



Published in final edited form as:

Stat Med. 2016 May 10; 35(10): 1637–1653. doi:10.1002/sim.6820.

Robust Estimation of the Proportion of Treatment Effect Explained by Surrogate Marker Information

Layla Parast^{a,*}, Mary M. McDermott^b, and Lu Tian^c

^a1776 Main Street, RAND Corporation, Santa Monica, CA 90401

^bNorthwestern University, Department of Medicine, Chicago, IL 60611

^cStanford University, Department of Health, Research and Policy, Stanford, CA 94305

Abstract

In randomized treatment studies where the primary outcome requires long follow-up of patients and/or expensive or invasive obtainment procedures, the availability of a surrogate marker that could be used to estimate the treatment effect and could potentially be observed earlier than the primary outcome would allow researchers to make conclusions regarding the treatment effect with less required follow-up time and resources. The Prentice criterion for a valid surrogate marker requires that a test for treatment effect on the surrogate marker also be a valid test for treatment effect on the primary outcome of interest. Based on this criterion, methods have been developed to define and estimate the proportion of treatment effect on the primary outcome that is explained by the treatment effect on the surrogate marker. These methods aim to identify useful statistical surrogates that capture a large proportion of the treatment effect. However, current methods to estimate this proportion usually require restrictive model assumptions that may not hold in practice and thus may lead to biased estimates of this quantity. In this paper, we propose a nonparametric procedure to estimate the proportion of treatment effect on the primary outcome that is explained by the treatment effect on a potential surrogate marker and extend this procedure to a setting with multiple surrogate markers. We compare our approach to previously proposed model-based approaches and propose a variance estimation procedure based on a perturbation-resampling method. Simulation studies demonstrate that the procedure performs well in finite samples and outperforms model-based procedures when the specified models are not correct. We illustrate our proposed procedure using a dataset from a randomized study investigating a group-mediated cognitive behavioral intervention for peripheral artery disease participants.

Keywords

surrogate marker; kernel estimation; robust; nonparametric; treatment effect

*Correspondence to: 1776 Main Street, RAND Corporation, Santa Monica, CA 90401. parast@rand.org.

Supplementary Materials

The reader is referred to the on-line Supplementary Materials for technical appendices.

1. Introduction

Clinical trials aimed at identifying effective treatment and prevention strategies to reduce the risk of a clinical outcome often face a number of key challenges when estimating a treatment effect on outcome risk. In particular, studies often require long term follow-up of patients in order to observe a sufficient number of events to precisely estimate treatment effects [1, 2]. In such settings, the availability of a surrogate marker that could be used to estimate the treatment effect and could be observed earlier than the primary outcome or with less cost or invasiveness to the patient would potentially allow researchers to make conclusions regarding the treatment effect with less required follow-up time and/or less cost [3]. That is, validated surrogate markers could enable shorter randomized clinical trials and require smaller sample sizes, thus accelerating acquisition of clinical information [4].

In one of the most influential papers on the validation of surrogate markers, Prentice [5] defined a criterion for a valid surrogate marker which required that a test for treatment effect on the surrogate marker also be a valid test for treatment effect on the primary outcome of interest. Since his work, a substantial amount of research has led to the development of four major frameworks for evaluating and validating surrogate markers, as described in Joffe & Greene [6]: one based on conditioning on the observed surrogate marker, a second based on defining direct and indirect effects of the treatment on the primary outcome, a third based on a meta-analytic framework, and a fourth based on principal stratification, with the approach proposed by Prentice belonging to the first framework. Methods developed within the first two frameworks have often focused on defining and estimating the proportion of treatment effect on the primary outcome that is explained by the treatment effect on the surrogate marker. Motivated by the Prentice criterion, these methods aim to identify useful statistical surrogates (as opposed to principal surrogates [7]) as those which capture a large proportion of the treatment effect on the primary outcome, which is also the focus of this paper. However, available statistical methods to estimate this proportion have numerous limitations [8, 9, 10, 11]. In particular, current methods usually require restrictive model assumptions that may not hold in practice. For example, Freedman *et al.* [8] proposed to estimate this proportion by examining the change in the regression coefficient for treatment when the surrogate marker is added to a specified regression model. However, when this model is misspecified, the appropriate interpretation of this estimate is not clear [12, 13]. Wang & Taylor [9] propose a much more flexible approach to estimate the proportion of treatment effect explained by defining a quantity that attempts to capture what the effect of the treatment would be if the values of the surrogate marker in the treatment group were distributed as those in the control group. While modeling choices are still required, this approach accommodates various practical settings and has a causal interpretation under certain conditions [13].

