



Evaluating biomarkers for treatment selection from reproducibility studies

XIAO SONG* and KEVIN K. DOBBIN

Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia,
Athens, GA 30602, USA
xsong@uga.edu

SUMMARY

We consider evaluating new or more accurately measured predictive biomarkers for treatment selection based on a previous clinical trial involving standard biomarkers. Instead of rerunning the clinical trial with the new biomarkers, we propose a more efficient approach which requires only either conducting a reproducibility study in which the new biomarkers and standard biomarkers are both measured on a set of patient samples, or adopting replicated measures of the error-contaminated standard biomarkers in the original study. This approach is easier to conduct and much less expensive than studies that require new samples from patients randomized to the intervention. In addition, it makes it possible to perform the estimation of the clinical performance quickly, since there will be no requirement to wait for events to occur as would be the case with prospective validation. The treatment selection is assessed via a working model, but the proposed estimator of the mean restricted lifetime is valid even if the working model is misspecified. The proposed approach is assessed through simulation studies and applied to a cancer study.

Keywords: Conditional score; Measurement error; Predictive biomarker; SIMEX; Survival.

1. INTRODUCTION

A biomarker used to predict the response to a treatment is called a predictive biomarker. For example, patients with colon cancer can be treated by surgery alone or surgery plus chemotherapy. Surgery alone is less expensive and has fewer side effects than surgery plus chemotherapy, but it may be less effective as well, at least for some patients. For an individual patient, it is desirable to identify whether or not the patient will benefit more from the extra chemotherapy based on a biomarker or a set of biomarkers. A possible useful biomarker in this context is the c-myc gene, which is over-expressed in approximately 70 percent of human colonic tumors. Based on a study conducted by the Eastern Cooperative Oncology Group (ECOG), [Augenlicht and others \(1997\)](#) suggested that the c-myc gene may be of clinically prognostic importance in patients with colon cancer. Using a subset of the cases from this clinical trial, [Li and Ryan \(2006\)](#) found that there is an interaction between the c-myc gene expression levels and the two treatments for the response of disease progression-free survival; [Song and Zhou \(2011\)](#) investigated using the observed c-myc gene expression level for treatment selection.

*To whom correspondence should be addressed.

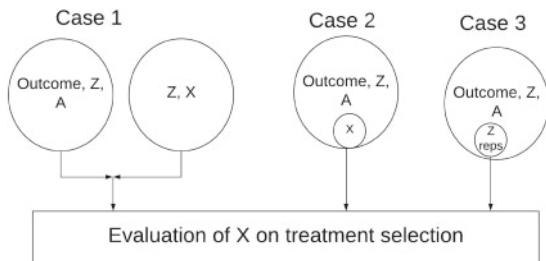


Fig. 1. Overview of the three cases, where Z denotes the standard biomarkers, X denotes the new biomarkers, and A denotes the treatment.

To evaluate the potential of biomarkers for treatment selection, approaches have been proposed that try to minimize the population event rate under (optimal) treatment selection criteria (Song and Pepe, 2004; Brinkley and others, 2010; Cai and others, 2011; Zhang and others, 2012; Janes and others, 2011, 2014) or maximize the population mean (restricted lifetime) (Song and Zhou, 2011).

However, biomarker measurements may contain measurement error, or the current biomarkers may not be very effective for treatment selection. In the ECOG study, the c-myc gene expression levels were measured with error (Li and Ryan, 2006). It is of interest to evaluate the amount of gain that could be achieved with respect to treatment selection if the biomarkers were accurately measured. In another aspect, a new technology may improve the measurement of biomarkers or a new biomarker may be identified with better predictability capacity. For example, with the development of polymerase chain reaction technique, the measurement of c-myc gene expression level may become more accurate. If we call the measurement of c-myc gene level in the original study the *standard biomarker*, and the improved measurement using advanced techniques the *new biomarker*, then we would like to assess the capacity of the new biomarker for making treatment selections. With a slight abuse of terminology, we refer both newly identified biomarkers and more accurately measured biomarkers as new biomarkers. Ideally, one would like to re-run the previous study, or perform a new study very similar to it, so that one could measure the new biomarkers using prospectively collected samples. But this is often not feasible. Such a procedure would entail large additional costs associated with obtaining samples of tumor tissues with measurement of disease progression-free survival through a multi-year randomized clinical trial. Moreover, it may not even be feasible to perform such a study. We propose a more efficient approach which requires only conducting a reproducibility study where the new biomarkers and the standard biomarkers will be measured on a set of patient samples or replicated measurements of the inaccurately measured standard biomarkers. Importantly, there is no need to re-run the clinical trial. This makes the study easier to conduct and much less expensive. In addition, it will make it possible to perform the estimation of the clinical performance of the new biomarkers quickly, since there will be no requirement to wait for events to occur. The idea of our approach is summarized in Figure 1. The outcome and the standard biomarkers are observed in the clinical trial, while the new biomarkers and the standard biomarkers are observed in the reproducibility study. We make inference on the new biomarkers versus the outcome to assess the capacity of the new biomarkers on treatment selection. There are related studies in the literature (Boosstra and others, 2013a, 2013b) aiming to predict outcome with new biomarkers observed on a subset in the original study, but their objectives are different from what is considered in this article.

In this article, we consider the setting in which the outcome is time to an event of interest (survival time), which may be subject to right censoring. To characterize the inter-relations between the biomarkers and the treatment arms, we adopt a working proportional hazards model with interactions between the treatment and the biomarkers. The relation between the standard markers and the new markers is modeled through

a classical measurement error model. Various approaches have been proposed to estimate the regression coefficients in the presence of covariate measurement error under the proportional hazards model, however, most were derived under the assumption of linear covariate effects. These include regression calibration (Prentice, 1982; Dafni and Tsiatis, 1998; Wang *and others*, 2000), SIMEX (Greene and Cai, 2004), likelihood based approaches (Wulfsohn and Tsiatis, 1997; Faucett and Thomas, 1996; Henderson *and others*, 2000; Xu and Zeger, 2001; Song *and others*, 2002b), conditional score (Tsiatis and Davidian, 2001; Song *and others*, 2002a), and correction approaches (Huang and Wang, 2000), among others. Here, we extend the conditional score approach to the proportional hazards model with interactions. In addition, we propose to use the mean restricted lifetime to evaluate the performance of the predictive biomarkers and derive the optimal treatment selection strategy under the working model. To estimate the mean restricted lifetime, we propose a SIMEX estimator and establish the asymptotic properties using the empirical process and stochastic integral techniques.

The novelty of this article includes the following aspects. First, our idea of evaluating new biomarkers without re-running the clinical trial is novel, which could greatly reduce the study time and cost. Second, the adoption of the measurement error model and techniques under this circumstance is novel. Third, we propose well-justified resampling-based inference which extends the technique of Peng and Huang (2008). To the best of our knowledge, the overlay of the resampling-based inference with the already resampling-based SIMEX approach is new.

The article is organized as follows. In Section 1, we give the model definition. We derive an empirical estimator for the optimal treatment selection and propose an approach to evaluate and compare the new biomarkers and standard biomarkers on treatment selection in Section 2. We investigate the finite sample performance of the proposed approach in Section 3, and we apply the approach to the ECOG data in Section 4. Some discussions are given in Section 5. The regularity conditions and sketched proofs are given in the [supplementary material](#) available at *Biostatistics* online.

