Title	Testing overall and subpopulation treatment effects with measurement errors
Author(s)	Ma, Y; Yin, G
Citation	Statistica Sinica, 2013, v. 23, p. 1019-1042
Issued Date	2013
URL	http://hdl.handle.net/10722/189460
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Statistica Sinica **23** (2013), 1019-1042
doi:http://dx.doi.org/10.5705/ss.2012.049

TESTING OVERALL AND SUBPOPULATION TREATMENT EFFECTS WITH MEASUREMENT ERRORS

Yanyuan Ma and Guosheng Yin

Texas A&M University and The University of Hong Kong

Abstract: There is a growing interest in the discovery of important predictors from many potential biomarkers for therapeutic use. In particular, a biomarker has predictive value for treatment if the treatment is only effective for patients whose biomarker values exceed a certain threshold. However, biomarker expressions are often subject to measurement errors, which may blur the biomarker's predictive capability in patient classification and, as a consequence, may lead to inappropriate treatment decisions. By taking into account the measurement errors, we propose a new testing procedure for the overall and subpopulation treatment effects in the multiple testing framework. The proposed method bypasses the permutation or other resampling procedures that become computationally infeasible in the presence of measurement errors. We conduct simulation studies to examine the performance of the proposed method, and illustrate it with a data example.

Key words and phrases: Biomarker study, clinical trial, measurement error, multiple testing, predictive marker, subgroup analysis, treatment effect.

1. Introduction

To a large extent, decisions on cancer treatment rely upon some specific biomarkers of patients. Due to patient heterogeneity, even for those with the same type of cancer and the same stage of disease, certain biomarkers may be differentially expressed which, in turn, may cause patients to respond differently to the same treatment. Most cancer treatments can be generally classified as cytotoxic agents and cytostatic agents. Cytotoxic agents may benefit patients by directly destroying cancer cells, while cytostatic agents function through inhibiting tumor growth instead of shrinking the tumor. Targeted therapies are typically cytostatic, and they may work only on a subset of the patient population with specific biomarker expression (Korn et al. (2001); Simon (2009); and Ratain and Sargent (2009)). Hence, it is critical to detect the treatment effect not only for the entire population, but also for a subpopulation identified by specific biomarkers. This is especially important in the development of personalized medicine where the treatment is expected to work only for certain subpopulations.

We focus on biomarkers measured at the baseline that would provide information for better treatment decisions. According to their functionalities in diagnosis and treatment selection for cancer patients, biomarkers are considered therapeutically useful if they have prognostic or predictive values. Prognostic markers reflect patients' prognosis such as their health status or disease stages, and they are associated with the disease outcome regardless of the presence of treatment. Predictive markers can predict differential treatment effects for patients belonging to different biomarker groups. In particular, these biomarkers have the capability of predicting which group of patients, say either marker positive or marker negative, is more likely to benefit from the treatment. Several examples of prognostic and predictive biomarkers are described as follows. In breast cancer, estrogen receptor (ER) overexpression may be used as a prognostic marker because ER positive patients have longer survival in the absence of systemic therapy. In addition, ER may also be used as a predictive marker because ER positive patients would benefit from anti-estrogens such as tamoxifen, and ER negativity predicts benefits from some cytotoxic chemotherapies. Also for breast cancer patients, the human epidermal growth factor receptor 2 (HER2) amplification is a predictive marker for treatment benefits from trastuzumab. In colorectal cancer, patients with KRAS mutations appear to be poor candidates for treatment with epidermal growth factor receptor (EGFR) antibodies. For example, cetuximab and panitumumab only benefit colorectal cancer patients with the wild-type KRAS gene status, but not those with mutant KRAS.

In oncology, targeted therapies are often putative or tumor growth-inhibitory agents. From the trial design perspective, the goal is to identify the subpopulation of patients who would benefit from the treatment (Sargent et al. (2005)). Freidlin and Simon (2005) proposed an adaptive signature design to identify sensitive patients through an assay or a signature (the gene-expression classifier). To enhance efficiency of the adaptive signature design, Freidlin, Jiang, and Simon (2010) proposed a cross-validation approach to combining the prospectively developed sensitive patient classifier and the properly powered test to maximize the overall treatment effects. Suppose that we compare an experimental drug with a control, and the experimental treatment only benefits patients with a high level of certain biomarker expression. Nevertheless, we have a continuous measurement for the biomarker expression but no binary classifier to categorize patients into marker positive (the biomarker expression is higher than a threshold), or marker negative (the biomarker expression is lower than a threshold) groups. In this situation, Jiang, Freidlin, and Simon (2007) proposed two procedures for hypothesis testing. Through subgroup analysis, they examined whether the treatment effects are the same or greater in patients with a specific feature or risk factor, so that subsequent marker-specific treatment decisions can be made.

However, biomarker expressions are often measured with errors that prevent us from classifying patients into marker positive or negative groups even if we are given a specific threshold. As naive methods ignoring measurement errors generally result in biased estimation, various methods have been developed to correct such bias; see Carroll et al. (2006) for a comprehensive description. In our case, the mismeasured biomarker value lies inside an indicator function, which is used to identify the "sensitive" patients who would respond to the treatment. Thus, it is in some sense related to measurement errors with change points, for which the estimation and inference are known to be much more difficult than the usual measurement error problems and, to our best knowledge, no approach has been proposed beyond the linear regression context (Ma (2011)). In this paper, we investigate the effects of measurement errors on the classification of patients, propose consistent estimators to handle the measurement errors inside the indicator function, and further develop hypothesis testing for both the entire population and the subpopulations based on the mismeasured biomarker values. The validity of our testing procedure does not require any resampling procedure which may hamper the applicability of the test due to the high computational cost.

The rest of the paper is organized as follows. In Section 2, we briefly review the two methods proposed by Jiang, Freidlin, and Simon (2007) that lay out the general framework for identifying predictive biomarkers. In Section 3, we investigate the situation where the biomarker values are measured with errors. In particular, we propose a new hypothesis testing procedure and an efficient way to estimate the cutoff value for determining marker positive or marker negative groups. In Section 4, we present a simulation study to evaluate the finite-sample properties of the proposed method. We illustrate the new method using the Framingham data in Section 5, and conclude with a brief discussion in Section 6. Technical details are delineated in the Appendix.

2. Identification of Predictive Biomarkers

From a statistical point of view, identification of predictive biomarkers or similar studies along the line, can be formulated as a multiple testing problem. For simplicity, we assume that the primary endpoint is binary; that is, $Y_i = 1$ if patient i has responded to treatment, and $Y_i = 0$ for a nonresponder. Let $Z_i = 1$ if patient i is treated with the experimental treatment, and $Z_i = 0$ if the subject is treated with the standard treatment; and let X_i denote the biomarker expression. We model Y_i with respect to both Z_i and X_i via the generalized linear model,

$$\Pr(Y_i = 1|X_i, Z_i) = G(\beta_1 + \beta_2 X_i + \beta_3 Z_i), \tag{2.1}$$

where the link function may be logit $G(u) = e^u/(1+e^u)$, or probit $\Phi(u)$, with $\Phi(u)$ denoting the cumulative distribution function of the standard normal distribution.

Model (2.1) characterizes the effectiveness of the overall treatment for all the patients as reflected by testing whether β_3 is zero. However, the treatment may work only for a subpopulation with biomarker expression satisfying $X_i > c$, where c is an unknown cutoff value, called a cutpoint of the biomarker. As a result, we may fit a more flexible model,

$$\Pr(Y_i = 1|X_i, Z_i) = G\{\beta_1 + \beta_2 X_i + \beta_3 Z_i I(X_i > c)\}. \tag{2.2}$$

Even if the treatment is not effective for all patients, it may still be effective for a subset of patients whose biomarker values are larger than c, as reflected by $\beta_3 \neq 0$ in model (2.2). Without loss of generality, we assume that the range of possible biomarker values is $0 \leq c \leq 1$. When $c = 0$, model (2.2) reduces to model (2.1). In the targeted therapy development, we need to determine whether

(i) $\beta_3 \neq 0$ in model (2.1),

(ii) $\beta_3 \neq 0$ in model (2.2) for at least one cutpoint c, and

(iii) $\beta_3 = 0$ in both (2.1) and (2.2).

