($\bar{V}$), is made up of observations with only information on S and X. The interest is to estimate some population parameter, β, that explains the relationship between Y and X. The primary statistical problem in a surrogate endpoint study is how to make efficient use of the surrogate endpoint to recover information about β from the nonvalidation data.

The validation sample and the nonvalidation sample can be considered to be random samples from the same population. This is appropriate, for example, when the two samples are data collected in a cohort study or in a survey study, which is designed to measure the surrogate on all subjects and to measure the true endpoint on a subset of the cohort. Venkatraman and Begg (1999) described a study on the influence of growth hormones in girls with Turner syndrome, where this assumption is valid. In this situation β can be estimated consistently using data from the validation sample only. However, the nonvalidation sample also contains information on β, and it is preferable to use the nonvalidation sample to enhance the estimation precision of β. In earlier work Pepe (1992) approached this problem semiparametrically by considering the likelihood

$$L(\gamma, \beta) = \prod_{i \in V} P(y_i|x_i, \beta) P(s_i|x_i, y_i, \gamma) \prod_{i \in \bar{V}} P(s_i|x_i, \gamma, \beta),$$

where it is assumed that the conditional probability of Y, given X, is parameterized by β, whereas the conditional probability of Y, S, given X, is parameterized by parameters γ and β. The augmented maximum likelihood estimate of β is obtained under the usual regularity conditions as long as the separate components in the likelihood are known or can be estimated consistently. The first component is merely that of the likelihood due to the validation data. For the second component Pepe (1992) decomposed $P(s_i|x_i, \gamma, \beta)$ as $\int P(y|x_i, \beta) P(s_i|x_i, y)\, dy$ and then estimated $P(s_i|x_i, y)$ by kernel smoothing. Fleming, Prentice, Pepe, and Glidden (1994) extended Pepe's work to survival-type data. They considered Pepe's "augmented" likelihood as well as an "augmented" score approach. The augmented score approach uses a similar idea as the augmented likelihood approach but instead of likelihood, score functions from the validation and nonvalidation samples are calculated separately and then combined to make inference about the unknown parameters. Rotnitzky and Robins (1995) considered a method using weighted estimating equations. Let $\hat{\eta}$ be the maximum likelihood estimate of $\eta = (\eta_1, \eta_2)$ in the model for the conditional probability $P(i \in V|s_i, x_i) = \rho_i(\eta)$ of selection into the validation sample

$$\text{logit}\{\rho_i(\eta)\} = \eta_1 + \eta_2' h(s_i, x_i),$$

Song Xi Chen is Associate Professor, Department of Statistics and Applied Probability, National University of Singapore, Singapore and Department of Statistics, Iowa State University, Ames, IA 50011-1210. Denis Leung is Associate Professor, School of Economics and Social Sciences, Singapore Management University, Singapore (E-mail: *denisleung@smu.edu.sg*). Jing Qin is Assistant Member, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10021. All authors contributed equally to this article. This work is based on an earlier draft by Denis Leung and Jing Qin when the former was at Memorial Sloan-Kettering Cancer Center. We thank the referees and the associate editor for their insightful comments that have led to a substantially improved article. Song Chen's research was supported by a National University of Singapore research grant (R-115-000-018-112). Denis Leung's research was partially funded by the Wharton-SMU Research Center at Singapore Management University (grant 01-C208-SMU-001).

Journal of the American Statistical Association
December 2003, Vol. 98, No. 464, Theory and Methods
DOI 10.1198/016214503000000972

where $h(s_i, x_i)$ is an arbitrary function of S and X. They showed that one can always improve upon inference based on solving a set of estimating equations that uses the validation data only, namely,

$$\sum_{i \in V} U(y_i|x_i, \beta) = 0.$$

by considering estimators that solve the weighted estimating equations

$$\sum_{i \in V} \frac{1}{\rho_i(\hat{\eta})} U(y_i|x_i, \beta) = 0.$$

When the validation sample is a random sample from the population, $\eta_2' = 0$. Therefore, the model is guaranteed to be correct. Nevertheless, as pointed out by the authors, modeling $\rho_i(\eta)$ is still useful in this case because it allows information about the parameters of interest to be extracted from the nonvalidation sample.

Chen and Chen (2000) considered a method based on the regression estimate. The method allows both the covariates and the endpoint to be measured with error. It is assumed that information relating the primary endpoint and the covariates is summarized in a set of estimating equations, involving the unknown parameter, say β, that is of interest. There is a second set of estimating equations that summarizes the relationship between the surrogate endpoint and the surrogate covariates in terms of another set of parameters, η. The two sets of equations are individually solved based on the validation sample. If $\hat{\beta}$ and $\hat{\eta}$ represent the solutions to the equations, the estimate of β they proposed is the conditional mean of $\hat{\beta}$, given $\hat{\eta}$. Their method is asymptotically equivalent to the method of Rotnitzky and Robins (1995).

In this article, the inference about β is obtained using a two-sample empirical likelihood (EL) based on estimating equations from both the validation and the nonvalidation samples. EL is a computer-intensive nonparametric method of inference introduced by Owen (1988) as an alternative to the bootstrap. Instead of resampling with equal probability weights for all data values, as in the bootstrap, EL chooses the weights by profiling a multinomial likelihood under a set of constraints. The constraints reflect the characteristics of the quantity of interest. Empirical likelihood has been shown in a wide range of situations to have properties analogous to a real likelihood, for example, Wilks' theorem and Bartlett's correction; see Hall and La Scala (1990), Owen (1991), Chen (1993), Qin and Lawless (1994), Li (1995a,b), Kitamura (1997), and the references given in Owen (2001), which provides a comprehensive overview of EL.

In the current study the EL assigns weights that reflect the information about β contained in the nonvalidation data. The efficiency of the maximum EL estimator is asymptotically equivalent to those of Rotnizky and Robins (1995) and Chen and Chen (2000). However, the method of Rotnizky and Robins (1995) uses a logistic regression to extract information from the surrogate. In practice, especially in small or medium samples, it is not uncommon for the logistic regression estimates to be infinite; see Albert and Anderson (1984). On the other hand, the method of Chen and Chen (2000) requires evaluating the conditional mean of $\hat{\beta}$ under asymptotic normality, which may not be simple in some cases. For example, if information from the surrogate could be summarized in terms of its median, then the conditional mean would involve the density at the median, which may not be easily estimated.

An attractive feature of the proposed method is that the EL provides a unified framework for producing both point estimates and confidence regions for β. The confidence regions are obtained from a Wilks' theorem for the log EL ratio. The regions, similar to other EL-based regions, have natural shapes and orientations and, more important, are free of any secondary procedures. The latter property is due to EL's ability to Studentize internally. We also show that the coverage of the confidence regions can be improved by implementing an empirical Bartlett correction, which reduces the difference between the means of the log EL ratio and the limiting χ^2 distribution. In contrast, confidence regions proposed in Rotnitzky and Robins (1995) and Chen and Chen (2000) are based on asymptotic normality. The shapes and orientations have to be subjectively chosen and the asymptotic covariance matrix has to be estimated. It is possible that the estimated covariance matrix is not invertible. Our simulation indicates that the proposed EL confidence regions have better coverage than those of Rotnitzky and Robins (1995) and Chen and Chen (2000) in nearly all the cases considered, and particularly so when the percentage of nonvalidation data is high.

The rest of this article is organized as follows. The method and its large-sample results are given in Section 2. Some finite-sample simulation results are reported in Section 3. An example illustrating how the method works is given in Section 4. The results of this article are discussed in Section 5. Proofs are given in the Appendix.

2. METHOD AND MAIN RESULTS

We assume that both Y and S are continuous, and we let X be a vector representing a set of covariates (such as treatment or prognostic factors). We assume the data consist of a total of N observations of which n come from the validation sample and $m = N - n$ are from the nonvalidation sample. The data thus comprise (y_i, s_i, x_i), $i = 1, \ldots n$, from the validation sample and (s_{n+j}, x_{n+j}), $j = 1, \ldots, m$, from the nonvalidation sample. We assume that $n/N \to \rho$ for a constant $\rho \in (0, 1)$ as $N \to \infty$.

Suppose the relationship between Y and X is summarized in an unknown p-dimensional parameter β via a p-dimensional zero-mean estimation function $U(Y, X, \beta)$. The interest is in drawing inference on β. The auxiliary information is given by the surrogate, S, and the covariate, X, whose relationship can be summarized in another r-dimensional zero-mean estimating equation $h(S, X, \gamma)$ via an r-dimensional unknown parameter γ.