2. MODEL DEFINITION

Let T denote the survival time, and C denote the censoring time. The observed survival data are $V = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Let Z denote a vector of K continuous standard biomarkers, and X denote a vector of K continuous new biomarkers. Remember that, with a slight abuse of terminology, we refer both newly identified biomarkers and more accurately measured biomarkers as new biomarkers. Let A denote the treatment, where $A = 0$ denotes the control or standard treatment, and $A = 1$ denotes the new treatment. Suppose a randomized clinical trial has been conducted to evaluate the standard biomarkers for treatment selection with the observed data $\{(V_i, \Delta_i, A_i, Z_i) : i = 1, \dots, n\}$. We are interested in evaluating the treatment selection capacity of the new biomarkers X .

For an individual with the new biomarkers X , intuitively, we may assign the subject to treatment $A = 1$ if

$$E(T|A = 1, X) > E(T|A = 0, X)$$

and $A = 0$ otherwise. That is, $A_{\text{opt}}(X) = I(D(X) > 0)$, where

$$D(X) = E(T|A = 1, X) - E(T|A = 0, X).$$

This is an extension of the treatment rule for binary outcomes where the probability of success is compared for the two treatments (Janes *and others*, 2014). However, when censoring exists, the mean (unrestricted) survival time may not be estimable if the largest observed survival time is censored without some tail correction on the estimated survival function (Klein and Moeschberger, 2003). Alternatively, we consider

the mean restricted survival time (lifetime) up to a given time L . The technique of restricting survival time has been used previously in estimating the mean lifetime and quality-adjusted lifetime (Zhao and Tsitais, 1997; Chen and Tsitais, 2001). Specifically, Let $T^* = \min(T, L)$ be the restricted survival time, and

$$D^*(X) = E(T^*|A = 1, X) - E(T^*|A = 0, X),$$

then the optimal treatment $A_{\text{opt}}^*(X) = I(D^*(X) > 0)$; that is, if $D^*(X) > 0$, select $A = 1$; otherwise, select $A = 0$. The capacity of treatment selection based on $A_{\text{opt}}^*(X)$ can be evaluated by the population mean restricted lifetime under the optimal treatment selection, that is,

$$\Theta_X(A_{\text{opt}}^*) = E[T^*|A = A_{\text{opt}}^*(X)]. \quad (2.1)$$

Without loss of generality, we assume an additive measurement error model

$$Z = X + e, \quad (2.2)$$

where $e \sim N(0, \Sigma)$, and e is independent of (X, A, T, C) . This is a natural model when the new biomarkers are obtained by improving the accuracy of measurement, but may represent a more general relationship between standard biomarkers and “true” new biomarkers. For example, if we have standard biomarkers Z^* and new biomarkers X^* and there exist functions g_1 and g_2 , which could be vector valued, such that

$$g_1(Z^*) = g_2(X^*) + e, \quad (2.3)$$

it reduces to model (2.2) with $Z = g_1(Z^*)$ and $X = g_2(X^*)$. For simplicity of presentation, we assume that g_1 and g_2 are known. When g_1 and g_2 are unknown, the relationship between the standard biomarkers and the new biomarkers can be estimated as discussed in Section 6.

To ensure the identifiability of model (2.2), we need to have either validation data or replicated data on Z . We consider three cases. In case 1, a validation data set is available from an external reproducibility study. The observations in the reproducibility study are $\{(X_i^\#, Z_i^\#) : i = 1, \dots, m\}$. In case 2, an internal validation data set of a size m is available in the original data set. Although we may directly evaluate the new marker using the validation set in this case, it would be more efficient to use the whole data set. In case 3, replicated error-contaminated observations are available on some subjects in the original study. Case 3 is only feasible when the new biomarkers are obtained by improving the accuracy of measurement while cases 1 and 2 also cover the situation when the new biomarkers are truly different variables. To unify the notations in the three cases, the observed data in the original study is denoted by $\{(V_i, \Delta_i, A_i, Z_{ij}, R_i X_i) : i = 1, \dots, n, j = 1, \dots, J_i\}$, where J_i denotes the number of replicates for subject i , which always equals one in cases 1 and 2; $R_i = 1$ for a subset of $i \in \Upsilon \subset \{1, \dots, n\}$. The set Υ contains m elements for case 2, while it is empty for cases 1 and 3.

3. ESTIMATION

3.1. Estimation of the optimal treatment

To estimate Θ_X , we adopt a working model, which assumes the survival time depends on the new biomarkers X and the treatment A through a proportional hazards model

$$\lambda(t|X, A) = \lambda_0(t) \exp(\beta_1^T X + \beta_2 A + \beta_3^T X A), \quad (3.1)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and $\beta = (\beta_1^T, \beta_2, \beta_3^T)^T$ are the regression parameters. Extension to more flexible survival models is discussed in Section 6.

Under model (3.1), we have

$$D^*(X) = \int_0^L S(t|\beta, \Lambda_0, 1, X) dt - \int_0^L S(t|\beta, \Lambda_0, 0, X) dt,$$

where $S(t|\beta, \Lambda_0, A, X) = \exp\{-\Lambda_0(t) \exp(\beta_1^T X + \beta_2 A + \beta_3^T A X)\}$ with $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ being the baseline cumulative hazard function. Thus $D^*(X) > 0$ if and only if $\beta_2 + \beta_3^T X < 0$, which implies $A_{\text{opt}}^*(X) = A(\beta, X) \triangleq I(\beta_2 + \beta_3^T X < 0)$. In fact, it can be easily seen that $A_{\text{opt}}(X) = A_{\text{opt}}^*(X)$ under model (3.1); that is, the optimal treatment based on the mean restricted survival time equals the optimal treatment based on the mean unrestricted survival time.

If X were observed, an ideal estimator $\hat{\beta}^I$ of β could be obtained by the standard partial likelihood approach, and the ideal estimator of $A_{\text{opt}}(X)$ is $\hat{A}_{\text{opt}}^I(X) = A(\hat{\beta}^I, X)$. Here and henceforth, we use the superscript “ I ” to denote the ideal approach. Since X is not observed in the original study or only observed in a subset Υ , we may estimate β through measurement error approaches.

We adopt the conditional score approach (Song and others, 2002a) for cases 1 and 3 as it is simple to compute. The conditional score estimator was originally derived for the proportional hazards model without interactions. Here we extend it to model (3.1). Specifically, assume Σ is known for now. Following similar arguments as those in Song and others (2002b), we may obtain the “complete sufficient statistic” for X_i , $H_i = \bar{Z}_i + J_i^{-1} \Sigma (\beta_1^T + \beta_3^T A_i) dN_i(u)$. Here $\bar{Z}_i = J_i^{-1} \sum_{j=1}^{J_i} Z_{ij}$. The conditional score estimating equation can be written as

$$U_n(\beta, \Sigma, H, A) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ (H_i^T, A_i, H_i^T A_i)^T - \frac{\hat{G}_1(t, \beta, \Sigma, H, A)}{\hat{G}_0(t, \beta, \Sigma, H, A)} \right\} dN_i(t) = 0, \quad (3.2)$$

where $\tau \geq L$ is a fixed time, $\hat{G}_{r, \beta, H, A}(t, \beta, \Sigma, H, A) = n^{-1} \sum_{i=1}^n G_{ri}(t, \beta, \Sigma, H, A)$,

$$\begin{aligned} G_{ri}(t, \beta, \Sigma, H, A) &= Y_i(t) (H_i^T, A_i, H_i^T A_i)^{T \otimes r} \\ &\times \exp \left\{ \beta_1^T H_i + \beta_2 A_i + \beta_3^T H_i^T A_i - (1 - \rho_i) (\beta_1^T + \beta_3^T A_i) (\Sigma / J_i) (\beta_1 + \beta_3 A_i) / 2 \right\}, \end{aligned}$$

with $a^{\otimes r} = 1, a, aa^T$ for a vector a and $r = 0, 1, 2$, respectively; $N_i(t) = I(V_i \leq t, \Delta_i = 1)$ is the counting process for the events, and $Y_i(t) = I(V_i \geq t)$ is the at-risk process. The error variance Σ may be estimated by the method of moments estimator $\hat{\Sigma}$ from the validation data or the replicated data (Song and others, 2002a).