Toward these goals, Jiang, Freidlin, and Simon (2007) proposed procedures A and B that we outline as follows.

Procedure A splits the problem into two separate hypothesis tests. The first test concerns (2.1), with the null and alternative hypotheses given by

$$H_0 : \beta_3 = 0 \quad \text{versus} \quad H_1 : \beta_3 \neq 0.$$

The second test concerns model (2.2), with the same null and alternative hypotheses. If the overall type I error rate is set at α, a typical choice is to split α in a ratio of 4:1 between the two tests, $\alpha_1 = 0.8\alpha$ for the first test and $\alpha_2 = 0.2\alpha$ for the second. If the first test rejects H_0 at a significance level of α_1, the procedure terminates and the treatment is deemed to be effective for all the patients as in case (i). If the first test fails to reject H_0, the second test is subsequently performed at a significance level of α_2. If the second test rejects H_0, then the treatment is deemed to be effective only for a subset of patients whose biomarker expression levels exceed a certain cutpoint as in case (ii). If the second test also fails to reject H_0, then the treatment is considered ineffective at all as in case (iii). For the second hypothesis testing, we may construct a test statistic $S(c)$ for a given cutpoint c, and take the final test statistic as $T_A = \max_{0<c\leq 1} S(c)$.

Procedure B is more sophisticated. It aims at testing $H_0 : \beta_3 = 0$ for all c in model (2.2) versus $H_1 : \beta_3 \neq 0$ for at least one value of c. The test statistic is

constructed as $T_B = \max\{S(0) + R,\ \max_{0<c\leq 1} S(c)\}$, where R is a prespecified constant to upweigh the contribution of the overall treatment effect in order to achieve a suitable balance. It is typically recommended that $R = 2.2$, the difference between the 95th and 80th percentiles of the chi-squared distribution with one degree of freedom. If H_0 is not rejected, then the treatment is considered to be ineffective for any subgroup as in case (iii). If H_0 is rejected, then the treatment is considered to work for at least a subset of patients depending on the value of c, while this subset of patients could also be the entire population if $c = 0$.

In the second test of procedure A and in procedure B of Jiang, Freidlin, and Simon (2007), the test statistics T_A and T_B take the maximum over a range of values of c, and thus it is difficult to derive their asymptotic null distributions. As outlined in the Appendix, a bootstrap-based procedure can be used to calculate the p-value. If both cases (i) and (iii) are excluded in procedure A, or if case (iii) is excluded in procedure B, the next step is to find an appropriate cutpoint c, in order to identify which subset of patients would be suitable for the treatment. The cutpoint c can be estimated by maximizing the profile likelihood

$$PL(c) = \prod_{i=1}^{n} G\{\hat{\beta}_1(c) + \hat{\beta}_2(c)X_i + \hat{\beta}_3(c)Z_i I(X_i > c)\}^{Y_i} \times [1 - G\{\hat{\beta}_1(c) + \hat{\beta}_2(c)X_i + \hat{\beta}_3(c)Z_i I(X_i > c)\}]^{1-Y_i},$$

where $\hat{\boldsymbol{\beta}}(c) = (\hat{\beta}_1(c), \hat{\beta}_2(c), \hat{\beta}_3(c))^{\mathrm{T}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}}$ for a fixed value of c.

3. Proposed Methods

3.1. Biomarkers with measurement errors

The procedures A and B in Jiang, Freidlin, and Simon (2007) are easy to implement if the biomarker value can be measured precisely. However, it is known that biomarker values are usually subject to measurement errors. Instead of observing the true biomarker value X_i for subject i, we might observe a surrogate value $W_i = X_i + U_i$, where U_i is the error incurred in the measurement of X_i. Typically, U_i is assumed to follow a normal distribution with mean zero and variance σ_U^2. In practice, σ_U can be estimated using repeated measurements or validation observations; for simplicity we take σ_U as known provisionally. Under the assumption of normal measurement errors and a logit link function $G(\cdot)$, we can derive a consistent estimator for $\boldsymbol{\beta}$ by applying the conditional score method to model (2.1). Specifically, following Stefanski and Carroll (1987), for the first

test in procedure A, we have

$$\sum_{i=1}^{n}[Y_i - G\{\beta_1 + \beta_2(W_i + Y_i\sigma_U^2\beta_2) + \beta_3 Z_i - \beta_2^2\frac{\sigma_U^2}{2}\}]\begin{pmatrix} 1 \\ W_i + Y_i\sigma_U^2\beta_2 \\ Z_i \end{pmatrix} = \mathbf{0}.$$

Consequently, we can construct a Wald test statistic, and conduct hypothesis testing at the desired significance level α_1 in a straightforward way. However, in the presence of measurement errors, the second test of procedure A and the test in procedure B are nontrivial because the covariate subject to measurement errors lies inside an indicator function in model (2.2), and the traditional methods in Carroll et al. (2006) are not directly applicable. Although regression calibration (Carroll and Stefanski (1990) and Gleser (1990)) or SIMEX (Cook and Stefanski (1995)) can still be used, these methods are approximate and typically do not provide consistent estimation.

Following the semiparametric approach of Tsiatis and Ma (2004), we can construct the estimating equation for $\boldsymbol{\beta}$ for a fixed value of c,

$$\sum_{i=1}^{n}\boldsymbol{\phi}(W_i, Y_i, Z_i; \boldsymbol{\beta}) = \mathbf{0},$$

where

$$\begin{aligned}\boldsymbol{\phi}(W_i, Y_i, Z_i; \boldsymbol{\beta}) &= \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, Z_i; \boldsymbol{\beta}) - E^*\{\mathbf{a}(X_i, Z_i; \boldsymbol{\beta})|W_i, Y_i, Z_i\} \\ \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, Z_i; \boldsymbol{\beta}) &= E^*\{\mathbf{S}^F_{\boldsymbol{\beta}}(X_i, Y_i, Z_i; \boldsymbol{\beta})|W_i, Y_i, Z_i\},\end{aligned}$$

$\mathbf{S}^F_{\boldsymbol{\beta}}(X_i, Y_i, Z_i; \boldsymbol{\beta})$ is the score function of the logistic model in (2.2), and $\mathbf{a}(X_i, Z_i; \boldsymbol{\beta})$ satisfies

$$E\{\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, Z_i; \boldsymbol{\beta})|X_i, Z_i\} = E[E^*\{\mathbf{a}(X_i, Z_i; \boldsymbol{\beta})|W_i, Y_i, Z_i\}|X_i, Z_i].$$

The evaluation of expectation $E(\cdot|W_i, Y_i, Z_i)$ requires the probability density function of X_i, which is not available. Our procedure replaces $E(\cdot|W_i, Y_i, Z_i)$ with $E^*(\cdot|W_i, Y_i, Z_i)$, calculated under a proposal density function of X_i. The proposal model could be misspecified, yet the resulting estimator is still consistent due to the fact that

$$\begin{aligned}&E\{\boldsymbol{\phi}(W_i, Y_i, Z_i; \boldsymbol{\beta})\} \\ &\quad= E[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, Z_i; \boldsymbol{\beta}) - E^*\{\mathbf{a}(X_i, Z_i; \boldsymbol{\beta})|W_i, Y_i, Z_i\}] \\ &\quad= E\big(E\{\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, Z_i; \boldsymbol{\beta})|X_i, Z_i\} - E[E^*\{\mathbf{a}(X_i, Z_i; \boldsymbol{\beta})|W_i, Y_i, Z_i\}|X_i, Z_i]\big) \\ &\quad= \mathbf{0}.\end{aligned}$$

In addition, the estimator $\hat{\boldsymbol{\beta}}$ has a root-n rate and the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has a typical sandwich form of $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}}$, where

$$\mathbf{A} = E\left\{\frac{\partial\boldsymbol{\phi}(W_i, Y_i, Z_i; \boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\} \quad \text{and} \quad \mathbf{B} = \mathrm{Cov}\{\boldsymbol{\phi}(W_i, Y_i, Z_i; \boldsymbol{\beta})\}.$$

For a more elaborate description of estimating equation approaches to measurement error models and sandwich variance estimation, see Tsiatis and Ma (2004).