2.1 Proposed Method and Large-Sample Properties

Let $p_1, \ldots, p_n$ be nonnegative weights allocated to the validation sample $\{(y_i, s_i, x_i)\}_{i=1}^n$ and let $q_1, \ldots, q_m$ be nonnegative weights allocated to the nonvalidation sample $\{(s_{n+j}, x_{n+j})\}_{j=1}^m$. The two-sample EL for the parameter (β, γ) is

$$L(\beta, \gamma) = \max \prod_{i=1}^{n} p_i \prod_{j=1}^{m} q_j \tag{1}$$

subject to

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{j=1}^{m} q_j = 1, \quad \sum_{i=1}^{n} p_i \big(U^T(y_i, x_i, \beta), h^T(s_i, x_i, \gamma)\big)^T = 0, \quad \sum_{j=1}^{m} q_j h(s_{n+j}, x_{n+j}, \gamma) = 0. \tag{2}$$

To simplify the notation, let $U_i(\beta) = U(y_i, x_i, \beta)$, $U_i'(\beta) = \partial U_i(\beta)/\partial\beta$, $h_k(\gamma) = h(s_k, x_k, \gamma)$, and $h_k'(\gamma) = \partial h_k(\gamma)/\partial\gamma$ for $i = 1, \ldots, n$ and $k = 1, \ldots, n+m$. By introducing Lagrange multipliers and following standard derivations in EL, the optimal p_i's and q_j's for given (β, γ) are

$$p_i = \frac{1}{n}\frac{1}{1 + \lambda_1^T U_i(\beta) + \lambda_2^T h_i(\gamma)} \quad \text{for } i = 1, \ldots, n \tag{3}$$

and

$$q_j = \frac{1}{m}\frac{1}{1 + \lambda_3^T h_{n+j}(\gamma)} \quad \text{for } j = 1, \ldots, m, \tag{4}$$

where the Lagrange multipliers λ_j, $j = 1, 2$, and 3, satisfy the following equations:

$$\sum_{i=1}^{n} \frac{U_i(\beta)}{1 + \lambda_1^T U_i(\beta) + \lambda_2^T h_i(\gamma)} = 0, \tag{5}$$

$$\sum_{i=1}^{n} \frac{h_i(\gamma)}{1 + \lambda_1^T U_i(\beta) + \lambda_2^T h_i(\gamma)} = 0, \tag{6}$$

$$\sum_{j=1}^{m} \frac{h_{n+j}(\gamma)}{1 + \lambda_3^T h_{n+j}(\gamma)} = 0. \tag{7}$$

From (3) and (4) we have the negative log EL

$$\begin{aligned}\ell(\beta, \gamma) &= -2\log\{L(\beta, \gamma)\} \\ &= 2\sum_{i=1}^{n} \log\{1 + \lambda_1^T U_i(\beta) + \lambda_2^T h_i(\gamma)\} \\ &\quad + 2\sum_{j=1}^{m} \log\{1 + \lambda_3^T h_{n+j}(\gamma)\} \\ &\quad - 2n\log(n) - 2m\log(m).\end{aligned} \tag{8}$$

Differentiating (8) with respect to β and γ and using (5)–(7) lead to

$$\lambda_1^T \sum_{i=1}^{n} \frac{U_i'(\beta)}{1 + \lambda_1^T U_i(\beta) + \lambda_2^T h_i(\gamma)} = 0, \tag{9}$$

$$\lambda_2^T \sum_{i=1}^{n} \frac{h_i'(\gamma)}{1 + \lambda_1^T U_i(\beta) + \lambda_2^T h_i(\gamma)} + \lambda_3^T \sum_{j=1}^{m} \frac{h_{n+j}'(\gamma)}{1 + \lambda_3^T h_{n+j}(\gamma)} = 0. \tag{10}$$

Let $(\hat\beta, \hat\gamma, \hat\lambda_1, \hat\lambda_2, \hat\lambda_3)$ be the solutions to (5)–(7) and (9), and (10). Then $(\hat\beta, \hat\gamma)$ is the maximum EL estimate of (β, γ). Let β_* and γ_* be the true parameter values of β and γ. To further simplify the notation, let $V(U) = E\{U_i(\beta_*)U_i^T(\beta_*)\}$, $V(h) = E\{h_i(\gamma_*)h_i^T(\gamma_*)\}$, $E(Uh^T) = E\{U_i(\beta_*)h_i^T(\gamma_*)\}$, $U' = U'(\beta_*)$, and $h' = h'(\gamma_*)$. The following theorem establishes the asymptotic normality of $(\hat\beta, \hat\gamma)$.

Theorem 1. Under the conditions given in (A.1), as $N \to \infty$,

$$\sqrt{n}(\hat\beta - \beta_*) \xrightarrow{d} N\big(0, E(U'^T)^{-1}\{V(U) - (1-\rho)E(Uh^T) \times V^{-1}(h)E(hU^T)\}E(U')^{-1}\big). \tag{11}$$

Remark 1. Note that $E(U'^T)^{-1}V(U)E(U')^{-1}$ is the asymptotic variance of the estimate of β based on the validation sample only. The theorem indicates that, asymptotically, using surrogate endpoints leads to more precise parameter estimates. The gain in efficiency is given by the term $(1-\rho)E(U'^T)^{-1}E(Uh^T)V^{-1}(h)E(hU^T)E(U')^{-1}$, which suggests that the gain is higher when ρ is smaller. In other words, when the nonvalidation sample is large, the gain is high. The expression $E(Uh^T)V^{-1}(h)E(hU^T)$ represents a "correlation" between the score functions U and h. Therefore, if the information contained in h is highly correlated to that contained in U, then the gain is higher.

Remark 2. Even when the surrogate S and the true endpoint Y are orthogonal, our method can still produce an asymptotically unbiased estimate with the same efficiency as the estimation based on the validation sample only. Hence, asymptotically, there is no loss in incorporating surrogate information. This can be observed by noting in (11) that if Y and S are orthogonal $E[U(y,x)h(s,x)|x] = E[U(y,x)|x]E[h(s,x)|x] = 0$. The same can be said when X and Y are orthogonal. This is because $E[U(y,x)h(s,x)|x] = E[U(y,x)\{Eh(s,x)|x\}]$, which means that $E[U(y,x)h(s,x)] = E\{U(y,x)Eh(s,x)|x]\} = 0$.

Remark 3. We note that the asymptotic distribution of our method is the same as those given by Rotnitzky and Robins (1995) and Chen and Chen (2000), when the same h is applied to all three methods. Therefore, all three methods have the same asymptotic efficiency as far as point estimation is concerned. However, as shown in the simulation studies later, there are substantial differences in the confidence regions produced by the three methods. Using a Bartlett correction, the coverage of the EL confidence regions can be substantially more accurate than those of the other methods.

Let $\tilde\gamma = \tilde\gamma(\beta_*) = \arg\min_\gamma \ell(\beta_*, \gamma)$ be the maximum likelihood estimate of γ when β is fixed at the true value β_*. The log EL ratio for β evaluated at β_* is

$$\ell(\beta_*) = \ell(\beta_*, \tilde\gamma) - \ell(\hat\beta, \hat\gamma).$$

The following is a nonparametric version of the Wilks theorem in conventional parametric likelihood theory.

Theorem 2. Under the conditions given in (A.1), $\ell(\beta_*) \xrightarrow{d} \chi_p^2$ as $N \to \infty$.

The theorem can be used to construct confidence regions for β. Let c_α be the upper α percentile of χ^2_p for $\alpha \in (0, 1)$. Then an α-level confidence region is $\mathrm{CR}_\alpha = \{\beta | \ell(\beta) \le c_\alpha\}$. Theorem 2 assures that CR_α attains the nominal coverage level α asymptotically.

In parametric likelihood theory it is known that a simple mean adjustment to the parametric likelihood ratio reduces the coverage error by one order of magnitude. This is the Bartlett correction for parametric likelihood. For the EL the Bartlett correction has been established for a range of situations; see DiCiccio, Hall, and Romano (1991) for smooth functions of means and Chen (1993) for linear regression. The idea of the Bartlett correction is to adjust the mean of the EL ratio $\ell(\beta_*)$ such that its difference from the mean of the limiting χ^2_p becomes a smaller order than n^{-1}. It turns out that this simple mean adjustment improves the approximation of the likelihood ratio to the χ^2_p by one order of magnitude. As a formal Bartlett correction for the current two-sample EL defined by generalized estimating equations with the nuisance parameter γ requires very lengthy derivations of the cumulants, we propose instead an empirical Bartlett correction based on the bootstrap. Recently, Chen and Cui (2002) formally established a Bartlett correction for one-sample EL defined by generalized estimating equations with nuisance parameters, which is a case closely related to the current situation. This together with the Bartlett correction for the two-sample EL established by Jing (1995) leads us to conjecture that the Bartlett correction would be valid for the current situation as well. This conjecture is well supported by the simulation results in Section 3. The proposed empirical Bartlett correction is the following:

Step 1. Generate bootstrap resamples of sizes n and m by sampling with replacement from the validation sample $\{(y_i, s_i, x_i)\}_{i=1}^n$ and the nonvalidation sample $\{(s_{n+j}, x_{n+j})\}_{j=1}^m$, respectively; compute the EL ratio at $\hat\beta$ based on the resamples and denote it as $\ell^*(\hat\beta)$.