In case 2, X is observed in a subset Υ . The partial likelihood estimator $\hat{\beta}^\Upsilon$ of β may be obtained using observations in Υ only. But this approach is not efficient as the information not in Υ is not used. To improve the efficiency, following Wang and Song (2016), an improved estimator can be obtained. Specifically, it is the best linear combination of $\hat{\beta}^\Upsilon$ and $\hat{\beta}^{\Upsilon, N}$, which equals $\hat{\beta}^\Upsilon - \hat{\Omega}(\hat{\beta}^{\Upsilon, N} - \hat{\beta}^N)$, where $\hat{\beta}^{\Upsilon, N}$ and $\hat{\beta}^N$ are the naive estimates of β obtained by substituting W for X using the observations in Υ and the whole data set, respectively, and $\hat{\Omega}$ is given in Appendix A in the [supplementary material](#) available on *Biostatistics* online. For simplicity, both this estimator and the conditional score estimator are referred to as error-corrected estimators henceforth.

Denote the error-corrected estimators of β by $\hat{\beta}^C$. The optimal treatment can be estimated by $\hat{A}_{\text{opt}}^C(X) = A(\hat{\beta}^C, X)$. Here and henceforth, we use the superscript “ C ” to denote the error-corrected approach. For now, assume model (3.1) holds.

PROPOSITION 1 Under the conditions C1–C4 given in Appendix B available on *Biostatistics* online, almost surely, $\hat{\beta}^C$ exists and converges to β . In addition, $n^{1/2}(\hat{\beta}^C - \beta)$ converges to a mean zero normal distribution.

If model (3.1) is not the true model, it is used as a working model to obtain the treatment selection criterion. Fine (2002) showed that even if model (3.1) does not hold, the partial likelihood estimator $\hat{\beta}^I$ still converges to some constant $\beta^* = (\beta_1^{*T}, \beta_2^*, \beta_3^{*T})$. It can be shown that the error-corrected estimators are consistent estimators of β^* . Given the value of X , $A(\hat{\beta}^I, X)$ will converge to a valid treatment selection criterion $A(\beta^*, X)$, which is equal to $A_{\text{opt}}(X)$ when model (3.1) is correctly specified. When model (3.1) is misspecified, $A(\beta^*, X)$ is the optimal treatment under (3.1), but may not equal to the optimal treatment $A_{\text{opt}}(X)$. We consider evaluating $A(\beta^*, X)$ and propose an empirical estimator of $\Theta_X(A(\beta^*, X))$.

3.2. Estimation of Θ_X

With some algebra, it can be shown that

$$\begin{aligned}\Theta_X(A(\beta^*, X)) &= E[E\{T^*|A = A(\beta^*, X)\}] \\ &= E(A(\beta^*, X)) \int_0^L \exp\{-\Lambda(t|A = 1, A(\beta^*, X) = 1)\} dt \\ &\quad + \{1 - E(A(\beta^*, X))\} \int_0^L \exp\{-\Lambda(t|A = 0, A(\beta^*, X) = 0)\} dt,\end{aligned}\tag{3.3}$$

where for $r = 0, 1$, $\Lambda(t|A = r, A(\beta, X) = r)$ is the cumulative hazard function conditional on $A = r$ and $A(\beta, X) = r$. If X were observed in the clinical trial, $E(A(\beta^*, X))$ could be estimated by the empirical estimator $\hat{E}(A(\hat{\beta}^I, X)) = n^{-1} \sum_{i=1}^n I(\hat{\beta}_2^I + \hat{\beta}_3^I X_i < 0)$, and $\Lambda(t|A = r, A(\beta^*, X) = r)$ could be estimated by Nelson–Aalen type estimators $\hat{\Lambda}(t|A = r, A(\hat{\beta}^I, X) = r)$, where

$$\hat{\Lambda}(t|A = r, A(\hat{\beta}^I, X) = r) = \sum_{i=1}^n \int_0^t \frac{dI(V_i \leq u, \Delta_i = 1, A_i = r, A(\beta, X_i) = r)}{\sum_{j=1}^n I(V_j \geq u, A_j = r, A(\beta, X_j) = r)}.$$

Then an estimator of $\Theta_X(A(\beta^*, X))$ could be

$$\begin{aligned}\hat{\Theta}_X^I &= \hat{E}(A(\hat{\beta}^I, X)) \int_0^L \exp\{-\hat{\Lambda}(t|A = 1, A(\hat{\beta}^I, X) = 1)\} dt \\ &\quad + \{1 - \hat{E}(A(\hat{\beta}^I, X))\} \int_0^L \exp\{-\hat{\Lambda}(t|A = 0, A(\hat{\beta}^I, X) = 0)\} dt.\end{aligned}\tag{3.4}$$

However, X is not observed in the original study or only observed in a subset Υ . To deal with the measurement error, we may apply the SIMEX approach (Carroll and others, 2006). Assuming Σ is known for now, for an increasing sequence of value of ζ starting from 0, for example, $\zeta = 0, 0.25, 0.5, \dots, 2$, and $b = 1, 2, \dots, B$, let $W_{\zeta i}^b = R_i X_i + (1 - R_i) \bar{Z}_i + \sqrt{\zeta} J_i^{-1/2} \Sigma^{1/2} \varepsilon_{\zeta i}^b$, where $\Sigma^{1/2}(\Sigma^{1/2})^T = \Sigma$, and $\varepsilon_{\zeta i}^b \sim N(0, I)$. Calculate the naive estimator $\hat{\Theta}_{\zeta}^b$ by replacing X_i by $W_{\zeta i}^b$ and $\hat{\beta}^I$ by $\hat{\beta}^C$ in (3.4). Let

$$\hat{\Theta}_{\zeta} = B^{-1} \sum_{b=1}^B \hat{\Theta}_{\zeta}^b = B^{-1} \sum_{b=1}^B \hat{E}_W^b(T^*|A = A(\hat{\beta}^C, W_{\zeta i}^b); W_{\zeta i}^b).$$

Extrapolate $(\zeta, \hat{\Theta}_{\zeta})$ to $\zeta = -1$ to get the SIMEX estimator $\hat{\Theta}_X^C$. A regression model is usually adopted for the extrapolation, such as the quadratic and nonlinear (rational linear) extrapolation (Carroll and others,

2006). Specifically, suppose that $\hat{\Theta}_\zeta = h(\theta, \zeta) + \epsilon_\zeta$, where ϵ_ζ is an error term with mean 0. Let $\hat{\theta}$ be the least square estimator of θ . Then the SIMEX estimator can be written as

$$\hat{\Theta}_X^C = h(\hat{\theta}, -1).$$

When Σ is not known, it can be replaced by the estimator $\hat{\Sigma}$ obtained from the validation data or the replicated data. Specifically, $\hat{\Sigma} = m^{-1} \sum_{j=1}^m (Z_j^\# - X_j^\#)^{\otimes 2}$ in case 1, $\hat{\Sigma} = \{\sum_{i=1}^n (J_i - 1)\}^{-1} \sum_{i=1}^n I(J_i > 1) \sum_{j=1}^{J_i} (Z_{ij} - \bar{Z}_i)^{\otimes 2}$ in case 3 and $\hat{\Sigma} = m^{-1} \sum_i^n R_i (Z_{i1} - X_i)^{\otimes 2}$ in case 2.