Based on $\hat{\boldsymbol{\beta}}$ and the estimated variance-covariance matrix $\hat{\mathbf{V}}$, a Wald test statistic may be constructed as $S(c) = n\hat{\beta}_3^2(c)/\hat{V}_{33}(c)$, where $\hat{V}_{33}(c)$ is the $(3,3)$ entry of $\hat{\mathbf{V}}$. We can further define $T_A = \max_{0<c\leq 1} S(c)$ for the second test in procedure A or $T_B = \max\{S(0) + R,\ \max_{0<c\leq 1} S(c)\}$ in procedure B. The p-value can be calculated following the same bootstrap procedure as described in the Appendix.

Although the adaption of the testing procedure A or B from the error-free case to the measurement-error case is conceptually achievable, the computation is extremely intense. For each fixed cutpoint value c, we need to apply the Newton-Raphson algorithm to estimate the parameter $\boldsymbol{\beta}$. Within each iteration of the Newton-Raphson procedure, we must evaluate the estimating equation and its derivative, which requires solving an integral equation to obtain the function $\mathbf{a}(X, Z; \boldsymbol{\beta})$. Since the lower and upper quantiles of the null distribution for the test statistic are needed for hypothesis testing via the bootstrap, this in turn increases the computational burden. In what follows, we develop a different testing procedure based on the asymptotic distribution of a new test statistic that alleviates the computational effort.

3.2. Hypothesis testing on subpopulations

In the implementation of procedure B of Jiang, Freidlin, and Simon (2007), the test statistic is to be $T_K = \max\{S(0) + R,\ \max_{c_2,\ldots,c_K} S(c)\}$ for a set of candidate values of c, $\mathbf{c} = \{c_1 \equiv 0, c_2, \ldots, c_K\}$. The difference between the original T_B and the implemented T_K is that T_B is based on all values of c continuously for $0 \leq c \leq 1$, and thus it is used to test no treatment effect at any value of c in the range of $[0, 1]$, while T_K is used to test no treatment effect at a set of specific values of c, namely $\mathbf{c} = \{c_1 \equiv 0, c_2, \ldots, c_K\}$. We observe that in model (2.2) if (and only if) the treatment is effective for the true cutpoint c_0, it is effective for any other cutpoint c. We illustrate this point. First, we consider that $H_1 : \beta_3 \neq 0$ is true in a two-arm study. Note that in model (2.2) β_1 represents the treatment effect of the standard arm. Figure 1 shows that the intercept is indeed shifted by β_3 in the subset $\{(X_i, Z_i) : X_i > c_0 \text{ and } Z_i = 1\}$ from its complementary region, and the intercept of the set $\{(X_i, Z_i) : X_i > c\ (c \neq c_0),\ \text{and } Z_i = 1\}$ is also shifted. As an illustration, suppose $\beta_3 > 0$, and take $c > c_0$. In this case, β_1 calculated under c is larger than that obtained under c_0, since it is inflated by the set $\{(X_i, Z_i) : c > X_i > c_0 \text{ and } Z_i = 1\}$, while it is still smaller than $\beta_1 + \beta_3$ because the effect of $X_i > c$ and $Z_i = 1$ does not contribute to the calculation of β_1. This means that β_3, calculated under c, is still positive, although it is smaller than that obtained under c_0 as shown in the second plot of Figure 1. On

the other hand, if we take $c < c_0$, then the value of β_1 remains the same under c or under c_0, while β_3, calculated under c, would be smaller than that under c_0, because some of the "irrelevant" X_i values would also be considered "relevant", and thus the treatment effect is diluted; see the third plot in Figure 1. Second, we consider the case that $H_0 : \beta_3 = 0$ is true; so the treatment is not effective for any cutpoint value of c, including those in the chosen set $\mathbf{c} = \{c_1 = 0, c_2, \ldots, c_K\}$. Thus, using the candidate cutpoint set to form the test statistic still yields a consistent testing procedure. The difference between T_B and T_K is only reflected in statistical power; power would be less if we take the value of c to be different from the true cutpoint value c_0.

If there is no treatment effect, then there is no treatment effect for any cutpoint value; so $\boldsymbol{\beta}_3(\mathbf{c}) \equiv \{\beta_3(c_1), \ldots, \beta_3(c_K)\}^{\mathrm{T}} = \mathbf{0}$ for a chosen set of candidate cutpoint values $\{c_1, \ldots, c_K\}$. Specifically, we can estimate $\boldsymbol{\beta}_3(\mathbf{c})$ and its variance-covariance matrix $\mathbf{V}_3$ using the method described in Section 3. Based on the sample estimates $\hat{\boldsymbol{\beta}}_3(\mathbf{c})$ and $\hat{\mathbf{V}}_3$, the test statistic is constructed as $T(\mathbf{c}) = \hat{\boldsymbol{\beta}}_3(\mathbf{c})^{\mathrm{T}} \hat{\mathbf{V}}_3^{-1} \hat{\boldsymbol{\beta}}_3(\mathbf{c})$, which is chi-squared with K degrees of freedom under the null hypothesis. This new test becomes a standard hypothesis testing problem that does not involve multiple comparisons. It allows us to compute the p-value and conduct hypothesis testing without the need to resort to a bootstrap procedure.

3.3. Estimation of biomarker cutpoint

If the test concludes that the treatment is effective for a subset of patients with biomarker values satisfying $X > c$, the next step is to determine the cutpoint c. When the biomarker value is measured precisely, c can be estimated by maximizing the profile likelihood. However, the same procedure cannot be applied when X is measured with errors because the likelihood cannot be derived without assuming a specific distribution for X. Instead, we can estimate c by maximizing the Wald test statistic at a fixed cutpoint value c,

$$\hat{c} = \arg\max_c \frac{\hat{\beta}_3^2(c)}{\hat{V}_{33}(c)}.$$

In fact, it is more convenient to use the score test statistic. For ease of exposition, we denote the first two components of $\boldsymbol{\phi}$ as $\boldsymbol{\phi}_a$ and the last component of $\boldsymbol{\phi}$ as ϕ_3, and similarly $\boldsymbol{\beta}_a = (\beta_1, \beta_2)^{\mathrm{T}}$. Let

$$\hat{U}(c) = n^{-1/2} \sum_{i=1}^{n} \phi_3(W_i, Z_i, Y_i; \widetilde{\boldsymbol{\beta}}_a, 0, c),$$

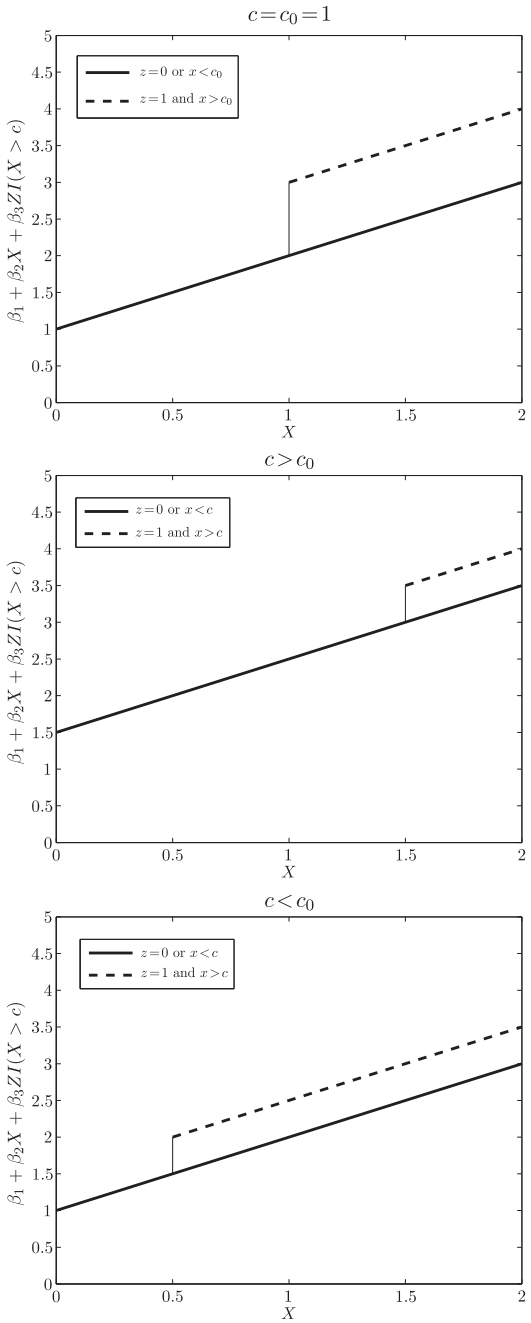

Figure 1. Illustration on the changes of the regression lines when a candidate cutpoint instead of the true cutpoint is used under the alternative H_1. The top panel is based on the true cutpoint c_0, the middle panel uses a candidate cutpoint $c > c_0$, and the bottom panel uses a candidate cutpoint $c < c_0$.