Step 2. Repeat Step 1 B times to obtain $\ell^{*1}(\hat\beta), \ldots, \ell^{*B}(\hat\beta)$ and $B^{-1}\sum_{b=1}^B \ell^{b*}(\hat\beta)$, which is the bootstrap estimate of $E\{\ell(\hat\beta)\}$.

The empirical Bartlett factor η is $p^{-1}B^{-1}\sum_{b=1}^B \ell^{b*}(\hat\beta)$ and the Bartlett-corrected EL confidence region is $\mathrm{CR}_{\alpha,\mathrm{BC}} = \{\beta | \ell(\beta) \le c_\alpha \times \eta\}$. As the EL ratio tends to take larger mean values than p, the Bartlett correction shifts the body of the distribution of $\ell(\beta)$ to the left and makes it closer to the χ^2_p distribution.

2.2 The Choice of $h(s_i, x_i, \gamma)$

Although a zero-mean $h(s_i, x_i, \gamma)$ is necessary to ensure the consistency of the resulting estimates, the efficiencies of the resulting estimates are dependent on the choice of h. Rotnitzky and Robins (1995) showed that the optimal h, in terms of minimizing the asymptotic variance of the estimate of β, is given by $\rho^{-1}E(U(y|x, \beta)|s, x)$. This choice, however, is usually not attainable because it depends on the unknown probability distribution generating the data Bang and Tsiatis (2000); Chen and Chen (2000). The choice of h in the simulations will be discussed further in Section 3.

3. SIMULATIONS

In this section we report results from two broad simulation studies designed to study the finite-sample properties of the proposed point and interval estimation of β by using information from the surrogate endpoint.

In the first study we assumed that the true endpoint (Y), the surrogate (S), and the covariate (X) were related according to the following linear models:

$$Y = \beta_1 + \beta_2 X + \epsilon_Y \quad \text{and} \quad S = \gamma_1 + \gamma_2 Y + \gamma_3 X + \epsilon_S$$

where ϵ_Y and ϵ_S were independent standard normal random errors, and X also followed a standard normal distribution independent of ϵ_Y and ϵ_S. The simulation study used the following four models for the relationship between S and (Y, X):

$$\begin{aligned} &\text{Model 1} && S = Y + \epsilon_S, \\ &\text{Model 2} && S = 1 + X + \epsilon_S, \\ &\text{Model 3} && S = 1 + Y + X + \epsilon_S, \\ &\text{Model 4} && S = 1 + 5Y + X + \epsilon_S. \end{aligned}$$

Model 1 corresponds to the situation where S is unbiased for Y. Model 2 corresponds to the situation where S is associated with X but not informative about Y. Note that, except for Model 1, S is biased for Y.

In the second simulation study, we considered three nonlinear models. In the first two models (Models 5 and 6), we assumed Y and S were exponential random variables with means $\beta_1 + \beta_2 X$ and $\beta_1 + \beta_2 X + \gamma\epsilon_s$, respectively, where X was $.25\chi^2_1$ and ϵ_s was a standard normal error. We chose $\gamma = 1$ and .5 which, respectively, defined Models 5 and 6. In the last model (Model 7), we assumed Y was a binary variable generated by a logistic model

$$P(Y = 1|X) = 1/\{1 + \exp(-\beta_1 - \beta_2 X)\},$$

where X was $.25\chi^2_1$ and S was binary with $P(S = 1|Y) = .6Y + .1$.

In both simulation studies the true parameter values $\beta_* = (\beta_{*1}, \beta_{*2})$ were fixed at $(0, 1)$. We used a total of 5,000 simulations in each of these studies. Each simulation was carried out by generating a pair of (Y, X) first. Then S was generated under each of the seven models, given (Y, X). This yielded an observation (Y, S, X). This exercise was then repeated $N = 200$ or 400 times to obtain a simulated sample. The nonvalidation sample was generated by randomly deleting Y in $(1 - \rho) \times 100\%$ of the observations. This defined the level of missingness in Y. Three levels of missingness—25%, 50%, and 75%—were used. The empirical Bartlett correction was based on $B = 100$ bootstrap resamples. All simulations were carried out using FORTRAN programs with IMSL routines, which can be obtained from the authors upon request.

For each model considered the estimates of β were obtained from the following methods: (1) the MLE using all N pairs of (X, Y) as if they were all observed [MLE(N)]; (2) the MLE using the pairs of (X, Y) in the validation sample only [MLE(N_V)]; (3) the three estimates of β using the methods of Rotnitzky and Robins (1995) (RR), Chen and Chen (2000) (CC), and the

current EL proposal (EL). For all models except Model 7, the estimating equations for RR, CC, and EL were

$$U(Y, X, \beta) = \big(Y - \beta_1 - \beta_2 X, (Y - \beta_1 - \beta_2 X)X\big)^T$$

and

$$h(S, X, \beta) = \big(Y - \gamma_1 - \gamma_2 X, (Y - \gamma_1 - \gamma_2 X)X\big)^T.$$

For Model 7 we used

$$U(Y, X, \beta) = \big(Y - 1/\{1 + \exp(-\beta_1 - \beta_2 X)\},$$
$$\big[Y - 1/\{1 + \exp(-\beta_1 - \beta_2 X)\}\big]X\big)^T$$

and

$$h(S, X, \gamma) = \big(S - 1/\{1 + \exp(-\gamma_1 - \gamma_2 X)\},$$
$$\big[S - 1/\{1 + \exp(-\gamma_1 - \gamma_2 X)\}\big]X\big)^T.$$

Apart from evaluating the proposed EL confidence regions for β, those proposed by CC and RR were also examined. CC used a normal approximation with the variance-covariance estimated by

$$D_1^{-1} C_{11} D_1^{-1} - (1 - \rho) D_1^{-1} C_{12} C_{22}^{-1} C_{12}^T D_1^{-1},$$

where $D_1 = n^{-1} \sum_{i=1}^n U_i'(\hat{\beta})$, $C_{11} = n^{-1} \sum_{i=1}^n U_i(\hat{\beta}) U_i'(\hat{\beta})$, $C_{12} = n^{-1} \sum_{i=1}^n U_i(\hat{\beta}) h_i'(\hat{\gamma})$, and $C_{22} = n^{-1} \sum_{i=1}^n h_i(\hat{\gamma}) h_i'(\hat{\gamma})$; RR also used a normal approximation but suggested a variance-covariance estimate of $D_1^{-1} C [D_1^T]^{-1}$, where $C = n^{-1} \times \sum_{i=1}^n \mathrm{Res}(U_i(\hat{\beta}), h_i(\hat{\gamma}))$ with $\mathrm{Res}(A, B) = A - E(AB^T) \times \{E(BB^T)\}^{-1} B$, the residual from the least squares regression of A over B. Note that the estimated variance-covariance matrices in CC and RR are not guaranteed to be invertible and when that is the case, the corresponding confidence regions are not available. Indeed, this had been encountered in up to 20% of the simulations. Also, the confidence regions produced by CC and RR are elliptical, whereas those based on the EL have naturally determined shapes and orientations.

The point estimate results are summarized in Tables 1–7 (corresponding to the seven models). A general observation from Tables 1–4 is that all methods give approximately unbiased results in all situations—the magnitude of the squared bias in each case is of a lower order than the variance. Therefore, we can focus our discussion on the standard errors of the estimates. Note that the parameter values chosen in the simulations induced a correlation between Y and S of $1/\sqrt{2}, 0, 1/\sqrt{2}$, and $\sqrt{25/26}$, respectively, for Models 1–4. Table 1 corresponds to the situation where S is unbiased for Y. From the bottom panel in that table, we can make two observations. First, the relative efficiencies of RR, CC, and EL to MLE(N_V) are larger than 1 in all cases studied. Second, the relative superiority of all three methods over MLE(N_V) increases with the proportion of data with missing Y. These observations also apply to the results in the other tables, except in Table 2, where S is uncorrelated to Y, given X. In Table 2 the relative efficiency between the proposed estimator and MLE(N_V) is nearly 1 in all cases. This indicates that there is little loss in using the proposed method compared to MLE(N_V) even when the surrogate is totally uninformative. Similar results are seen for CC and RR. Table 3 corresponds to the situation where S is related to Y, given X, but it is biased for Y. All three methods (RR, CC, and EL) are more efficient than MLE(N_V) in all cases studied, and the gain in efficiency increases with the proportion of missing data. There is little difference among the three methods. Finally, Table 4 represents the situation where S is highly correlated with Y, given X. Once again, RR, CC, and EL are more efficient than MLE(N_V) in all cases. However, we note that EL is the best performer, achieving almost the same efficiency as MLE(N), the MLE *as if all* (X, Y) *were observed*. The results given in Tables 5 and 6 for the nonlinear models (Models 5 and 6) demonstrate that EL is

Table 1. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under the Linear Model 1 With $S = Y + \epsilon_S$