PROPOSITION 2 Under the regularity conditions C1–C3 given in Appendix B available on *Biostatistics* online, $\hat{\Theta}_X^I$ is a consistent estimator of $\Theta_X(A(\beta^*, X))$ and $n^{1/2}(\hat{\Theta}_X^I - \Theta_X(A(\beta^*, X)))$ converges to a mean zero normal distribution. Further, with the additional conditions C4–C6, $\hat{\Theta}_X^C$ is a consistent estimator of $\Theta_X(A(\beta^*, X))$ and $n^{1/2}(\hat{\Theta}_X^C - \Theta_X(A(\beta^*, X)))$ converges to a mean zero normal distribution.

Proposition 2 indicates that even if the model (3.1) is misspecified, the empirical estimator $\hat{\Theta}_X^C$ is still a valid estimator of the mean restricted lifetime under the treatment selection criterion constructed based on the working model.

Model-based estimation. If the working model is the true model, a model-based estimator can be obtained. Noting that

$$E(T^*|A) = E [E(T^*|A, X)|A] = E \left[\int_0^L S(t|\beta, \Lambda_0, A, X) dt \middle| A \right]. \quad (3.5)$$

If X_i were observed, an ideal estimator of Θ_X could be

$$\hat{\Theta}_{m,X}^* = \frac{1}{n} \sum_{i=1}^n \int_0^L S(t|, \hat{\beta}^I, \hat{\Lambda}_0^I, \hat{A}_{opt,i}^I, X_i) dt, \quad (3.6)$$

where $\hat{\Lambda}_0^I(t)$ is the Breslow estimator of $\Lambda_0(t)$. This cannot be applied directly since X is not observed in the clinical trial. By analogy to the empirical estimator, the SIMEX approach can be adopted to estimate Θ_X when X is not observed.

If model (3.1) is misspecified, it can be shown that the model-based estimator $\hat{\Theta}_{m,X}^*$ actually estimate $\Theta_{mis} = E[E\{\int_0^L S^*(t|\beta, \Lambda_0^*, A, X) dt|A = A(\beta^*, X)\}]$ with $S^*(t|\beta, \Lambda_0^*, A, X) = \exp\{-\Lambda_0^*(t) \exp(\beta_1^* X + \beta_2^* A + \beta_3^* XA)\}$ and $\Lambda_0^*(t) = \int_0^L dE\{N(t)\}/E(Y(t) \exp\{\beta_1^* X + \beta_2^* A + \beta_3^* XA\})$, which is different from the optimal mean restricted lifetime $\Theta_X(A(\beta^*, X))$ under the working model. For example, under the scenario considered in our simulation studies in Section 4, if the true model is $\lambda(t|X, A) = \lambda_0(t) \exp(-2.5X - 0.5X^2 + 0.2A + 5.0XA + X^2A)$, the optimal mean restricted lifetime is 6.82, the optimal mean restricted lifetime is 6.79 under working model (3.1), while $\Theta_{mis} = 6.60$. Therefore the model-based estimator may not work well for estimation of the mean restricted lifetime under the working model in this case, while the empirical estimator is still a valid estimator.

We will focus on the empirical estimation for its robustness. This approach may be applied to the standard biomarkers as described in Section 3.4, which will facilitate the comparison of the standard biomarkers and the new biomarkers.

3.3. Marker-independent treatment selection

Marker-independent treatment selection would assign all subjects to one treatment $A = a$ ($a = 0$ or 1) if it has been shown significantly better than the other treatment without taking into consideration of

their marker values, for example, through the log-rank test. The corresponding mean restricted lifetime is $E(T^*|A = a)$, which can be estimated by $\int_0^L \exp\{-\widehat{\Lambda}(t|A = a)\}dt$, where

$$\widehat{\Lambda}(t|A = a) = \sum_{i=1}^n \int \frac{dI(V_i \leq t, \Delta_i = 1, A_i = a)}{\sum I(V_i \geq t, A_i = a)}$$

is the Nelson–Aalen estimator of the cumulative hazard function $\Lambda(t|A = a)$ given $A = a$.

3.4. Compare the standard and new biomarkers

To compare the new biomarkers to the standard biomarkers, we need to estimate the capacity of the standard biomarkers Z on treatment selection. This can be evaluated by a working model that replaces X by Z in (3.1), that is,

$$\lambda(t|Z, A) = \lambda_0^\#(t) \exp(\beta_1^T Z + \beta_2 A + \beta_3^T Z A). \quad (3.7)$$

This approach is called the naive approach under the literature of measurement error models. Let $\hat{\beta}^N$ be the partial likelihood estimator based on the working model (3.7), where the superscript “ N ” is used to denote naive estimators. It can be shown that $\hat{\beta}^N$ converges to some constant $\beta^\#$ (Fine, 2002). Given Z , the optimal treatment selection criterion under the working model is $\Theta_Z = A(\beta^\#, Z) = I(\beta_2^\# + \beta_3^{\#T} Z < 0)$ and can be estimated by $A(\hat{\beta}^N, Z)$. The mean restricted lifetime $E(T^*|A = A(\beta^\#, Z))$ can be estimated empirically by $\hat{\Theta}_Z^N = \hat{E}_{e,Z}(T^*|A = A(\hat{\beta}^N, Z); Z)$, which is obtained by substituting Z for X and $\hat{\beta}^N$ for $\hat{\beta}^I$ in (3.4).

Another way to evaluate a treatment selection rule A is to evaluate the probabilities of subjects misassigned to the non-optimal treatment. If the optimal treatment is 1, the misassignment probability is

$$P_0(A) = P(A = 0 \text{ and } A_{\text{opt}} = 1), \quad (3.8)$$

and if the optimal treatment is 0, the misassignment probability is

$$P_1(A) = P(A = 1 \text{ and } A_{\text{opt}} = 0). \quad (3.9)$$

The overall misclassification probability $P(A) = P_0(A) + P_1(A) = P(A \neq A_{\text{opt}})$. Thus, we may compare the treatment selection rules $A(\beta^*, X)$ and $A(\beta^\#, Z)$ based on (3.8) and (3.9).