where $\widetilde{\boldsymbol{\beta}}_a$ is the estimator of $\boldsymbol{\beta}_a$ under the null hypothesis $H_0 : \beta_3 = 0$ in model (2.2), so it does not depend on c. We denote the ith observation as $\mathcal{O}_i = (W_i, Z_i, Y_i)$, and let

$$\begin{aligned}
\mathbf{A}_1 &= E\left\{\frac{\partial \boldsymbol{\phi}_a(\mathcal{O}_i, \boldsymbol{\beta}_a, 0)}{\partial \boldsymbol{\beta}_a^{\mathrm{T}}}\right\}\Bigg|_{\boldsymbol{\beta}_a=\boldsymbol{\beta}_{a0}},\\
\mathbf{A}_2(c) &= E\left\{\frac{\partial \phi_3(\mathcal{O}_i, \boldsymbol{\beta}_a, 0, c)}{\partial \boldsymbol{\beta}_a^{\mathrm{T}}}\right\}\Bigg|_{\boldsymbol{\beta}_a=\boldsymbol{\beta}_{a0}},\\
v(c) &= \mathrm{Var}\{\phi_3(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c) - \mathbf{A}_2(c)\mathbf{A}_1^{-1}\boldsymbol{\phi}_a(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0)\},
\end{aligned}$$

where $\boldsymbol{\beta}_{a0}$ denotes the true value of $\boldsymbol{\beta}_a$. For a fixed value of c, these quantities can be calculated by replacing expectations and covariance matrices with their empirical counterparts and inserting the estimator $\widetilde{\boldsymbol{\beta}}_a$. If we denote the estimate of $v(c)$ by $\hat{v}(c)$, we can estimate c by maximizing $\hat{U}^2(c)/\hat{v}(c)$. In the Appendix, we show that the variance of $\hat{c}$ can be estimated by $\hat{\sigma}_c^2 = \hat{\xi}^{-2}\widehat{\mathrm{Var}}(\zeta_i)/n$, evaluated at $(\widetilde{\boldsymbol{\beta}}_a, 0, \hat{c})$, where ξ and ζ_i are given in (A.1) and (A.2), respectively. Both ξ and ζ_i involve the first or second derivative of ϕ_3 with respect to c. However, ϕ_3 is not a continuous function of c, which makes the variance estimation for the cutpoint very difficult. This is similar to the situation encountered in quantile regression, yet here the issue of discontinuity is more severe and we resort to a bootstrap procedure to estimate the variance of $\hat{c}$.

After obtaining $\hat{c}$, we can proceed to estimate $\boldsymbol{\beta}$ using the semiparametric estimation procedure described in Section 3. In addition to the original variance-covariance matrix while assuming c fixed, an extra source of variation should be considered due to estimating c, and thus the variance-covariance matrix of $\hat{\boldsymbol{\beta}}(\hat{c})$ is estimated by

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}(\hat{\mathbf{A}}^{-1})^{\mathrm{T}} + \hat{\sigma}_c^2\left\{\frac{\partial \hat{\boldsymbol{\beta}}(c)}{\partial c}\right\}^{\mathrm{T}}\frac{\partial \hat{\boldsymbol{\beta}}(c)}{\partial c},$$

where c is evaluated at $\hat{c}$. Note that estimation of a cutpoint is a very difficult problem due to the discontinuity caused by the indicator function. This is true when the covariates are measured precisely, for example, see Luo, Turnbull, and Clark (1997), Pons (2003), and Kosorok and Song (2007) for estimation of a changepoint in survival models. The problem becomes even more challenging when the covariates are measured with errors.

4. Simulation Study

We conducted extensive simulation studies to evaluate the performance of the proposed estimation and testing procedure. The true parameter values were

$\beta_1 = -1.5$ and $\beta_2 = 1.0$ under the null model, and were $\beta_1 = -1.5$, $\beta_2 = 1.5$ and $\beta_3 = 1.0$ with the true cutpoint $c_0 = 1.0$ under the alternative model. The covariate X_i was uniform on $[0, 3]$, and the measurement error was normal with a standard deviation of 0.1732, which corresponds to the noise-to-signal ratio of 20%. The sample size is typically large in these studies in order to identify a responsive subset of subjects, as such a subset may be small, say, 20% or 30% of the total population. We took sample sizes $n = 500$ and 1,000, and replicated 1,000 data sets.

In the implementation of the proposed estimation and testing procedure, we posited the distribution of the true biomarker expression X to be either uniform, normal, or exponential, corresponding to the true and two misspecified cases, respectively. We took the set of candidate cutpoint values $\mathbf{c} = \{0, 0.6, 1.2, 1.8, 2.4\}$ that does not include the true cutpoint $c_0 = 1.0$. For comparison, we also present simulation results using a naive approach, for which the measurement errors are completely ignored. In the absence of measurement errors, the estimation procedure reduces to that of Jiang, Freidlin, and Simon (2007).

The upper panel of Table 1 summarizes the results for the type I error rate and power. We can see that the proposed tests are generally consistent, as reflected by the closeness between the sample proportions of rejecting the null hypothesis and the corresponding nominal levels. Even when $f^*(x)$, the assumed distribution of X, is misspecified as a normal or exponential distribution, the type I error rates are still maintained at the nominal levels. By contrast, the naive testing procedure is severely biased, leading to an inflation of the type I error rate. Because the estimation variability is inevitably higher with measurement errors, some power loss is incurred for the proposed method compared with its naive counterpart. To gain more insight into the influence of selection of the candidate cutpoint set on power, we further explored the case that the true cutpoint $c_0 = 1.0$ happened to be included in the candidate set by taking the candidate set $\mathbf{c} = \{0, 0.6, 1.0, 1.8, 2.4\}$. Comparing the lower and upper panels of Table 1, we can see that there is indeed some power loss when the true cutpoint is not included in the candidate cutpoint set. However, the power loss is not substantial, in general within the 5% difference. This is certainly encouraging, as the true cutpoint is never known in practice.

Tables 2 and 3 show the parameter estimates when the true cutpoint c_0 is not contained in the candidate set, with the sample size $n = 500$ and 1,000, respectively. Clearly, the estimates using the naive method are biased which, in turn, causes inconsistency of the naive testing procedure. The biases are especially large for the intercept term under the null hypothesis, and for both the intercept and slope terms under the alternative hypothesis. By contrast, the proposed method provides consistent estimates of model parameters with

Table 1. Levels of precision and power of the test with sample size $n = 500$ and $1,000$. The true cutpoint c_0 is or is *not* included in the candidate set, and the proposal distribution for the mismeasured covariate is uniform (the true model), normal, and exponential, respectively.