Method	Parameter	Percentage missing, $(1-\rho) \times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.10 (7.02)	
	β_2		−.05 (6.98)	
MLE(N_V)	β_1	−.09 (8.29)	−.01 (10.17)	.00 (14.38)
	β_2	−.10 (8.19)	−.03 (10.08)	.04 (14.61)
RR	β_1	−.05 (7.62)	−.04 (8.72)	.28 (11.49)
	β_2	−.03 (7.68)	−.03 (8.87)	−.02 (12.19)
CC	β_1	−.06 (7.61)	−.05 (8.70)	.21 (11.36)
	β_2	−.03 (7.65)	−.02 (8.75)	−.02 (11.82)
EL	β_1	−.03 (7.65)	−.04 (8.70)	.25 (11.31)
	β_2	−.01 (7.74)	−.02 (8.84)	−.02 (11.81)
$N = 400$				
MLE(N)	β_1		−.01 (5.05)	
	β_2		−.01 (4.92)	
MLE(N_V)	β_1	.02 (5.78)	.04 (7.13)	.01 (10.10)
	β_2	.03 (5.73)	.04 (7.01)	.18 (10.13)
RR	β_1	.03 (5.44)	.03 (6.22)	−.02 (8.06)
	β_2	.04 (5.34)	.06 (6.10)	.15 (8.27)
CC	β_1	.04 (5.44)	.03 (6.24)	−.02 (8.02)
	β_2	.04 (5.33)	.06 (6.08)	.18 (8.13)
EL	β_1	.03 (5.45)	.02 (6.22)	−.02 (8.00)
	β_2	.03 (5.36)	.06 (6.09)	.16 (8.11)

Table 2. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under the Linear Model 2 With $S = 1 + X + \epsilon_S$

Method	Parameter	Percentage missing, $(1-\rho) \times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.10 (7.02)	
	β_2		−.05 (6.98)	
MLE(N_V)	β_1	−.09 (8.29)	−.01 (10.17)	.00 (14.38)
	β_2	−.10 (8.19)	−.03 (10.08)	.04 (14.61)
RR	β_1	−.09 (8.32)	.00 (10.14)	.10 (14.48)
	β_2	−.10 (8.24)	−.02 (10.24)	.06 (14.97)
CC	β_1	−.09 (8.31)	.00 (10.20)	.12 (14.63)
	β_2	−.10 (8.25)	−.04 (10.24)	.02 (15.40)
EL	β_1	−.08 (8.31)	.00 (10.13)	.10 (14.49)
	β_2	−.10 (8.23)	−.02 (10.22)	.03 (14.97)
$N = 400$				
MLE(N)	β_1		−.01 (5.05)	
	β_2		−.01 (4.92)	
MLE(N_V)	β_1	.02 (5.78)	.04 (7.13)	.01 (10.10)
	β_2	.03 (5.73)	.04 (7.01)	.18 (10.13)
RR	β_1	.02 (5.79)	.02 (7.12)	.01 (10.14)
	β_2	.03 (5.75)	.04 (7.04)	.15 (10.28)
CC	β_1	.02 (5.78)	.03 (7.14)	.02 (10.19)
	β_2	.03 (5.76)	.04 (7.08)	.19 (10.34)
EL	β_1	.02 (5.78)	.02 (7.12)	.00 (10.14)
	β_2	.03 (5.75)	.04 (7.04)	.15 (10.27)

Table 3. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under the Linear Model 3 With $S = 1 + Y + X + \epsilon_S$

Method	Parameter	Percentage missing, $(1-\rho) \times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.10 (7.02)	
	β_2		−.05 (6.98)	
MLE(N_V)	β_1	−.09 (8.29)	−.01 (10.17)	.00 (14.38)
	β_2	−.10 (8.19)	−.03 (10.08)	.04 (14.61)
RR	β_1	−.05 (7.62)	−.05 (8.72)	.27 (11.49)
	β_2	−.03 (7.68)	−.03 (8.87)	−.02 (12.19)
CC	β_1	−.06 (7.61)	−.56 (8.70)	−.82 (11.36)
	β_2	−.03 (7.65)	−.01 (8.75)	−.02 (11.82)
EL	β_1	−.03 (7.65)	−.04 (8.70)	.25 (11.31)
	β_2	−.01 (7.74)	−.02 (8.84)	−.02 (11.81)
$N = 400$				
MLE(N)	β_1		−.01 (5.05)	
	β_2		−.01 (4.92)	
MLE(N_V)	β_1	.02 (5.78)	.04 (7.13)	.01 (10.10)
	β_2	.03 (5.73)	.04 (7.01)	.18 (10.13)
RR	β_1	.03 (5.44)	.02 (6.22)	−.02 (8.06)
	β_2	.04 (5.34)	.06 (6.10)	.15 (8.27)
CC	β_1	.04 (5.44)	−.22 (6.24)	−.53 (8.02)
	β_2	.04 (5.33)	.06 (6.08)	.18 (8.12)
EL	β_1	.03 (5.45)	.02 (6.22)	−.02 (8.00)
	β_2	.03 (5.36)	.06 (6.09)	.16 (8.11)

Table 5. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under Model 5: Y and S Are Exponential Random Variables With Means $\beta_1 + \beta_2 X$ and $\beta_1 + \beta_2 X + \epsilon_S$, Respectively

Method	Parameter	Percentage missing, $(1-\rho) \times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.23 (11.55)	
	β_2		.88 (47.05)	
MLE(N_V)	β_1	−.18 (13.26)	1.55 (16.48)	3.66 (23.23)
	β_2	1.11 (54.08)	−.82 (65.95)	−3.57 (93.22)
RR	β_1	.18 (12.52)	.50 (14.28)	.90 (19.19)
	β_2	−1.12 (49.37)	−3.29 (54.45)	−7.58 (70.84)
CC	β_1	−.16 (12.48)	.84 (14.82)	1.96 (23.35)
	β_2	.98 (49.23)	.16 (56.04)	−1.76 (90.30)
EL	β_1	.64 (12.12)	.92 (13.77)	1.27 (18.28)
	β_2	−6.13 (47.14)	−5.80 (50.76)	−6.36 (64.23)
$N = 400$				
MLE(N)	β_1		−.13 (8.36)	
	β_2		.72 (33.67)	
MLE(N_V)	β_1	−.28 (9.62)	.59 (11.74)	1.30 (16.72)
	β_2	1.27 (38.79)	−.40 (47.67)	.03 (67.93)
RR	β_1	.09 (9.05)	.39 (10.07)	.60 (13.36)
	β_2	−.40 (35.43)	−2.37 (38.34)	−4.64 (48.86)
CC	β_1	−.12 (8.96)	.47 (10.17)	.94 (14.56)
	β_2	.76 (34.95)	−.28 (38.25)	−.66 (53.00)
EL	β_1	.45 (8.81)	.64 (9.77)	.73 (12.93)
	β_2	−3.42 (34.24)	−3.76 (36.34)	−3.47 (45.15)

more efficient than the other methods. For Model 7 the estimates for β_2 are quite severely biased when $N = 200$ and the proportion of missingness is high (75%). Despite a correlation of nearly .5 between S and Y, none of the methods improve significantly over MLE(N_V).

Table 8 gives the empirical coverage of the 90% confidence regions of CC, RR, and the proposed EL method. For EL we report the likelihood ratio-based confidence interval from Theorem 2 (EL) as well as the Bartlett-corrected confidence interval (EL_{BC}). Under Models 1–4 the EL confidence levels (EL), though smaller than the nominal 90% coverage, are very close to those produced by CC and are better than those produced by RR. The confidence regions of RR fail to produce the correct coverage for Models 4–6. Both EL and CC have severe undercoverage in Models 5 and 6. These are the situations where the Bartlett-correction is needed. Indeed, we see the

Table 4. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under the Linear Model 4 With $S = 1 + 5Y + X + \epsilon_S$

Method	Parameter	Percentage missing, $(1-\rho) \times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.10 (7.02)	
	β_2		−.05 (6.98)	
MLE(N_V)	β_1	−.09 (8.29)	−.01 (10.17)	.00 (14.38)
	β_2	−.10 (8.19)	−.03 (10.08)	.04 (14.61)
RR	β_1	−.09 (7.06)	−.11 (7.28)	.04 (8.08)
	β_2	−.03 (7.09)	−.03 (7.35)	−.11 (8.62)
CC	β_1	−.09 (7.04)	−.29 (7.15)	−.41 (7.46)
	β_2	−.04 (7.06)	−.01 (7.22)	.02 (8.04)
EL	β_1	−.08 (7.12)	−.11 (7.25)	.02 (7.52)
	β_2	−.01 (7.21)	−.02 (7.27)	−.06 (7.59)
$N = 400$				
MLE(N)	β_1		−.01 (5.05)	
	β_2		−.01 (4.92)	
MLE(N_V)	β_1	.02 (5.78)	.04 (7.13)	.01 (10.10)
	β_2	.03 (5.73)	.04 (7.01)	.18 (10.13)
RR	β_1	.00 (5.09)	.00 (5.19)	−.03 (5.59)
	β_2	.01 (4.95)	.02 (5.06)	.04 (5.67)
CC	β_1	.00 (5.09)	−.10 (5.17)	−.22 (5.33)
	β_2	.00 (4.94)	.01 (5.02)	.02 (5.42)
EL	β_1	.00 (5.12)	−.01 (5.19)	−.03 (5.37)
	β_2	.01 (4.99)	.02 (5.05)	.03 (5.29)