In the simulation studies, if β is estimated by $\hat{\beta}$, the misassignment probabilities $P_0(A(X))$ and $P_1(A(X))$ of using biomarker X can be estimated by $\hat{P}_0(A(\hat{\beta}, X)) = n^{-1} \sum_{i=1}^n I(A(\hat{\beta}, X) = 0 \text{ and } A_{\text{opt}} = 1)$ and $\hat{P}_1(A(\hat{\beta}, X)) = n^{-1} \sum_{i=1}^n I(A(\hat{\beta}, X) = 1 \text{ and } A_{\text{opt}} = 0)$, and the estimated misassignment probabilities of using Z are obtained by substituting Z for X . Note that $\hat{P}_0(A, \hat{\beta}, X)$ and $\hat{P}_1(A, \hat{\beta}, X)$ converge to zero for a consistent estimator $\hat{\beta}$ of β .

In practice, X is not observed in the original study or only observed in a subset Υ . If (3.1) is correctly specified, the overall misassignment probability of using Z equals

$$P(A(Z)) = 2[P(A(\beta^\#, Z) = 1) - P(A(\beta, X) = 1)],$$

where $P(A(\beta^\#, Z) = 1)$ can be estimated by $\tilde{P}(A(\hat{\beta}^N, Z) = 1) = n^{-1} \sum_{i=1}^n I(A(\hat{\beta}^N, Z) = 1)$, and $P(A(\beta, X) = 1)$ can be estimated by $\tilde{P}(A(\hat{\beta}^C, X) = 1) = n^{-1} \sum_{i=1}^n I(A(\hat{\beta}^C, X) = 1)$ if X were observed. Since X is not observed, we may apply the SIMEX approach to obtain the estimate of $P(A(\beta, X) = 1)$.

When the original data contain replicated observations, Z may be replaced by the mean \bar{Z} of the replicated observations, which has better performance than Z with reduced measurement error, and we may compare the treatment selection using \bar{Z} vs. X .

3.5. Resampling-based inference

Since the asymptotic variance for the empirical estimator depends on the unknown density and hazard functions, the estimation requires smoothing and may not work well when n is not large. We develop a resampling-based approach by analogy to that used in Peng and Huang (2008). We describe how to derive the variance estimator for $\hat{\Theta}_X^C$ in case 3; the process is similar in the other two cases. Specifically, we generate $\{\eta_i : i = 1, \dots, n\}$ from a known nonnegative distribution with mean 1. Using η_i as weights in the method of moment estimating equation, we first obtain the perturbed estimator $\hat{\Sigma}^P$ of Σ , where the superscript “ P ” stands for the perturbed estimator. Then using η_i as weights and replacing Σ by $\hat{\Sigma}^P$ in the conditional score estimating equation (3.2), we obtain the perturbed estimator $\hat{\beta}^{C,P}$ of β . Next, we obtain the perturbed estimator $\hat{\Theta}_{eX}^{C,P}$ through the perturbed SIMEX process where for each ζ and b , replace X by $W_{i\xi}^{b,P} = \bar{Z}_i^P + \sqrt{\zeta} J_i^{-1/2} \hat{\Sigma}^{P,1/2} \varepsilon_{i\xi}^b$ in $\hat{E}_X(T^*|A = \hat{A}_{\text{opt}}^I; X)$ with $\bar{Z}_i^P = n^{-1} \sum_{i=1}^n \eta_i Z_i$. By repeatedly generating $\{\eta_i : i = 1, \dots, n\}$, we obtain a large number of realization of $\hat{\Theta}_{X,r}^{C,P}$, denoted by $\{\hat{\Theta}_{X,r}^{C,P}\}_{r=1}^R$. It can be shown that conditional on the observed data, $n^{1/2} \{\hat{\Theta}_{X,r}^{C,P} - \hat{\Theta}_X^C\}$ has asymptotically the same distribution as $n^{1/2} \{\hat{\Theta}_X^C - \Theta_X\}$. Thus the variance of $\hat{\Theta}_X^C$ can be estimated by the sample variance of $\{\hat{\Theta}_{X,r}^{C,P}\}_{r=1}^R$, and the confidence interval of Θ_X can be constructed through Wald method or by the percentiles of $\hat{\Theta}_{X,r}^{C,P}$. When the error variance is estimated from the validation data in case 1, the perturbed estimator of $\hat{\Sigma}^P$ of Σ is obtained from the validation data with a separately generated set of perturbation variables $\{\eta_j^R : j = 1, \dots, m\}$.

4. SIMULATION STUDIES

We conducted simulation studies to evaluate the performance of the proposed approaches. Mimicking the case of c-myc gene in the ECOG study, we consider treatment selection using one biomarker. The new biomarker X was generated from a standard normal distribution, and the measurement error was generated from a normal distribution with mean 0 and variance $\Sigma = 0.1, 0.3, 0.5$, or 0.7 . The survival data contains observations of (V, Δ, A, Z) on $n = 500$ or 2000 subjects, where A was generated from a Bernoulli distribution with probability 0.5, mimicking treatment assignment in a randomized clinical trial. The survival time was generated based on model (3.1) with $\beta_1 = 2.5$, $\beta_2 = -0.2$, $\beta_3 = -5.0$, and $\lambda_0(t) = 0.4$. The optimal mean restricted lifetime under A_{opt} equals 6.93. The censoring time was generated from an exponential distribution with mean 20 and truncated at 12. The censoring rate was 30%. We considered three cases: (i) the error variance is estimated from an external validation data set that contains observations of (X, Z) on 500 subjects; (ii) the error variance is estimated from replicated observations of Z in the original data with $J_i = 2$ for $i = 1, \dots, n$; (iii) an internal validation data set is available with $m = 0.2n$. For each setting, 500 simulated data sets were generated.

We obtained the error-corrected estimators and naive estimators of β , A_{opt} , and $E(T^*|A = A_{\text{opt}})$ as described in Section 3. As a benchmark, we also obtained the ideal estimators of β , A_{opt} , and $E(T^*|A = A_{\text{opt}})$, assuming X is observed in the survival data. When the error is large, the root to the conditional score estimating equation $U_n(\beta) = 0$ may not exist, and there may exist some “outliers.” We calculate the bias based on the median of the estimates, and the standard error by the normalized median absolute deviation (MAD) via the resampling method. The 95% Wald confidence intervals were calculated correspondingly. For the SIMEX approaches, we adopted the rational linear extrapolation, and used the quadratic extrapolation as a backup if the rational extrapolation failed (Carroll and others, 2006). The

perturbation variable η was generated from a location-scale transformation of $\text{beta}(\sqrt{2} - 1, 1)$ and we set $R = 500$ in the resampling process and $B = 500$ in the SIMEX process.

The results of estimating β for $n = 500$ and 2000 are shown in Tables S1 and S2 in the [supplementary material](#) available at *Biostatistics* online, respectively. Compared to the ideal estimates, the naive estimates show clear bias with the coverage probabilities well below the nominal level; the coverage probabilities worsen when the sample size increases and error variance gets larger. The error-corrected estimates perform reasonably well for $n = 500$, and the performance improves when the sample size increases to $n = 2000$.