	Test size under H_0				Power under H_1			
Nominal level	0.01	0.05	0.1	0.15	0.01	0.05	0.1	0.15
$n = 500$, c_0 is *not* in the candidate set								
Uniform $f^*(x)$	0.023	0.057	0.100	0.153	0.289	0.509	0.640	0.718
Normal $f^*(x)$	0.022	0.055	0.096	0.143	0.272	0.477	0.590	0.673
Exponential $f^*(x)$	0.023	0.055	0.099	0.149	0.281	0.517	0.635	0.711
Naive method	0.016	0.073	0.134	0.177	0.336	0.579	0.704	0.780
$n = 1000$, c_0 is *not* in the candidate set								
Uniform $f^*(x)$	0.009	0.053	0.099	0.139	0.600	0.809	0.881	0.931
Normal $f^*(x)$	0.009	0.050	0.098	0.134	0.589	0.787	0.862	0.908
Exponential $f^*(x)$	0.009	0.053	0.104	0.142	0.628	0.820	0.882	0.920
Naive method	0.018	0.070	0.135	0.191	0.777	0.914	0.959	0.972
$n = 500$, c_0 is in the candidate set								
Uniform $f^*(x)$	0.022	0.067	0.110	0.154	0.309	0.533	0.643	0.717
Normal $f^*(x)$	0.023	0.066	0.112	0.144	0.323	0.518	0.648	0.731
Exponential $f^*(x)$	0.020	0.067	0.117	0.170	0.311	0.533	0.648	0.731
Naive method	0.018	0.074	0.123	0.178	0.389	0.640	0.770	0.827
$n = 1000$, c_0 is in the candidate set								
Uniform $f^*(x)$	0.011	0.059	0.105	0.147	0.637	0.818	0.900	0.933
Normal $f^*(x)$	0.013	0.058	0.106	0.152	0.628	0.828	0.899	0.937
Exponential $f^*(x)$	0.014	0.064	0.109	0.155	0.662	0.859	0.913	0.942
Naive method	0.019	0.071	0.122	0.174	0.836	0.939	0.967	0.980

negligible biases. In addition, the standard errors for the estimates of $\boldsymbol{\beta}$ are quite close to the standard deviations, and the coverage probabilities of the 95% confidence intervals are reasonably close to the nominal level. As the sample size increases, the estimation, particularly that for β_3, is much improved. For the cutpoint c, the point estimate is consistent with small bias, while the variance is often over-estimated because the estimation of c is extremely difficult due to the discontinuity caused by the indicator function.

We further experimented the situations where the true cutpoint happened to be included in the candidate set by considering the candidate set $\mathbf{c} = \{0, 0.6, 1.0, 1.8, 2.4\}$, and repeated the simulations as before. Tables 4 and 5 correspond to the results with sample sizes $n = 500$ and 1,000, respectively. The parameter and variance estimates, as well as the coverage probabilities, are similar to those when the candidate set does not contain the true cutpoint.

Table 2. Parameter estimation with sample size $n = 500$ when the true cutpoint c_0 is *not* contained in the candidate set. The median of the parameter estimates is $\hat{\beta}$, the empirical standard deviation is SD, the median of the estimated standard errors is SE, and the coverage probability of 95% confidence intervals is CP, in percentage.

	Estimation under H_0		Estimation under H_1			
True values	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_3 = 1.0$	$c = 1.0$
Uniform $f^*(x)$						
$\hat{\beta}$	−1.5057	1.0049	−1.5062	0.9746	1.2494	1.0237
SD	0.1945	0.1454	0.2469	0.2708	0.4671	0.1766
SE	0.1871	0.1382	0.2529	0.2823	0.5353	0.2695
CP(%)	94.7	94.5	96.7	96.0	98.0	97.4
Normal $f^*(x)$						
$\hat{\beta}$	−1.5057	1.0047	−1.5105	0.9857	1.2125	1.0187
SD	0.1935	0.1452	0.2360	0.2668	0.4538	0.1746
SE	0.1871	0.1382	0.2541	0.2827	0.5339	0.2791
CP(%)	94.9	94.4	96.9	96.7	97.8	97.3
Exponential $f^*(x)$						
$\hat{\beta}$	−1.5052	1.0045	−1.4974	0.9736	1.2228	0.9879
SD	0.1935	0.1456	0.2412	0.2711	0.4486	0.1786
SE	0.1862	0.1373	0.2527	0.2799	0.5341	0.2783
CP(%)	94.9	94.2	95.9	96.4	97.6	98.4
Naive method						
$\hat{\beta}$	−1.4523	0.9566	−1.4116	0.8657	1.2862	1.0319
SD	0.1866	0.1372	0.2095	0.2214	0.3606	0.1292
SE	0.1808	0.1309	0.2235	0.2323	0.3879	0.2091
CP(%)	92.7	93.1	94.3	91.0	93.3	98.6

5. Example

We applied the proposed method to the data from the Framingham study. This study contains 1,615 subjects, and the response of interest Y is the occurrence of coronary heart disease, a value of 1 or 0 indicating whether a subject has the disease or is free of it. One of the study objectives is to characterize how coronary heart disease is affected by the long-term average of systolic blood pressure (covariate X). Based on the literature, it is known that X is subject to an additive measurement error that is normally distributed with mean zero and standard deviation 0.08 (Carroll et al. (2006)). Previous studies have established that coronary heart disease is also related to smoking status (denoted as Z), in that smokers are more likely to develop heart disease. We are interested in determining whether there is a "safe" zone for the blood pressure region in which, as long as a subject's average blood pressure is below a certain threshold, smoking

Table 3. Parameter estimation with sample size $n = 1,000$ when the true cutpoint c_0 is *not* contained in the candidate set. The median of the parameter estimates is $\hat{\beta}$, the empirical standard deviation is SD, the median of the estimated standard errors is SE, and the coverage probability of 95% confidence intervals is CP, in percentage.

	Estimation under H_0		Estimation under H_1			
True values	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_3 = 1.0$	$c = 1.0$
Uniform $f^*(x)$						
$\hat{\beta}$	−1.5053	1.0032	−1.5074	0.9919	1.0881	0.9816
SD	0.1288	0.0955	0.1662	0.1844	0.2817	0.1529
SE	0.1316	0.0969	0.1759	0.1933	0.3035	0.2015
CP(%)	96.4	96.6	96.4	96.1	97.9	96.1
Normal $f^*(x)$						
$\hat{\beta}$	−1.5048	1.0027	−1.5098	0.9910	1.0871	0.9861
SD	0.1293	0.0958	0.1738	0.1908	0.2893	0.1596
SE	0.1316	0.0968	0.1759	0.1938	0.3038	0.2046
CP(%)	96.3	96.5	94.8	94.9	97.8	96.3
Exponential $f^*(x)$						
$\hat{\beta}$	−1.5051	1.0029	−1.5095	0.9938	1.0920	0.9594
SD	0.1287	0.0954	0.1678	0.1863	0.2851	0.1650
SE	0.1316	0.0968	0.1761	0.1945	0.3046	0.2037
CP(%)	96.4	96.7	96.5	97.0	97.8	96.5
Naive method						
$\hat{\beta}$	−1.4524	0.9556	−1.4208	0.8886	1.1520	1.0075
SD	0.1244	0.0905	0.1551	0.1570	0.2300	0.1198
SE	0.1273	0.0918	0.1569	0.1601	0.2500	0.1737
CP(%)	93.6	92.5	91.6	89.3	94.2	98.2

does not elevate the chance of developing heart disease. In other words, we aim to identify a subpopulation based on the measurement of the blood pressure such that the risk of developing heart disease for those subjects is not elevated even if they smoke. This is particularly relevant for smokers trying to find an "excuse" for their smoking behavior. For the purposes of modeling, we can view smoking behavior as a "treatment" indicator and blood pressure as a biomarker. If there is such a subpopulation, we would be interested in finding the cutpoint on blood pressure below which smoking does not make any difference in the risk of heart disease.

To implement our proposed procedure, we used five ($K = 5$) candidate cutpoint values equally spaced along the supporting range of the observed blood pressure values, $3.5528 \leq W \leq 5.2426$, with the true blood pressure X unobserved. Based on model (2.2), we obtained the test statistic $T_5 = 12.02$, which

Table 4. Parameter estimation with sample size $n = 500$ when the true cutpoint c_0 is contained in the candidate set. The median of the parameter estimates is $\hat{\beta}$, the empirical standard deviation is SD, the median of the estimated standard errors is SE, and the coverage probability of 95% confidence intervals is CP, in percentage.