Table 6. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under Model 6: Y and S Are Exponential Random Variables With Means $\beta_1 + \beta_2 X$ and $\beta_1 + \beta_2 X + .5\epsilon_S$, Respectively

Method	Parameter	Percentage missing, $(1-\rho) \times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.23 (11.55)	
	β_2		.88 (47.05)	
MLE(N_V)	β_1	−.18 (13.26)	1.55 (16.48)	3.66 (23.23)
	β_2	1.11 (54.08)	−.82 (65.95)	−3.57 (93.22)
RR	β_1	.06 (11.94)	.49 (12.76)	1.03 (15.73)
	β_2	−1.24 (47.90)	−3.89 (50.23)	−8.27 (60.52)
CC	β_1	−.12 (11.93)	.49 (12.95)	1.30 (16.61)
	β_2	1.06 (47.79)	.89 (50.26)	.94 (61.63)
EL	β_1	.70 (11.53)	.99 (12.22)	1.19 (14.57)
	β_2	−6.96 (45.71)	−6.73 (46.20)	−6.79 (51.69)
$N = 400$				
MLE(N)	β_1		−.13 (8.36)	
	β_2		.72 (33.67)	
MLE(N_V)	β_1	−.28 (9.62)	.59 (11.74)	1.31 (16.72)
	β_2	1.27 (38.79)	−.40 (47.67)	.03 (67.93)
RR	β_1	.07 (8.64)	.44 (9.13)	.82 (11.06)
	β_2	−.53 (34.50)	−2.48 (35.55)	−5.05 (42.17)
CC	β_1	−.05 (8.61)	.31 (9.12)	.72 (11.30)
	β_2	.72 (33.96)	.31 (34.92)	.61 (40.65)
EL	β_1	.59 (8.42)	.73 (8.82)	.75 (10.49)
	β_2	−4.05 (33.20)	−4.00 (33.46)	−3.72 (36.94)

Table 7. Bias (Standard Error) in Estimating $\beta = (\beta_1, \beta_2)$ Under Model 7: $P(Y = 1|X) = 1/\{1 + \exp(-\beta_1 - \beta_2 X)\}$ Where X is $.25\chi^2_1$ and S Is Binary With $P(S = 1|Y) = .6\,Y + .1$

Method	Parameter	Percentage missing, $(1-\rho)\times 100\%$ 25	50	75
$N = 200$		Bias $\times 10^2$ (SE $\times 10^2$)		
MLE(N)	β_1		−.54 (18.29)	
	β_2		3.95 (50.54)	
MLE(N_V)	β_1	−.53 (21.27)	−1.04 (26.06)	−2.77 (38.26)
	β_2	4.95 (59.49)	9.31 (76.96)	22.61 (125.68)
RR	β_1	−.77 (21.22)	−.87 (24.91)	−2.45 (35.14)
	β_2	6.03 (60.47)	9.38 (76.07)	21.48 (120.52)
CC	β_1	−.57 (20.48)	−.88 (24.27)	−2.63 (34.54)
	β_2	4.49 (58.15)	7.43 (73.76)	17.09 (117.34)
EL	β_1	−.56 (20.51)	−.80 (24.31)	−2.18 (34.71)
	β_2	5.10 (58.56)	8.61 (74.15)	19.80 (119.02)
$N = 400$				
MLE(N)	β_1		−.14 (12.82)	
	β_2		2.24 (34.06)	
MLE(N_V)	β_1	−.22 (14.91)	−.47 (18.36)	−.89 (26.05)
	β_2	3.02 (39.78)	4.92 (49.67)	9.02 (74.95)
RR	β_1	−.43 (14.73)	−.77 (17.27)	−1.03 (23.38)
	β_2	3.46 (39.94)	5.14 (48.61)	8.40 (71.72)
CC	β_1	−.27 (14.32)	−.62 (16.96)	−1.00 (23.18)
	β_2	2.68 (38.82)	4.04 (47.55)	6.05 (70.75)
EL	β_1	−.29 (14.34)	−.60 (16.98)	−.83 (23.23)
	β_2	3.04 (38.92)	4.64 (47.66)	7.44 (71.20)

Table 8. Empirical Coverage of 90% Confidence Regions From Three Different Methods

		$N = 200$			$N = 400$		
		Percentage missing, $(1-\rho)\times 100\%$					
Model	Method	25	50	75	25	50	75
1	CC	.892	.885	.867	.889	.889	.876
	RR	.736	.762	.783	.726	.767	.810
	EL	.889	.877	.858	.888	.890	.873
	EL_{BC}	.903	.895	.907	.892	.892	.893
2	CC	.887	.873	.850	.888	.887	.872
	RR	.886	.877	.851	.888	.889	.874
	EL	.890	.877	.843	.889	.892	.872
	EL_{BC}	.899	.897	.902	.892	.899	.894
3	CC	.893	.885	.867	.890	.889	.876
	RR	.736	.762	.783	.727	.767	.810
	EL	.889	.877	.858	.888	.890	.873
	EL_{BC}	.900	.898	.906	.891	.896	.895
4	CC	.892	.877	.777	.888	.883	.800
	RR	.197	.191	.326	.173	.179	.323
	EL	.879	.881	.878	.882	.887	.884
	EL_{BC}	.904	.898	.901	.891	.889	.894
5	CC	.828	.783	.675	.851	.820	.749
	RR	.613	.600	.565	.622	.628	.625
	EL	.823	.811	.771	.845	.830	.813
	EL_{BC}	.888	.878	.890	.885	.882	.881
6	CC	.830	.789	.700	.852	.822	.759
	RR	.413	.393	.402	.413	.409	.455
	EL	.803	.807	.789	.836	.835	.824
	EL_{BC}	.880	.879	.889	.885	.881	.885
7	CC	.905	.898	.884	.901	.899	.893
	RR	.816	.825	.833	.824	.845	.853
	EL	.886	.872	.835	.894	.891	.869
	EL_{BC}	.893	.891	.895	.897	.895	.897

Bartlett-correction (EL_{BC}) markedly improves the coverage of the confidence regions in Models 5 and 6. The remarkable performance of the Bartlett-correction is also seen in other models and all sample sizes considered.

4. EXAMPLE: HEART DATA

Besarab et al. (1998) reported on a randomized trial on 1,233 patients with clinical evidence of congestive heart failure or ischemic heart disease who were undergoing hemodialysis: 618 patients were assigned to receive increasing doses of epoetin to achieve and to maintain a hematocrit of 42%, and 615 were assigned to receive doses of epoetin sufficient to maintain a hematocrit of 30% throughout the study. The primary endpoint was the length of time to death or a first nonfatal myocardial infarction. In this study hematocrit level is a surrogate endpoint as it is believed that cardiac arrest is the most common cause of death among patients receiving dialysis and correction or maintenance of anemia reduces exercise-induced cardiac ischemia and ameliorates the left ventricular hypertrophy that predisposes patients to death and cardiac-related morbidity. After 29 months there were 183 deaths and 19 first nonfatal myocardial infarctions among the patients with increasing doses of epoetin and 150 deaths and 14 nonfatal myocardial infarctions among those in the group with lower doses of epoetin. Although the difference in event-free survival between the two groups did not reach the prespecified statistical stopping boundary, the study was halted. The causes of death in the two groups were similar. The mortality rates decreased with increasing hematocrit values in both groups. The patients in the increasing dose group had a decline in the adequacy of dialysis and received intravenous iron dextran more often than those in the low-hematocrit group, and the investigators concluded that in patients with clinically evident congestive heart failure or ischemic heart disease who were receiving hemodialysis, administration of epoetin to raise their hematocrit was not recommended. Therefore, in this case, even though the treatment is effective in correcting the surrogate endpoint (hematocrit level), it is ineffective in influencing the true endpoint (death).

We reanalyzed the data using the proposed EL method, CC, and RR. We also included results of the MLEs that use only the primary endpoints. We focused on mortality as the primary endpoint. The data consist of Y (binary: dead vs. alive), S (binary: hematocrit $< 33\%$ vs. $\geq 33\%$), hematocrit and X (binary: two regimens of epoetin). We assumed Y was generated by the logistic model:

$$P(Y = 1|X) = 1/\{1 + \exp(-\beta_1 - \beta_2 X)\}$$

and S was generated by $P(S = 1|X) = 1/\{1 + \exp(-\gamma_1 - \gamma_2 X)\}$. So U and h were based on the score functions of these models.