We also estimated the misassignment probabilities using the standard biomarker and the new biomarker for $A_{\text{opt}} = 0$ and 1 separately. The misassignment probabilities would be zero if the treatment assignment was based on the true model with the true regression coefficients. The estimated misassignment probabilities are shown in Table 1. The estimated misassignment probabilities of the error-corrected estimates based on the new marker are close to the ideal estimates, and they are much lower than those from the naive approach based on the standard biomarker. When $n = 2000$ and error standard variance increases from 0.1 to 0.7, the estimated total misassignment rate increases from 0.49% to 2.03% in case 1, 0.64% to 0.84% in case 2, and 0.44% to 0.93% in case 2 from the error-corrected approaches based on the new biomarker, while from 9.76% to 22.21% in case 1, 9.76% to 22.14% in case 2, and 6.96% to 17% in case 3 from the naive approach based on the standard marker. This shows the advantage of adopting the new biomarker, especially when the measurement error is large.

We further estimated the mean restricted lifetime within 10 years using the new biomarker and the standard biomarker. The results are shown in Table 2. The optimal mean restricted lifetime is 6.93 using the new marker, whereas when $\Sigma^{1/2}$ increases from 0.1 to 0.7, the mean restricted lifetime based on the standard marker decreases from 6.72 to 6.02 in cases 1 and 2 and 6.82 to 6.35 in case 3. The estimates perform reasonably well for $n = 500$ and improves when $n = 2000$.

Under this setting, the average survival time for treatment $A = 1$ is significantly longer than $A = 0$ without adjusted for the biomarker value. Therefore, marker-independent treatment selection will assign all subjects to treatment 1. The corresponding mean restricted lifetime within 10 years is 3.984, which is much shorter than marker based estimates. The corresponding misassignment rate is 48.4%, which is much higher than marker based estimates.

We also consider a simulation setting when the true model is misspecified as (3.1). The simulation setting is the same as case 1 above except that the true model is $\lambda(t|X, A) = \lambda_0(t) \exp(-2.5X - 0.5X^2 + 0.2A + 5.0XA + X^2A)$ with $\lambda_0(t) = 0.4$. In this case, the treatment selection based on the working model (3.1) is not optimal overall. Based on the working models (3.1) and (3.7), we estimated the misassignment probabilities and the mean restricted lifetime using the new biomarker and the standard biomarker, respectively. The results for $n = 2000$ are shown in Tables S3 and S4 in the [supplementary material](#) available at *Biostatistics* online. For estimation of the overall misassignment rate, the estimate is 3.53% from the ideal approach. When the error variance increases from 0.1 to 0.7, the estimates from the error-corrected approach based on the new biomarker are close to that from the ideal approach, while the estimates based on the standard biomarker increase considerably from 10.21% to 22.53% for the naive approach with the error variance increases from 0.1 to 0.7. The optimal mean restricted lifetime based on the new marker is 6.79 under model (3.1), while the mean restricted lifetime based on the standard marker decreases from 6.59 to 5.92 when Σ increases from 0.1 to 0.7. The estimates of the mean restricted lifetime perform well.

5. APPLICATION

We applied the proposed approach to a subset of the ECOG clinical trial, which was analyzed in [Li and Ryan \(2006\)](#), where c-myc expression level was measured via dot plots on 92 patients randomized to receive surgery alone or surgery plus chemotherapy, both progression-free survival and overall survival were

Table 1. Simulation results for estimation of treatment misassignment rate (%).

$\Sigma^{1/2}$	Marker	Method	$n = 500$				$n = 2000$			
			$A_{\text{opt}} = 0$		$A_{\text{opt}} = 1$		$A_{\text{opt}} = 0$		$A_{\text{opt}} = 1$	
			Est	SD	Est	SD	Est	SD	Est	SD
	New	Ideal	0.35	0.61	0.41	0.64	0.19	0.30	0.18	0.28
Case 1										
0.1	New	EC	0.49	0.79	0.56	0.89	0.26	0.38	0.23	0.36
	Standard	Naive	4.91	1.06	4.93	1.07	4.94	0.54	4.82	0.52
0.3	New	EC	0.77	1.24	0.75	1.21	0.40	0.62	0.38	0.57
	Standard	Naive	8.00	1.27	7.99	1.26	8.12	0.64	7.84	0.61
0.5	New	EC	1.02	1.62	1.07	1.81	0.69	1.17	0.54	0.87
	Standard	Naive	10.00	1.40	9.63	1.39	9.98	0.73	9.61	0.71
0.7	New	EC	1.31	2.33	1.92	3.62	1.04	2.04	0.99	1.97
	Standard	Naive	11.46	1.51	10.85	1.48	11.37	0.77	10.84	0.71
Case 2										
0.1	New	EC	0.60	0.95	0.66	0.96	0.32	0.47	0.32	0.48
	Standard	Naive	4.91	1.08	4.83	1.03	4.91	0.54	4.85	0.53
0.3	New	EC	0.71	1.07	0.75	1.08	0.38	0.55	0.38	0.57
	Standard	Naive	8.05	1.26	7.82	1.24	8.09	0.67	7.88	0.64
0.5	New	EC	0.73	1.11	0.78	1.12	0.41	0.59	0.41	0.61
	Standard	Naive	9.93	1.41	9.54	1.33	9.95	0.70	9.63	0.71
0.7	New	EC	0.74	1.13	0.80	1.14	0.42	0.61	0.42	0.63
	Standard	Naive	11.30	1.45	10.78	1.43	11.30	0.74	10.84	0.75
Case 3										
0.1	New	EC	0.47	0.73	0.43	0.71	0.21	0.33	0.23	0.35
	Standard	Naive	3.59	0.96	3.47	0.88	3.50	0.48	3.46	0.47
0.3	New	EC	0.60	0.94	0.57	0.91	0.25	0.40	0.31	0.45
	Standard	Naive	5.96	1.13	5.80	1.08	5.92	0.56	5.80	0.55
0.5	New	EC	0.75	1.27	0.71	1.09	0.31	0.49	0.39	0.56
	Standard	Naive	7.43	1.23	7.26	1.19	7.44	0.62	7.25	0.59
0.7	New	EC	0.78	1.28	0.96	1.56	0.47	0.76	0.46	0.72
	Standard	Naive	8.60	1.26	8.30	1.34	8.64	0.68	8.36	0.65

EC, error corrected; Est, estimate; SD, empirical standard deviation.

recorded. The results for overall survival are presented as follows while the results for progression-free survival are presented in the [supplementary material](#) available at *Biostatistics* online.

Based on the log-rank test, surgery plus chemotherapy is better than surgery alone for overall survival (p -value = 0.0186). Thus marker-independent treatment selection would assign all patients to surgery plus chemotherapy. Using this subset of data, [Li and Ryan \(2006\)](#) found marginally significant interaction between treatment $A = I$ (surgery plus chemotherapy) and log c-myc gene expression level X for progression-free survival (p -value = 0.073) and overall survival (p -value = 0.112) under model (3.1), and [Song and Zhou \(2011\)](#) evaluated bisecting the observed c-myc gene expression level Z for treatment selection. It is conjectured that a more accurate measurement of c-myc gene expression level might improve

Table 2. Simulation results for estimation of Θ_X and $\Theta_{\bar{Z}}$.