	Estimation under H_0		Estimation under H_1			
True values	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_3 = 1.0$	$c = 1.0$
Uniform $f^*(x)$						
$\hat{\beta}$	−1.5052	1.0043	−1.5146	0.9931	1.2155	1.0200
SD	0.1957	0.1467	0.2462	0.2707	0.4748	0.1763
SE	0.1863	0.1373	0.2527	0.2825	0.5535	0.2763
CP(%)	94.5	94.3	95.6	96.3	97.6	97.4
Normal $f^*(x)$						
$\hat{\beta}$	−1.5052	1.0040	−1.5012	0.9746	1.2420	1.0217
SD	0.1941	0.1463	0.2283	0.2565	0.4580	0.1794
SE	0.1861	0.1373	0.2516	0.2796	0.5284	0.2624
CP(%)	94.8	94.6	97.2	97.6	98.0	98.1
Exponential $f^*(x)$						
$\hat{\beta}$	−1.5063	1.0054	−1.4974	0.9736	1.2228	0.9879
SD	0.1938	0.1454	0.2412	0.2711	0.4486	0.1786
SE	0.1863	0.1374	0.2527	0.2799	0.5341	0.2783
CP(%)	94.8	94.3	95.9	96.4	97.6	98.4
Naive method						
$\hat{\beta}$	−1.4519	0.9562	−1.3984	0.8524	1.2952	1.0302
SD	0.1887	0.1387	0.2117	0.2211	0.3398	0.1224
SE	0.1800	0.1300	0.2203	0.2297	0.3824	0.2043
CP(%)	92.5	92.7	93.1	90.6	94.2	98.7

exceeds the critical constant 11.07, the 95th percentile of the chi-squared distribution with five degrees of freedom. As a result, we reject the null hypothesis and conclude that there exists a cutpoint on the blood pressure, above which smoking has an effect on the development of heart disease. We further performed estimation under the alternative model in (2.2) and obtained the estimates of the regression coefficients $\beta_1 = -14.69$, $\beta_2 = 2.68$, and $\beta_3 = 0.53$, with the corresponding estimated variances 3.49, 0.170, and 0.0636. The cutpoint is estimated as $\hat{c} = 3.1$, which is to the left of the entire region of the observed blood pressure $[3.5528, 5.2426]$. We emphasize here that whether $\hat{c}$ is also to the left of the region of the true blood pressure is unknown because X is not observed. Nevertheless, our finding indicates that although smoking status has an effect only in the region above the cutpoint $\hat{c} = 3.1$, this region contains all the observed values of blood pressure. Hence there is indeed no "safe" zone in terms of the observed

Table 5. Parameter estimation with sample size $n = 1,000$ when the true cutpoint c_0 is contained in the candidate set. The median of the parameter estimates is $\hat{\beta}$, the empirical standard deviation is SD, the median of the estimated standard errors is SE, and the coverage probability of 95% confidence intervals is CP, in percentage.

	Estimation under H_0		Estimation under H_1			
True values	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_1 = -1.5$	$\beta_2 = 1.0$	$\beta_3 = 1.0$	$c = 1.0$
Uniform $f^*(x)$						
$\hat{\beta}$	−1.5048	1.0027	−1.5125	0.9986	1.0837	0.9919
SD	0.1292	0.0957	0.1698	0.1899	0.2978	0.1524
SE	0.1316	0.0968	0.1759	0.1944	0.3054	0.2001
CP(%)	96.3	96.5	96.1	95.6	96.8	95.7
Normal $f^*(x)$						
$\hat{\beta}$	−1.5045	1.0021	−1.5081	0.9892	1.0955	0.9926
SD	0.1291	0.0957	0.1756	0.1925	0.2851	0.1577
SE	0.1316	0.0968	0.1761	0.1942	0.3067	0.2066
CP(%)	96.3	96.5	94.8	95.5	98.6	96.3
Exponential $f^*(x)$						
$\hat{\beta}$	−1.5050	1.0030	−1.5095	0.9938	1.0920	0.9594
SD	0.1280	0.0951	0.1678	0.1863	0.2851	0.1650
SE	0.1317	0.0968	0.1761	0.1945	0.3046	0.2037
CP(%)	96.4	96.8	96.5	97.0	97.8	96.5
Naive method						
$\hat{\beta}$	−1.4517	0.9546	−1.4192	0.8865	1.1600	1.0126
SD	0.1246	0.0908	0.1566	0.1597	0.2247	0.1175
SE	0.1273	0.0918	0.1562	0.1601	0.2509	0.1710
CP(%)	93.6	92.3	91.2	89.0	94.0	97.8

blood pressure levels for smoking. This is important information to motivate all smokers to quit smoking.

6. Discussion

We have developed a new testing procedure for treatment effects in either an overall population or a subpopulation identified by some biomarker that may be measured with errors. The proposed method differs from that of Jiang, Freidlin, and Simon (2007) in three important aspects. First, our testing procedure does not require the estimation of the true cutpoint c_0. Instead, it requires one to prespecify a set of candidate cutpoints and to construct a Wald test statistic. Second, the p-value can be calculated based on the asymptotic distribution of the test statistic, in contrast to using a bootstrap procedure that is computationally infeasible in the presence of measurement errors. Third, due to the difficulty

in constructing a likelihood in measurement error models, we propose to maximize the score-type statistic to estimate the cutpoint instead of using the profile likelihood approach.

One of the most interesting discoveries is that whether the set of candidate cutpoints contains the true cutpoint or not, the test remains consistent, and the power of detecting the treatment effect tends to one as the sample size increases. This is a direct consequence of the root-n rate of the estimation procedure in Section 3. However, in practice the selection of the candidate cutpoints has an impact on power. This means that if the alternative model holds, a properly chosen set can increase the power. In the ideal case, the best candidate set is the one that contains only the true cutpoint c_0. Of course the true cutpoint is unknown, hence we suggest including several reasonable cutpoints in the set based on scientific knowledge. When no such scientific information is available, a natural choice is to include a series of equally spaced candidate cutpoints. Although choosing more candidate cutpoints can increase the chance of capturing the true one and thus improve power, it also results in a larger number of degrees of freedom for the chi-squared test and thus diminishes power. Hence there is a delicate balance in choosing the number of cutpoints. In order to capture the potential global treatment effect, we suggest always including the lower limit of the range of W in the set of cutpoints. But we generally do not recommend including the upper limit of the range of W, because it can cause degeneration of model (2.2) if the upper limit point happens to lie out of the range of X. Based on our experience, choosing between two to five candidate cutpoints generally works well. However, establishing a theoretically justifiable method for the choice of candidate cutpoints requires more research involving higher order asymptotic properties. Finally, although the procedure is motivated from handling measurement errors, the proposed method is certainly applicable when biomarkers are measured precisely.

Acknowledgements

We would like to thank the referees, an associate editor, and Editor Naisyin Wang for many insightful suggestions which strengthened the work immensely. This work is partially supported by US NSF and NINDS grants (Y. Ma) and a grant (Grant No. 784010) from the Research Grants Council of Hong Kong (G. Yin).

Appendix

A.1. Bootstrap approximation of the p-value

Because the test statistic in Jiang, Freidlin, and Simon (2007) is constructed as $T_A = \max_{0<c\leq 1} S(c)$ in procedure A and as $T_B = \max\{S(0)+R, \max_{0<c\leq 1} S(c)\}$

in procedure B, the asymptotic distributions of T_A and T_B under H_0 cannot be obtained easily. A bootstrap procedure can be used to determine the p-value in this case. Specifically, we randomly permute the $Z_1, \ldots, Z_n$ values to the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ to form a new data set $\{(X_1, Y_1, Z^*_{b,1}), \ldots, (X_n, Y_n, Z^*_{b,n})\}$, and then construct the same test statistic T^*_b. Repeating the procedure a large number of times, say B times, the p-value can be calculated as $\sum_{b=1}^{B} I(T^*_b > T)/B$. Because the procedure to generate the data under the null hypothesis is permutation-based, the resulting bootstrap test is also called the permutation test.