We randomly deleted the true endpoint from subsets of the patients. The results based on deleting 25%, 50%, and 75% of the true endpoints from the data are given in Table 9. There is little difference among the parameter estimates of EL, CC, and RR. The main difference among them is in the confidence intervals. These results are consistent with the findings in the simulation studies. The EL confidence intervals for β_1 are much shorter than those of RR, CC, and MLE(N_V), whereas the lengths of CC and RR are almost the same as those

Table 9. Parameter Estimates and 95% Confidence Interval (Width of CI) of Parameters for Heart Data Example

MLE(N)	β_1	−1.131[a] −1.309, −.954[b] (.355[c])
	β_2	.266 .021, .510 (.490)

		Percentage missing, $(1-\rho)\times 100\%$		
Method	*Parameter*	*25*	*50*	*75*
MLE(N_V)	β_1	−1.127 −1.335, −.918 (.417)	−1.126 −1.372, −.880 (.492)	−1.057 −1.382, −.732 (.650)
	β_2	.230 −.058, .518 (.576)	.238 −.107, .583 (.690)	.252 −.203, .707 (.909)
CC	β_1	−1.127 −1.335, −.919 (.417)	−1.122 −1.368, −.877 (.492)	−1.036 −1.360, −.712 (.648)
	β_2	.231 −.057, .518 (.576)	.237 −.107, .582 (.689)	.223 −.230, .675 (.905)
RR	β_1	−1.127 −1.335, −.919 (.415)	−1.122 −1.368, −.877 (.491)	−1.038 −1.363, −.713 (.650)
	β_2	.231 −.056, .518 (.574)	.238 −.106, .581 (.687)	.225 −.228, .678 (.906)
EL	β_1	−1.127 −1.283, −.981 (.302)	−1.122 −1.306, −.954 (.352)	−1.038 −1.284, −.817 (.467)
	β_2	.231 −.057, .519 (.576)	.238 −.106, .583 (.689)	.225 −.228, .681 (.909)

[a]Parameter estimate.
[b]95% confidence interval.
[c]Width of 95% confidence interval.

of MLE(N_V). None of the methods shows a significant difference in treatment effects between the treatment arms ($\beta_2 \neq 0$). Note that the treatment effects estimates are in the right direction (i.e., the same sign as that using the true endpoints only [MLE(N_V)]) in all cases considered. For example, for the case where 50% of the true endpoints are deleted, the p values using MLE(N_V), CC, RR, and EL are all equal to .176, to three significant digits. Even though these results are not significant, as opposed to the results from all 1,233 patients, where the result is significant at $p = .04$, the important point is all four methods show that high doses of epoetin are harmful. The reason for the nonsignificant results can be attributed to the loss of statistical power from the deleted cases. This can be seen by observing that the confidence intervals using the full data from all 1,233 patients [MLE(N)] are approximately shorter than the corresponding ones from the other methods by a factor of $1/\sqrt{(1-\rho)}$.

In cases such as the current example, where the surrogate is negatively correlated with the true endpoint, there may be concerns that using a method that incorporates the surrogate may give misleading results. As shown in this example, however, such is not the case. The important point is that the parameter estimates are in the right direction. This is an example of an augmented surrogate study (Begg and Leung 2000), where there exist a validation and a nonvalidation sample. The primary endpoint data in the validation sample will guarantee that the parameter estimates are in the right directions. The surrogate endpoint data in the nonvalidation sample only serve to improve the precision of the parameter estimates (due to the surrogate's correlation to the primary endpoint). In fact, the problem is symmetric; that is, a surrogate S that has a negative correlation with Y is just as good as another surrogate, say S', with a positive correlation of the same magnitude. This is because data in S can be converted to data in S' by simply taking $S' = -S$.

5. DISCUSSION

In this article, we introduce a method based on the EL for recovering information from a study where the primary endpoint is missing on some observations, but a surrogate endpoint is measured on all observations. As in previous works discussed here, the method assumes that the validation and the nonvalidation samples are random samples from the population. We show that the method gives unbiased estimates and it improves upon the MLE that ignores the surrogate data. The relative improvement is higher when the nonvalidation sample is large relative to the validation sample. Furthermore, when the information provided by the surrogate is highly "correlated" to that using the true endpoint, the method is more efficient.

As suggested in Fleming and Demets (1996) and Begg and Leung (2000), one of the major concerns of using a method that uses surrogate endpoints in the inference procedure is the robustness of the method. Our simulations (Model 2) and analysis in the example in Section 4 give us some assurance that, even in cases where the surrogate is unrelated to or negatively related to the primary endpoint, the proposed method will not lead to biased results. This robustness property can be partly attributed to the fact that the setup considered here is that of an augmented surrogate study (Begg and Leung 2000). In an augmented surrogate study, validation data are available to provide a benchmark on the directions of the parameter estimates. This is contrary to a pure surrogate study in which only nonvalidation data are available and no such benchmarking is possible. The use of any method in a pure surrogate study can be risky. Our study supports the use of surrogates only in an augmented surrogate study.

When compared with two existing methods—Rotnitzky and Robins (1995) (RR) and Chen and Chen (2000) (CC)—the proposed EL method gives the same asymptotic efficiency for point

estimation. However, we find in the simulations that the proposed method compares favorably with the methods of CC and RR in moderate samples. In particular, in the linear model situation, the proposed method performs similarly to CC in most cases and is better than CC when the correlation between the surrogate and the true endpoint is high and there are a lot of missing data. In the exponential nonlinear model situation, the EL method performs much better than CC and RR in terms of efficiency and coverage probabilities. In the binary logistic model, the results are in favor of EL and CC.

One of the main advantages of the proposed EL method over the existing methods is the EL method's ability to obtain efficient point estimates and confidence regions in a single framework. The resulting confidence intervals are also Bartlett correctable for coverage improvement. The proposed empirical Bartlett correction is simple to implement and its performance is quite remarkable. The proposed EL method can be extended to handle the situation where surrogates for both the endpoint and the covariates are available. In that case the function h would be in terms of a surrogate for the endpoint and the surrogates for the covariates. In this article, we assume the validation sample is a random subsample of the data. It would be useful to extend the investigation to situations where the validation is nonrandom.

APPENDIX: PROOFS

The conditions needed to establish Theorems 1 and 2 are the following:

(A.1)

(a) Both $V(U) = E\{U(\beta_*)U^T(\beta_*)\}$ and $V(h) = E\{h(\gamma_*)h^T(\gamma)\}$ are positive definite, and the ranks of $E(\frac{\partial U(\beta_*)}{\partial \beta})$ and $E(\frac{\partial h(\gamma_*)}{\partial \gamma})$ are, respectively, p and r;

(b) $\frac{\partial^2 U(\beta)}{\partial \beta \partial \beta^T}$ is continuous in a neighborhood of β_*, and in this neighborhood both $\|\frac{\partial U(\beta)}{\partial \beta}\|$ and $\|U(\beta)\|^3$ are bounded;

(c) $\frac{\partial^2 h(\gamma)}{\partial \gamma \partial \gamma^T}$ is continuous in a neighborhood of γ_*, and in this neighborhood both $\|\frac{\partial h(\gamma)}{\partial \gamma}\|$ and $\|h(\gamma)\|^3$ are bounded;

(d) n and $m \to \infty$, and $n/N \to \rho \in (0, 1)$ as $N \to \infty$.

A.1 Proof of Theorem 1

By a similar derivation to that in Qin and Lawless (1994), it may be shown that $\hat{\lambda}_i = O_p(n^{-1/2})$ for $i = 1, 2$, and 3. Using Taylor series expansion on the left sides of the five equations (5), (6), (7), (9), and (10) at $(\beta_*, \gamma_*, 0, 0, 0)$, where β_* and γ_* are the true parameter values, we have, by ignoring terms of $O_p(n^{-1})$,

$$\sum_{i=1}^{n} U_i'(\beta_*)\hat{\lambda}_1 = 0, \tag{A.2}$$

$$\sum_{i=1}^{n} h_i'(\gamma_*)\hat{\lambda}_2 + \sum_{j=1}^{m} h_{n+j}'(\gamma_*)\hat{\lambda}_3 = 0, \tag{A.3}$$

$$\sum_{i=1}^{n} U_i'(\beta_*)(\hat{\beta} - \beta_*) - \sum_{i=1}^{n} U_i(\beta_*)U_i^T(\beta_*)\hat{\lambda}_1 - \sum_{i=1}^{n} U_i(\beta_*)h_i^T(\gamma_*)\hat{\lambda}_2 = -\sum_{i=1}^{n} U_i(\beta_*), \tag{A.4}$$

$$\sum_{i=1}^{n} h_i'(\gamma_*)(\hat{\gamma} - \gamma_*) - \sum_{i=1}^{n} h_i(\gamma_*)U_i^T(\beta_*)\hat{\lambda}_1 - \sum_{i=1}^{n} h_i(\gamma_*)h_i^T(\gamma_*)\hat{\lambda}_2 = -\sum_{i=1}^{n} h_i(\gamma_*), \tag{A.5}$$

$$\sum_{j=1}^{m} h_{n+j}'(\gamma_*)(\hat{\gamma} - \gamma_*) - \sum_{j=1}^{m} h_{n+j}(\gamma_*)h_{n+j}^T(\gamma_*)\hat{\lambda}_3 = -\sum_{j=1}^{m} h_{n+j}(\gamma_*). \tag{A.6}$$