$\Sigma^{1/2}$	True	Method	$n = 500$				$n = 2000$			
			B	MAD	RMAD	CP	B	MAD	RMAD	CP
Estimation of Θ_X										
	6.931	Ideal	-0.034	0.245	0.247	0.950	-0.013	0.122	0.123	0.982
					Case 1					
0.1	6.931	EC	0.007	0.379	0.474	0.948	-0.007	0.168	0.197	0.978
0.3	6.931	EC	-0.041	0.498	0.637	0.942	-0.032	0.242	0.272	0.950
0.5	6.931	EC	-0.101	0.582	0.805	0.913	-0.104	0.283	0.345	0.948
0.7	6.931	EC	-0.211	0.632	0.894	0.922	-0.141	0.333	0.411	0.910
					Case 2					
0.1	6.931	EC	0.013	0.402	0.452	0.906	-0.007	0.176	0.195	0.964
0.3	6.931	EC	-0.033	0.440	0.603	0.940	-0.041	0.211	0.256	0.962
0.5	6.931	EC	-0.016	0.578	0.776	0.920	-0.064	0.286	0.324	0.936
0.7	6.931	EC	-0.059	0.700	0.966	0.890	-0.086	0.320	0.394	0.936
					Case 3					
0.1	6.931	EC	0.001	0.353	0.412	0.934	0.015	0.149	0.179	0.952
0.3	6.931	EC	-0.019	0.367	0.477	0.964	-0.012	0.186	0.211	0.958
0.5	6.931	EC	-0.026	0.425	0.549	0.970	-0.025	0.228	0.249	0.958
0.7	6.931	EC	-0.043	0.550	0.676	0.920	-0.013	0.265	0.293	0.938
Estimation of $\Theta_{\bar{Z}}$										
					Case 1					
0.1	6.720	Naive	-0.039	0.255	0.255	0.940	-0.009	0.114	0.127	0.958
0.3	6.411	Naive	-0.032	0.283	0.263	0.968	-0.040	0.119	0.132	0.956
0.5	6.186	Naive	-0.053	0.271	0.270	0.942	-0.013	0.142	0.134	0.964
0.7	6.017	Naive	-0.035	0.264	0.274	0.938	-0.011	0.135	0.136	0.974
					Case 2					
0.1	6.720	Naive	-0.036	0.262	0.255	0.940	-0.011	0.125	0.127	0.958
0.3	6.411	Naive	-0.026	0.257	0.265	0.952	-0.015	0.121	0.131	0.952
0.5	6.186	Naive	-0.022	0.255	0.271	0.942	-0.008	0.126	0.134	0.954
0.7	6.017	Naive	-0.019	0.261	0.274	0.940	-0.011	0.121	0.136	0.962
					Case 3					
0.1	6.818	Naive	-0.023	0.251	0.253	0.958	-0.001	0.121	0.126	0.940
0.3	6.629	Naive	-0.023	0.243	0.258	0.958	0.000	0.135	0.129	0.942
0.5	6.476	Naive	-0.016	0.258	0.262	0.954	0.003	0.131	0.131	0.954
0.7	6.347	Naive	-0.018	0.274	0.267	0.938	-0.005	0.136	0.132	0.948

EC, error corrected; B, empirical bias based on the median; MAD, empirical median absolute deviation; RMAD, resampling median absolute deviation; CP, empirical coverage probability of 95% confidence interval.

its capacity for treatment selection. In this subset, 26 subjects have replicated c-myc gene expression measurements with the estimated measurement error standard deviation equal to 0.20. The plot of the residuals $Z_{ij} - \bar{Z}_i$ and the corresponding Q-Q plot indicate that it is reasonable to assume the error is normal with constant variance (Figure 2).

We checked the proportional hazards assumption in model (3.7) using the method in Therneau and Grambsch (2006, Chapter 6.2). To check the proportional hazards assumption in model (3.1), as X is not

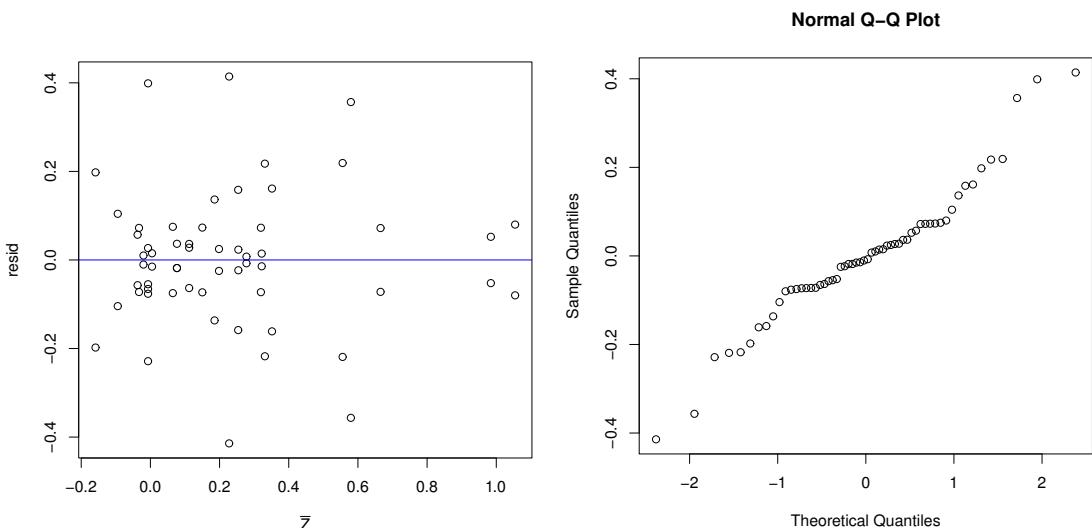


Fig. 2. Left, residual plot; right, Q–Q plot of the residuals.

Table 3. Analysis of the ECOG data for overall survival.

Estimation of Cox model coefficients			
	log(c-myc) Est (RMAD)	TRT Est (RMAD)	log(c-myc)×TRT Est (RMAD)
Model based on X	0.715 (0.548)	-0.027 (0.488)	-1.355 (0.975)
Model based on Z	0.582 (0.396)	-0.105 (0.428)	-1.126 (0.691)
Estimation of 5-year restricted lifetime			
	Est	CI	
All assigned to surgery alone	3.85	(3.41–4.29)	
All assigned to surgery+chemotherapy	4.10	(3.66–4.53)	
Treatment selection using X	4.39	(3.79–5.00)	
Treatment selection using Z	4.13	(3.71–4.55)	

lable, we adopted SIMEX approach to obtain the error-corrected p-values. There is no evidence of violation of the proportional hazards assumptions in both models.

We would like to assess the amount of gain of improving c-myc gene expression level measurement on treatment selection and to compare treatment selection based on the true value X and the error-contaminated observation. To reduce variation from SIMEX and bootstrap, we took larger values of $R = 5000$ in the resampling process and $B = 5000$ in the SIMEX process. Under the working proportional hazards model (3.1) and (3.7), the estimated coefficients and standard errors are shown in Table 3. The error-corrected estimates of log c-myc level and the interaction are larger in magnitude. Among the 92 subjects, 48 (47.8%) patients were assigned to surgery alone and 44 to surgery plus chemotherapy. An estimate of 87% patients would be assigned to surgery plus chemotherapy if the treatment is selected

using Z , and 82% if the treatment is selected using X . This suggested that 10% patients might be assigned to the wrong treatment if model (3.1) is true and the treatment is selected using Z . The estimated mean restricted lifetime within 5 years is 3.85 if all patients are assigned to surgery alone, and 4.10 if all are assigned to surgery plus chemotherapy. The estimated mean restricted lifetime within 5 years is 4.40 when the treatment is selected using X while 4.13 when the treatment is selected using Z . There is no significant difference between the mean restricted life times (95% confidence interval $-0.07, 0.61$). With the consideration of the possible side effects of chemotherapy, the treatment selection based on the new marker seems to be better as less patients are assigned to surgery plus chemotherapy.