A.2. Variance estimation of the cutpoint

Maximizing $\hat{U}(c)^2/\hat{v}(c)$ to estimate c is equivalent to solving

$$2\hat{U}'(c)\hat{v}(c) - \hat{U}(c)\hat{v}'(c) = 0,$$

where $\hat{U}'(c)$ and $\hat{v}'(c)$ correspond to the derivatives of $\hat{U}(c)$ and $\hat{v}(c)$ with respect to c. Denote the maximizer of $\hat{U}(c)^2/\hat{v}(c)$ as $\hat{c}$, and let $\phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_a, 0, c) = \partial\phi_3(\mathcal{O}_i; \boldsymbol{\beta}_a, 0, c)/\partial c$. We assume that c_0 satisfies

$$E\left\{\phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0)v'(c_0) - 2\phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0)v(c_0)\right\} = 0.$$

Then we have

$$\begin{aligned}
0 &= \hat{U}(\hat{c})\hat{v}'(\hat{c}) - 2\hat{U}'(\hat{c})\hat{v}(\hat{c}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_3(\mathcal{O}_i; \widetilde{\boldsymbol{\beta}}_a, 0, \hat{c})\hat{v}'(\hat{c}) - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\phi'_{3c}(\mathcal{O}_i; \widetilde{\boldsymbol{\beta}}_a, 0, \hat{c})\hat{v}(\hat{c}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0)v'(c_0) - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0)v(c_0) \\
&\quad + \frac{\hat{v}'(\hat{c})}{\sqrt{n}}\sum_{i=1}^{n}\left\{\phi_3(\mathcal{O}_i; \widetilde{\boldsymbol{\beta}}_a, 0, \hat{c}) - \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0)\right\} \\
&\quad + \frac{\hat{v}'(\hat{c}) - v'(c_0)}{\sqrt{n}}\sum_{i=1}^{n}\phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) - \frac{2\{\hat{v}(\hat{c}) - v(c_0)\}}{\sqrt{n}}\sum_{i=1}^{n}\phi'_{3c}(\mathcal{O}_i; \widetilde{\boldsymbol{\beta}}_a, 0, \hat{c}) \\
&\quad - \frac{2v(c_0)}{\sqrt{n}}\sum_{i=1}^{n}\left\{\phi'_{3c}(\mathcal{O}_i; \widetilde{\boldsymbol{\beta}}_a, 0, \hat{c}) - \phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0)\right\}.
\end{aligned}$$

We now consider each term of the above equation separately. First, take

$$\phi_b(\mathcal{O}_i, \boldsymbol{\beta}_a, 0, c) = \phi_3(\mathcal{O}_i, \boldsymbol{\beta}_a, 0, c) - \mathbf{A}_2(c)\mathbf{A}_1^{-1}\boldsymbol{\phi}_a(\mathcal{O}_i, \boldsymbol{\beta}_a, 0),$$

and $\phi'_{bc}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c) = \partial\phi_b(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c)/\partial c$. The third term is then

$$\begin{aligned}
&\frac{\hat{v}'(\hat{c})}{\sqrt{n}} \sum_{i=1}^{n} \Big\{ \phi_3(\mathcal{O}_i; \widetilde{\boldsymbol{\beta}}_a, 0, \hat{c}) - \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \Big\} \\
&\quad = \hat{v}'(\hat{c}) \left\{ \mathbf{A}_2(c_0) + o_p(1) \right\} \sqrt{n}(\widetilde{\boldsymbol{\beta}}_a - \boldsymbol{\beta}_{a0}) \\
&\qquad + \hat{v}'(\hat{c}) \left[E \left\{ \phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \right\} + o_p(1) \right] \sqrt{n}(\hat{c} - c_0) \\
&\quad = v'(c_0) \mathbf{A}_2(c_0) \sqrt{n}(\widetilde{\boldsymbol{\beta}}_a - \boldsymbol{\beta}_{a0}) + v'(c_0) E \left\{ \phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \right\} \sqrt{n}(\hat{c} - c_0) + o_p(1) \\
&\quad = -v'(c_0) \mathbf{A}_2(c_0) \mathbf{A}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\phi}_a(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0) \\
&\qquad + v'(c_0) E \left\{ \phi'_{3c}(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \right\} \sqrt{n}(\hat{c} - c_0) + o_p(1).
\end{aligned}$$

The fourth term is

$$\begin{aligned}
&\frac{\hat{v}'(\hat{c}) - v'(c_0)}{\sqrt{n}} \sum_{i=1}^{n} \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \\
&= [E \{ \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \} + o_p(1)] \sqrt{n} \left\{ \hat{v}'(\hat{c}) - \hat{v}'(c_0) + \hat{v}'(c_0) - v'(c_0) \right\} \\
&= [E\{ \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \} + o_p(1)][\{ v''(c_0) + o_p(1) \} \sqrt{n}(\hat{c} - c_0) + \sqrt{n}\{ \hat{v}'(c_0) - v'(c_0) \}] \\
&= E \{ \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \} v''(c_0) \sqrt{n}(\hat{c} - c_0) + E \{ \phi_3(\mathcal{O}_i; \boldsymbol{\beta}_{a0}, 0, c_0) \} \\
&\quad \times \sqrt{n} \left\{ \hat{v}'(c_0) - v'(c_0) \right\} + o_p(1).
\end{aligned}$$

Since

$$\begin{aligned}
&\sqrt{n} \left\{ \hat{v}'(c_0) - v'(c_0) \right\} \\
&= \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \Big[\phi_b(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \phi'_{bc}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \\
&\quad - E \Big\{ \phi_b(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \phi'_{bc}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \Big\} \Big] \\
&\quad - 2\sqrt{n} \Big\{ \frac{1}{n} \sum_{i=1}^{n} \phi_3(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \Big\} \Big\{ \frac{1}{n} \sum_{i=1}^{n} \phi'_{3c}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \Big\} \\
&\quad + 2\sqrt{n} E \{ \phi_3(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \} E \left\{ \phi'_{3c}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \right\} \\
&= \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \Big[\phi_b(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \phi'_{bc}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \\
&\quad - E \Big\{ \phi_b(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \phi'_{bc}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \Big\} \Big] \\
&\quad - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} [\phi_3(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) - E \{ \phi_3(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \}] \Big\{ \frac{1}{n} \sum_{i=1}^{n} \phi'_{3c}(\mathcal{O}_i, \boldsymbol{\beta}_{a0}, 0, c_0) \Big\}
\end{aligned}$$

$$
\begin{aligned}
&-2\sqrt{n}E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\frac{1}{n}\sum_{i=1}^{n}\left[\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\Big[\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\\
&\quad-E\left\{\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\Big]\\
&\quad-\frac{2E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&\quad-\frac{2E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&\quad+o_p(1),
\end{aligned}
$$

the fourth term can be written as

$$
\begin{aligned}
&\frac{\hat{v}'(\hat{c})-v'(c_0)}{\sqrt{n}}\sum_{i=1}^{n}\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\\
&=E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}v''(c_0)\sqrt{n}(\hat{c}-c_0)+\frac{2E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\\
&\quad\times\sum_{i=1}^{n}\Big[\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\\
&\quad-E\left\{\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\Big]\\
&\quad-\frac{2E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\\
&\quad\times\sum_{i=1}^{n}\left[\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&\quad-\frac{2\left[E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]^2}{\sqrt{n}}\sum_{i=1}^{n}\Big[\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\Big]\\
&\quad+o_p(1).
\end{aligned}
$$