Let

$$s_{12} = \begin{pmatrix} n^{-1}\sum_{i=1}^{n} U_i'(\beta_*) & 0 & 0 \\ 0 & n^{-1}\sum_{i=1}^{n} h_i'(\gamma_*) & n^{-1}\sum_{j=1}^{m} h_{n+j}'(\gamma_*) \end{pmatrix},$$

$$s_{22} = \begin{pmatrix} n^{-1}\sum_{i=1}^{n} U_i(\beta_*)U_i^T(\beta_*) & n^{-1}\sum_{i=1}^{n} U_i(\beta_*)h_i^T(\gamma_*) & 0 \\ n^{-1}\sum_{i=1}^{n} h_i(\gamma_*)U_i^T(\beta_*) & n^{-1}\sum_{i=1}^{n} h_i(\gamma_*)h_i^T(\gamma_*) & 0 \\ 0 & 0 & n^{-1}\sum_{j=1}^{m} h_{n+j}(\gamma_*)h_{n+j}^T(\gamma_*) \end{pmatrix},$$

and

$$S_n = \begin{pmatrix} 0 & s_{12} \\ s_{12}^T & -s_{22} \end{pmatrix}.$$

Throughout this article, we use the same symbol 0 for either the scalar quantity 0 or a null matrix of 0's. In the latter case we leave the dimension of the null matrix unspecified because it can be determined easily from the other components of the matrices.

Equations (A.2)–(A.6) imply that

$$S_n\big((\hat{\beta} - \beta_*)^T, (\hat{\gamma} - \gamma_*)^T, \hat{\lambda}_1^T, \hat{\lambda}_2^T, \hat{\lambda}_3^T\big)^T = -Q_n, \tag{A.7}$$

where

$$Q_n = \left(0, 0, n^{-1}\sum_{i=1}^{n} U_i^T(\beta_*), n^{-1}\sum_{i=1}^{n} h_i^T(\gamma_*), n^{-1}\sum_{j=1}^{m} h_{n+j}^T(\gamma_*)\right)^T.$$

It may be shown by applying algebra on block matrices that

$$S_n^{-1} = \begin{pmatrix} (s_{12}s_{22}^{-1}s_{12}^T)^{-1} & (s_{12}s_{22}^{-1}s_{12}^T)^{-1}s_{12}s_{22}^{-1} \\ s_{22}^{-1}s_{12}^T(s_{12}s_{22}^{-1}s_{12}^T)^{-1} & -s_{22}^{-1} + s_{22}^{-1}s_{12}^T(s_{12}s_{22}^{-1}s_{12}^T)^{-1}s_{12}s_{22}^{-1} \end{pmatrix}.$$

Let

$$T_n = \left(n^{-1}\sum_{i=1}^{n} U_i^T(\beta_*), n^{-1}\sum_{i=1}^{n} h_i^T(\gamma_*), n^{-1}\sum_{j=1}^{m} h_{n+j}^T(\gamma_*)\right)^T.$$

From (A.7)

$$\begin{pmatrix} \hat{\beta} - \beta_* \\ \hat{\gamma} - \gamma_* \end{pmatrix} = -(s_{12}s_{22}^{-1}s_{12}^T)^{-1}s_{12}s_{22}^{-1}T_n. \tag{A.8}$$

Define $U = U_i(\beta_*)$ and $h = h_i(\gamma_*)$,

$$\Sigma_{12} = \begin{pmatrix} E\left(\frac{\partial U}{\partial \beta}\right) & 0 & 0 \\ 0 & E\left(\frac{\partial h}{\partial \gamma}\right) & (1-\rho)\rho^{-1}E\left(\frac{\partial h}{\partial \gamma}\right) \end{pmatrix},$$

$$\Sigma_{22} = \begin{pmatrix} E(UU^T) & E(Uh^T) & 0 \\ E(hU^T) & E(hh^T) & 0 \\ 0 & 0 & (1-\rho)\rho^{-1}E(hh_{n+j}) \end{pmatrix},$$

and

$$\Sigma = \begin{pmatrix} 0 & \Sigma_{12} \\ \Sigma_{12}^T & -\Sigma_{22} \end{pmatrix}.$$

Standard asymptotics show that, as $N \to \infty$,

$$s_{12} \xrightarrow{p} \Sigma_{12}, \qquad s_{22} \xrightarrow{p} \Sigma_{22}, \qquad S_n \xrightarrow{p} \Sigma. \tag{A.9}$$

Because $n^{1/2}T_n \xrightarrow{d} N(0, \Sigma_{22})$, from (A.8) and (A.9), we immediately have

$$n^{1/2}\big((\hat\beta - \beta_*)^T, (\hat\gamma - \gamma_*)^T\big) \xrightarrow{d} N\big(0, (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}\big). \tag{A.10}$$

By standard matrix manipulation,

$$\begin{aligned}
&(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1} \\
&\quad = \begin{pmatrix} E^{-1}\left(\dfrac{\partial U}{\partial \beta}\right) & 0 \\ 0 & E^{-1}\left(\dfrac{\partial h}{\partial \gamma}\right) \end{pmatrix} \\
&\qquad \times \begin{pmatrix} V(U) - (1-\rho)E(Uh^T)V^{-1}(h)E(hU^T) & \rho E(Uh^T) \\ \rho E(hU^T) & \rho V(h) \end{pmatrix} \\
&\qquad \times \begin{pmatrix} E^{-1}\left(\dfrac{\partial U}{\partial \beta}\right) & 0 \\ 0 & E^{-1}\left(\dfrac{\partial h}{\partial \gamma}\right) \end{pmatrix}.
\end{aligned} \tag{A.11}$$

Therefore, as $N \to \infty$,

$$\begin{aligned}
\hat\beta - \beta_* \xrightarrow{d} N\Bigg(0, E^{-1}\left(\frac{\partial U}{\partial \beta}\right)\big\{V(U) - (1-\rho)E(Uh^T) \\
\times V^{-1}(h)E(hU^T)\big\}E^{-1}\left(\frac{\partial U}{\partial \beta}\right)\Bigg),
\end{aligned} \tag{A.12}$$

which completes the proof of Theorem 1.

A.2 Proof of Theorem 2

We start by deriving an expansion for $\ell(\hat\beta, \hat\gamma)$. Applying Taylor expansion on (8) and ignoring the constant $-2n\log(n) - 2m\log(m)$, we have

$$\begin{aligned}
\ell(\hat\beta, \hat\gamma) &= 2\hat\lambda_1 \sum_{i=1}^n U_i(\hat\beta) + 2\hat\lambda_2 \sum_{i=1}^n h_i(\hat\gamma) \\
&\quad - \Bigg\{\hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)U_i^T(\hat\beta)\hat\lambda_1 \\
&\qquad + 2\hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)h_i^T(\hat\gamma)\hat\lambda_2 + \hat\lambda_2^T \sum_{i=1}^n h_i(\hat\gamma)h_i^T(\hat\gamma)\hat\lambda_2\Bigg\} \\
&\quad + 2\hat\lambda_3 \sum_{j=1}^m h_{n+j}(\hat\gamma) - \hat\lambda_3^T \sum_{j=1}^m h_{n+j}(\hat\gamma)h_{n+j}^T(\hat\gamma)\hat\lambda_3 + o_p(1).
\end{aligned} \tag{A.13}$$

From (A.4)–(A.6)

$$\begin{aligned}
\hat\lambda_1 \sum_{i=1}^n U_i(\hat\beta) &= -\hat\lambda_1^T \sum_{i=1}^n \frac{\partial U_i(\beta_*)}{\partial \beta}(\hat\beta - \beta_*) \\
&\quad + \hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)U_i^T(\hat\beta)\hat\lambda_1 + \hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)h_i^T(\hat\gamma)\hat\lambda_2,
\end{aligned}$$

$$\begin{aligned}
\hat\lambda_2 \sum_{i=1}^n h_i(\hat\gamma) &= -\hat\lambda_2^T \sum_{i=1}^n \frac{\partial h_i(\gamma_*)}{\partial \gamma}(\hat\gamma - \gamma_*) \\
&\quad + \hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)h_i^T(\hat\gamma)\hat\lambda_2 + \hat\lambda_2^T \sum_{i=1}^n h_i(\hat\gamma)h_i^T(\hat\gamma)\hat\lambda_2,
\end{aligned}$$

and

$$\begin{aligned}
\hat\lambda_3 \sum_{j=1}^m h_{n+j}(\hat\gamma) &= -\hat\lambda_3^T \sum_{j=1}^m \frac{\partial h_{n+j}(\gamma_*)}{\partial \gamma}(\hat\gamma - \gamma_*) \\
&\quad + \hat\lambda_3^T \sum_{j=1}^m h_{n+j}(\hat\gamma)h_{n+j}^T(\hat\gamma)\hat\lambda_3 + o_p(1).
\end{aligned}$$