6. DISCUSSION

We have proposed a novel method to evaluate a new biomarker based on data from a reproducibility study. Our approach assumes that the reproducibility study samples and the clinical trial samples are randomly drawn from a common target population. It is beyond the scope of this article to address the impact of violations of this assumption.

We have aimed on maximizing the population mean-restricted lifetime. This method can be extended to other statistical measures, for example, L -year survival rate, or a more flexible utility function that incorporates notions of cost and quality of life. Although we have focused on survival time as outcomes, the approach can be adapted to discrete and continuous outcomes with minor modifications.

For simplicity, we have adopted proportional hazards models as working models. The estimation of mean restricted lifetime under the treatment selection criterion based on the working model is still valid even if the model is misspecified. Our approaches may also be extended to more flexible models, for example, by allowing time-varying treatment or covariate effects, or including nonparametric functions of covariates in the survival model. Other types of survival models, such as the accelerated failure time model or the additive hazards model, may also be used. Such extensions may warrant further investigation.

In addition, we have assumed a classical measurement error model between the new marker and the standard marker. A more flexible model such as (2.3) with g_1 and g_2 unknown may be adopted and could be estimated using data from the reproducibility study with validation data based on parametric models or spline approximation. In some situations, both markers might involve measurement error while the measurement error for the new marker might be smaller. It would be of interest to extend our approaches to accommodate such complexity.

The model of equation (2.2) assumes that the new assay reduces error compared to the existing assay. This is motivated by the fact that often newer technologies provide better measurements than existing technologies. However, the new technologies are often more expensive as well. There is a cost–benefit tradeoff for many new assays. Our modeling approach can help in the quantification of how large the potential benefit might be from a new assay that is under consideration. This can be used to make the “go” versus “no go” decision on whether or not to switch to the new marker in the future. In other words, by quantifying the benefit, our method gives the information needed to make this decision. Another application of our approach is to the setting where a gold standard treatment-selection biomarker exists along with an approximation to the biomarker. The approximation may be based for example on pathological evaluations, like the Magee score as an approximation to the OncotypeDX score (Farrugia and others, 2017). Our approach provides a framework in which to understand the relationship between two such scores as predictors of a clinically important survival outcome.

7. SOFTWARE

The R code and a sample data set are available on GitHub (<https://github.com/xsong88/Evaluate-Biomarkers>).

SUPPLEMENTARY MATERIAL

Supplementary material is available online at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

National Institutes of Health (CA201207 to X.S. and K.K.D.; R43GM134768 to X.S.); and NSF (DMS-1916411 to X.S.).

REFERENCES

AUGENLICHT, L., WADLER, S., CORNER, G., RICHARDS, C., RYAN, L., MULTANI, A., PATHAK, S., BENSON, A., HAILER, D., and HEERDT, B. (1997). Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Research* **99**, 1769–1775.

BOONSTRA, P., MUKHERJEE, B., AND TAYLOR, J. (2013). Bayesian shrinkage methods for partially observed data with many predictors. *The Annals of Applied Statistics* **7**, 2272–2292.

BOONSTRA, P., TAYLOR, J., AND MUKHERJEE, B. (2013). Incorporating auxiliary information for improved prediction in high dimensional datasets: an ensemble of shrinkage approaches. *Biostatistics* **14**, 259–272.

BRINKLEY, J., TSIATIS, A. A., AND ANSTROM, K. J. (2010). A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics* **66**, 512–522.

CAI, T., TIAN, L., WONG, P., AND WEI, L. J. (2011). Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics* **12**, 270–282.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A., AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models*. New York: Chapman and Hall/CRC.

CHEN, P. Y. AND TSITAIIS, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57**, 1030–1038.

DAFNI, U. G. AND TSIATIS, A. A. (1998). Evaluating surrogate markers of clinical outcome measured with error. *Biometrics* **54**, 1445–1462.

FARRUGIA, D., LANDMANN, A., ZHU, L., DIEGO, E., JOHNSON, R., BONAVENTURA, M., SORAN, A., DABBS, D., CLAREK, B., PUHALLA, S. and others (2017). Magee equation 3 predicts pathologic response to neoadjuvant systemic chemotherapy in estrogen receptor positive, her2 negative/equivocal breast tumors. *Modern Pathology* **30**, 1078–1085.

FAUCETT, C. J. AND THOMAS, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.

FINE, J. (2002). Comparing nonnested Cox models. *Biometrika* **89**, 635–647.

GREENE, W. AND CAI, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics* **60**, 987–996.

HENDERSON, R., DIGGLE, P., AND DOBSON, A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **4**, 465–480.

HUANG, Y. AND WANG, C. Y. (2000). Cox regression with accurate covariates unascertainable: a nonparametric correction approach. *Journal of the American Statistical Association* **95**, 1209–1219.

JANES, H., BROWN, M. D., PEPE, M. S., AND HUANG, Y. (2014). An approach to evaluating and comparing biomarkers for patient treatment selection. *International Journal of Biostatistics* **10**, 99–121.

JANES, H., PEPE, M. S., BOSSUYT, P. M., AND BARLOW, W. E. (2011). Measuring the performance of markers for guiding treatment decisions. *Annals of Internal Medicine* **154**, 253–259.

KLEIN, J. P. AND MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

LI, Y. AND RYAN, L. (2006). Inference on survival data with covariate measurement error an imputation-based approach. *Scandinavian Journal of Statistics* **33**, 169–190.

PENG, L. AND HUANG, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103**, 637–649.

PRENTICE, R. (1982). Covariate measurement errors and parameter estimates in a failure time regression model. *Biometrika* **69**, 331–342.

SONG, X., DAVIDIAN, M., AND TSIATIS, A. A. (2002a). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* **3**, 511–528.

SONG, X., DAVIDIAN, M., AND TSIATIS, A. A. (2002b). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753.

SONG, X. AND PEPE, M. S. (2004). Evaluating markers for selecting a patient's treatment. *Biometrics* **60**, 874–883.

SONG, X. AND ZHOU, X. H. (2011). Evaluating markers for treatment selection based on survival outcome. *Statistics in Medicine* **30**, 2251–2264.

THERNEAU, T. M. AND GRAMBSCH, P. M. (2006). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

TSIATIS, A. AND DAVIDIAN, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458.

WANG, C. Y. AND SONG, X. (2016). Robust best linear estimator for Cox regression with instrumental variables in whole cohort and surrogates with additive measurement error in calibration sample. *Biometrical Journal* **58**, 1465–1484.

WANG, C. Y., WANG, N., AND WANG, S. (2000). Regression analysis when covariates are regression parameters of a random effect model for observed longitudinal measurements. *Biometrics* **56**, 487–495.

WULFSOHN, M. S. AND TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

XU, J. AND ZEGER, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.

ZHANG, B., TSIATIS, A. A., LABER, E. B., AND DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68**, 1010–1018.

ZHAO, H. AND TSIATIS, A. A. (1997). A consistent estimator for the distribution of quality adjusted lifetime. *Biometrika* **84**, 339–348.

[Received January 30, 2019; revised March 20, 2020; accepted for publication March 25, 2020]