The fifth term is

$$
\begin{aligned}
&\frac{2\{\hat{v}(\hat{c})-v(c_0)\}}{\sqrt{n}}\sum_{i=1}^{n}\phi'_{3c}(\mathcal{O}_i;\widetilde{\boldsymbol{\beta}}_a,0,\hat{c})\\
&\quad=2\sqrt{n}\{\hat{v}(\hat{c})-\hat{v}(c_0)+\hat{v}(c_0)-v(c_0)\}\left[E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}+o_p(1)\right]\\
&\quad=[2\{v'(c_0)+o_p(1)\}\sqrt{n}(\hat{c}-c_0)+2\sqrt{n}\{\hat{v}(c_0)-v(c_0)\}][E\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\}\\
&\qquad+o_p(1)]
\end{aligned}
$$

$$
\begin{aligned}
&= 2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}v'(c_0)\sqrt{n}(\hat{c}-c_0)\\
&\quad +2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\sqrt{n}\{\hat{v}(c_0)-v(c_0)\}+o_p(1).
\end{aligned}
$$

Since the variance can be expanded as

$$
\begin{aligned}
&\sqrt{n}\,\{\hat{v}(c_0)-v(c_0)\}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&\quad -\sqrt{n}\Big\{\frac{1}{n}\sum_{i=1}^{n}\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\Big\}^2+\sqrt{n}\left[E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]^2\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&\quad -\frac{2E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\}\right]+o_p(1),
\end{aligned}
$$

the fifth term can be further written as

$$
\begin{aligned}
&\frac{2\{\hat{v}(\hat{c})-v(c_0)\}}{\sqrt{n}}\sum_{i=1}^{n}\phi'_{3c}(\mathcal{O}_i;\widetilde{\boldsymbol{\beta}}_a,0,\hat{c})\\
&= 2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}v'(c_0)\sqrt{n}(\hat{c}-c_0)\\
&\quad +2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&\quad -\frac{4E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\\
&\quad \times\sum_{i=1}^{n}\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}+o_p(1).
\end{aligned}
$$

Finally, the last term is

$$
\begin{aligned}
&\frac{2v(c_0)}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\phi'_{3c}(\mathcal{O}_i;\widetilde{\boldsymbol{\beta}}_a,0,\hat{c})-\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\Big\}\\
&\quad = 2v(c_0)E\left\{\frac{\partial\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)}{\partial\boldsymbol{\beta}_a^{\mathrm{T}}}\right\}\sqrt{n}(\widetilde{\boldsymbol{\beta}}_a-\boldsymbol{\beta}_{a0})\\
&\qquad +2v(c_0)E\left\{\frac{\partial\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)}{\partial c}\right\}\sqrt{n}(\hat{c}-c_0)+o_p(1)\\
&\quad = -2v(c_0)\mathbf{A}_2'(c_0)\mathbf{A}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\phi}_a(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0)
\end{aligned}
$$

$$+2v(c_0)E\left\{\phi''_{3cc}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\sqrt{n}(\hat{c}-c_0)+o_p(1),$$

where $\phi''_{3cc}(\mathcal{O}_i;\boldsymbol{\beta}_a,0,c)=\partial\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_a,0,c)/\partial c$.

Combining these terms, we obtain

$$\begin{aligned}
0=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)v'(c_0)-\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)v(c_0)\\
&-v'(c_0)\mathbf{A}_2(c_0)\mathbf{A}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\phi}_a(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0)\\
&+v'(c_0)E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\sqrt{n}(\hat{c}-c_0)\\
&+E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}v''(c_0)\sqrt{n}(\hat{c}-c_0)+\frac{2E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\\
&\times\sum_{i=1}^{n}[\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\\
&-E\{\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\}]\\
&-\frac{2E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\\
&\sum_{i=1}^{n}\left[\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&-\frac{2\left[E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]^2}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&-2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}v'(c_0)\sqrt{n}(\hat{c}-c_0)\\
&-2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\left\{\phi_3^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\
&+\frac{4E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}E\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}}{\sqrt{n}}\\
&\times\sum_{i=1}^{n}\left\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)-E\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\\
&+2v(c_0)\mathbf{A}'_2(c_0)\mathbf{A}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\phi}_a(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0)\\
&-2v(c_0)E\left\{\phi''_{3cc}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\sqrt{n}(\hat{c}-c_0)+o_p(1).
\end{aligned}$$

Let

$$\begin{aligned}
\xi=&E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}v''(c_0)-v'(c_0)E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\\
&-2v(c_0)E\left\{\phi''_{3cc}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\},
\end{aligned}\tag{A.1}$$

$$\begin{aligned}\zeta_i = {} & \phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)v'(c_0) - 2\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)v(c_0)\\ & -v'(c_0)\mathbf{A}_2(c_0)\mathbf{A}_1^{-1}\boldsymbol{\phi}_a(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0)\\ & +2E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\left[\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right.\\ & \left.-E\{\phi_b(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\phi'_{bc}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\}\right]\\ & -2E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\left[\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right.\\ & \left.-E\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\}\right]\\ & -2\left[E\left\{\phi_3(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]^2\left[\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0) - E\left\{\phi'_{3c}(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\ & -2E\left\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\right\}\left[\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0) - E\left\{\phi_b^2(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\right\}\right]\\ & +4E\{\phi'_{3c}(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0,c_0)\}E\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\}\{\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\\ & -E\phi_3(\mathcal{O}_i,\boldsymbol{\beta}_{a0},0,c_0)\} + 2v(c_0)\mathbf{A}'_2(c_0)\mathbf{A}_1^{-1}\boldsymbol{\phi}_a(\mathcal{O}_i;\boldsymbol{\beta}_{a0},0). \qquad \text{(A.2)}\end{aligned}$$

This yields

$$-\xi\sqrt{n}(\widehat{c}-c_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i + o_p(1).$$

Hence, the variance of $\hat{c}$ can be estimated by the sample version $\hat{\sigma}_c^2 = \hat{\xi}^{-2}\widehat{\mathrm{Var}}(\zeta_i)/n$ evaluated at $(\widetilde{\boldsymbol{\beta}}_a, 0, \hat{c})$.

References

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd edition. CRC Press, London.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictors. *J. Amer. Statist. Assoc.* **85**, 652-663.

Cook, J. and Stefanski, L. A. (1995). A simulation extrapolation method for parametric measurement error models. *J. Amer. Statist. Assoc.* **89**, 1314-1328.

Freidlin, B. and Simon, R. (2005). Adaptive signature design: An adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clinical Cancer Research* **11**, 7872-7878.

Freidlin, B., Jiang, W. and Simon, R. (2010). The cross-validated adaptive signature design. *Clinical Cancer Research* **16**, 691-698.

Gleser, L. J. (1990). Improvements of the naive approach to estimation in non-linear errors-in-variables regression models. *Statistical Analysis of Measurement Error Models and Applications* (Edited by P. J. Brown and W. A. Fuller). American Mathematical Society, Providence.

Jiang, W., Freidlin, B. and Simon, R. (2007). Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. National Cancer Institute* **99**, 1036-1043.

Korn, E. L., Arbuck, S. G., Pluda, J. M., Simon, R., Kaplan, R. S. and Christian, M. C. (2001). Clinical trial designs for cytostatic agents: are new approaches needed? *J. Clinical Oncology* **19**, 265-272.

Kosorok, M. R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.* **35**, 957-989.

Luo, X., Turnbull, B. W. and Clark, L. C. (1997). Likelihood ratio tests for a changepoint with survival data. *Biometrika* **84**, 555-565.

Ma, Y. (2011). Testing change-point in logistic models with covariate measurement error. *J. Statist. Res.* **45**, 131-138.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann. Statist.* **31**, 442-463.

Ratain, M. J. and Sargent, D. J. (2009). Optimising the design of phase II oncology trials: the importance of randomisation. *European J. Cancer* **45**, 275-280.

Sargent, D. J., Conley, B. A., Allegra, C. and Collette, L. (2005). Clinical trial designs for predictive marker validation in cancer treatment trials. *J. Clinical Oncology* **23**, 2020-2027.

Simon, R. (2009). Advances in clinical trial designs for predictive biomarker discovery and validation. *Current Breast Cancer Reports* **1**, 216-221.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703-716.

Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91**, 835-848.

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.

E-mail: ma@stat.tamu.edu

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: gyin@hku.hk

(Received February 2012; accepted September 2012)