These simplify (A.13) to

$$\begin{aligned}
\ell(\hat\beta, \hat\gamma) &= -2\hat\lambda_1^T \sum_{i=1}^n \frac{\partial U_i(\beta_*)}{\partial \beta}(\hat\beta - \beta_*) \\
&\quad - 2\hat\lambda_2^T \sum_{i=1}^n \frac{\partial h_i(\gamma_*)}{\partial \gamma}(\hat\gamma - \gamma_*) \\
&\quad - 2\hat\lambda_3^T \sum_{j=1}^m \frac{\partial h_{n+j}(\gamma_*)}{\partial \gamma}(\hat\gamma - \gamma_*) \\
&\quad + \Bigg\{\hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)U_i^T(\hat\beta)\hat\lambda_1 \\
&\qquad + 2\hat\lambda_1^T \sum_{i=1}^n U_i(\hat\beta)h_i^T(\hat\gamma)\hat\lambda_2 + \hat\lambda_2^T \sum_{i=1}^n h_i(\hat\gamma)h_i^T(\hat\gamma)\hat\lambda_2\Bigg\} \\
&\quad + \hat\lambda_3^T \sum_{j=1}^m h_{n+j}(\hat\gamma)h_{n+j}^T(\hat\gamma)\hat\lambda_3 + o_p(1) \\
&= n\hat\lambda^T s_{22}\hat\lambda - 2\hat\lambda^T s_{12}\big((\hat\beta - \beta_*)^T, (\hat\gamma - \gamma_*)^T\big)^T + o_p(1),
\end{aligned} \tag{A.14}$$

where $\hat\lambda = (\hat\lambda_1^T, \hat\lambda_2^T, \hat\lambda_3^T)^T$. Because, by ignoring terms of $o_p(n^{-1/2})$,

$$\hat\lambda = n^{-1}\{-s_{22}^{-1} + s_{22}^{-1}s_{12}^T(s_{12}s_{22}s_{12}^T)^{-1}s_{12}s_{22}^{-1}\}T_n$$

and $\hat\lambda^T s_{12} = 0$,

$$\begin{aligned}
\ell(\hat\beta, \hat\gamma) &= n\hat\lambda^T s_{22}\hat\lambda + o_p(1) \\
&= nT_n^T\{\Sigma_{22}^{-1} - \Sigma_{22}^{-1}\Sigma_{12}^T(\Sigma_{12}\Sigma_{22}\Sigma_{12}^T)^{-1} \\
&\qquad \times \Sigma_{12}\Sigma_{22}^{-1}\}T_n + o_p(1).
\end{aligned} \tag{A.15}$$

Next we develop an expansion for $\ell\{\beta_*, \tilde\gamma(\beta_*)\}$. Let

$$\tilde\Sigma_{12} = \begin{pmatrix} 0 & E\left(\dfrac{\partial h}{\partial \gamma}\right) & (1-\rho)\rho^{-1}E\left(\dfrac{\partial h}{\partial \gamma}\right) \end{pmatrix}.$$

It may be shown by similar derivations to those leading to (A.14) that

$$\begin{aligned}
\ell\{\beta_*, \tilde\gamma(\beta_*)\} &= nT_n^T\{\Sigma_{22}^{-1} - \Sigma_{22}^{-1}\tilde\Sigma_{12}^T(\tilde\Sigma_{12}\Sigma_{22}\tilde\Sigma_{12}^T)^{-1} \\
&\qquad \times \tilde\Sigma_{12}\Sigma_{22}^{-1}\}T_n + o_p(1).
\end{aligned} \tag{A.16}$$

From (A.15) and (A.16) $\ell(\hat\beta, \hat\gamma) - \ell\{\beta_*, \tilde\gamma(\beta_*)\} = T_n^T\Sigma_{22}^{-1/2}A \times \Sigma_{22}^{-1/2}T_n + o_p(1)$, where

$$\begin{aligned}
A = \Sigma_{22}^{-1/2}\{&\Sigma_{12}^T(\Sigma_{12}\Sigma_{22}\Sigma_{12}^T)^{-1}\Sigma_{12} \\
&- \tilde\Sigma_{12}^T(\tilde\Sigma_{12}\Sigma_{22}\tilde\Sigma_{12}^T)^{-1}\tilde\Sigma_{12}\}\Sigma_{22}^{-1/2}.
\end{aligned}$$

Note that (1) $\sqrt{n}\Sigma_{22}^{-1/2}T_n \xrightarrow{d} N(0, I_p)$, where I_p is the $p \times p$ identity matrix; and (2) $\mathrm{tr}(A) = p$ and $A^2 = A$. These facts imply that $\ell(\hat{\beta}, \hat{\gamma}) - \ell\{\beta_*, \tilde{\gamma}(\beta_*)\}$ has a limiting χ_p^2 distribution, hence completing the proof.

[Received September 2000. Revised July 2003.]

REFERENCES

Albert, A., and Anderson, J. (1984), "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika*, 71, 1–10.

Bang, H., and Tsiatis, A. (2000), "Estimating Medical Costs With Censored Data," *Biometrika*, 87, 329–343.

Begg, C., and Leung, D. (2000), "On the Use of Surrogate Endpoints in Randomized Trials," *Journal of the Royal Statistical Society*, Ser. A, 163, 15–27.

Besarab, A., Bolton, W., Browne, J., Egrie, J., Nissenson, A., Okamoto, D., Schwab, S., and Goodkin, D. (1998), "The Effects of Normal as Compared With Low Hematocrit Values in Patients With Cardiac Disease Who Are Receiving Hemodialysis and Epoetin," *New England Journal of Medicine*, 339, 584–590.

Chen, S.-X. (1993), "On the Coverage Accuracy of Empirical Likelihood Confidence Regions for Linear Regression Model," *Annals of the Institute of Statistical Mathematics*, 45, 621–637.

Chen, S.-X., and Cui, H.-J. (2002), "On Bartlett Correction of Empirical Likelihood in the Presence of Nuisance Parameters," research report, National University of Singapore, Dept. of Statistics and Applied Probability.

Chen, Y., and Chen, H. (2000), "A Unified Approach to Regression Analysis Under Double-Sampling Designs," *Journal of the Royal Statistical Society*, Ser. B, 62, 449–460.

DiCiccio, T., Hall, P., and Romano, J. (1991), "Empirical Likelihood is Bartlett-Correctable," *The Annals of Statistics*, 19, 1053–1061.

Ellenberg, S., and Hamilton, J. (1989), "Surrogate Endpoints in Clinical Trials: Cancer," *Statistics in Medicine*, 8, 405–413.

Fleming, T., and Demets, D. (1996), "Surrogate Endpoints in Clinical Trials: Are We Being Misled?" *Annals of Internal Medicine*, 125, 605–613.

Fleming, T., Prentice, R., Pepe, M., and Glidden, D. (1994), "Surrogate and Auxiliary Endpoints in Clinical Trials: With Potential Applications in Cancer and Aids Research," *Statistics in Medicine*, 13, 955–968.

Hall, P., and La Scala, B. (1990), "Methodology and Algorithms of Empirical Likelihood," *International Statistical Reviews*, 58, 109–127.

Jing, B.-Y. (1995), "Two-Sample Empirical Likelihood Method," *Statistics and Probability Letters*, 24, 315–319.

Kitamura, Y. (1997), "Empirical Likelihood for Weakly Dependent Processes," *The Annals of Statistics*, 25, 2084–2102.

Li, G. (1995a), "Nonparametric Likelihood Ratio Estimation of Probabilities for Truncated Data," *Journal of the American Statistical Association*, 90, 997–1003.

——— (1995b), "On Nonparametric Likelihood Ratio Estimation of Survival Probabilities for Censored-Data," *Statistics and Probability Letters*, 25, 95–104.

Owen, A. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

——— (1991), "Empirical Likelihood for Linear Model," *The Annals of Statistics*, 19, 1725–1747.

——— (2001), *Empirical Likelihood*, Boca Raton, FL: Chapman & Hall/CRC Press.

Pepe, M. (1992), "Inference Using Surrogate Outcome Data and a Validation Sample," *Biometrika*, 79, 355–365.

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Functions," *The Annals of Statistics*, 22, 300–325.

Rotnitzky, A., and Robins, J. (1995), "Semiparametric Regression Estimation in the Presence of Dependent Censoring," *Biometrika*, 82, 805–820.

Venkatraman, E., and Begg, C. (1999), "Properties of a Nonparametric Test for Early Comparison of Treatments in Clinical Trials in the Presence of Surrogate Endpoints," *Biometrics*, 55, 1171–1176.

Wittes, J., Lakatos, E., and Probstfield, J. (1989), "Surrogate Endpoints in Clinical Trials: Cardiovascular Disease," *Statistics in Medicine*, 8, 415–425.