It is of great interest to investigate estimation procedures that allow for more flexible assumptions and do not rely on the correct specification of multiple models. Furthermore, when there are multiple surrogate markers of interest, it is difficult to capture the complex relationships between the surrogate markers using restrictive model-based methods. The approach proposed by Freedman *et al.* [8] for the single marker setting can be easily extended to a multiple marker setting by examining the change in the regression coefficient

for treatment when the surrogate markers are added to a specified regression model, though this approach still relies on the correct specification of the models. Xu & Zeger [14] have proposed methods to determine whether multiple markers can improve inference about the treatment effects on a clinical endpoint, but their approach requires specifying parametric models and using Markov Chain Monte Carlo to estimate model parameters.

In this paper, we propose a robust estimation procedure to estimate the proportion of treatment effect on the primary outcome that is explained by the treatment effect on one or more potential surrogate markers in order to identify useful statistical surrogates. For brevity, we will refer to ‘the proportion of treatment effect on the primary outcome that is explained by the treatment effect on a surrogate marker’ as ‘the proportion of treatment effect explained by a surrogate’. In Section 2 we introduce our setup and definitions in a potential outcomes framework and describe the quantity we aim to estimate. In Section 3 we propose our estimation procedure in a single marker and multiple marker setting. We first describe available approaches for estimating the quantity including the model-based approach proposed by Freedman *et al.* [8] and a more flexible though still model-based approach proposed by Wang & Taylor [9]. We propose to estimate the proportion of treatment effect explained by a single potential surrogate marker using a nonparametric approach and then extend this procedure to estimate the proportion of treatment effect explained by a combination of multiple potential surrogate markers. We focus on the setting in which the surrogate marker(s) and primary outcome are fully observed for all individuals in both treatment groups and the surrogate marker (at least one surrogate marker in the multiple surrogate case) is continuous, while the primary outcome can be any general fully observed (uncensored) outcome. In Section 4 we describe the asymptotic properties of our estimates and propose variance estimation procedures. In Section 5 we investigate the finite sample properties of our estimation procedure and compare our proposed procedure to other available methods using simulation studies. In Section 6 we illustrate this procedure using a dataset from a randomized study investigating a group-mediated cognitive behavioral intervention for peripheral artery disease participants.

2. Setup and Definitions in a Causal Inference Framework

Let G be the binary treatment indicator with $G = T$ for treatment and $G = C$ for control (or placebo) and we assume throughout that subjects are randomly assigned to treatment or control at baseline. Let Y and S denote the primary outcome measure and surrogate marker measure, respectively, observed for all subjects $i, i = 1, \dots, n$. Suppose Y is used to estimate and test for a treatment effect, but that Y is expensive or invasive to obtain while S is less expensive or invasive or that S can be obtained earlier than Y . We aim to measure the surrogacy of S by estimating the proportion of treatment effect explained by S .

To define our quantity of interest, we use potential outcomes notation such that $Y^{(g)}$ and $S^{(g)}$ denote the primary outcome and surrogate marker under treatment $G = g$. That is, $Y^{(T)}, Y^{(C)}, S^{(T)}$ and $S^{(C)}$ denote the measures for the primary outcome under the treatment, primary outcome under the control, surrogate marker under the treatment and surrogate marker under the control, respectively. In practice, we only observe $(Y, S) = (Y^{(T)}, S^{(T)})$ or $(Y^{(C)}, S^{(C)})$ depending on whether $G = T$ or C .

Throughout, we define the treatment effect, Δ , as the expected difference in Y under the treatment compared to Y under the control, $\Delta = E(Y^{(T)} - Y^{(C)})$. We aim to measure the surrogacy of a potential surrogate marker using contrasts between the actual treatment effect on Y and the residual treatment effect that would be observed if the surrogate marker is not affected by the treatment. The residual treatment effect can be defined as

$$\Delta_s = \int_{-\infty}^{\infty} E(Y^{(T)} - Y^{(C)} | S^{(T)} = S^{(C)} = s) dF_C(s) = \int_{-\infty}^{\infty} \Delta_s(s) dF_C(s), \quad (1)$$

where $\Delta_s(s) = E(Y^{(T)} - Y^{(C)} | S^{(T)} = S^{(C)} = s)$ and $F_C(\cdot)$ is the marginal cumulative distribution function of $S^{(C)}$, the surrogate marker measure under the control. Note that $F_C(s)$ could similarly be replaced by $F_T(s)$, the marginal cumulative distribution function of $S^{(T)}$ and we assume that the support of S_C and S_T are the same. However, $\Delta_s(s)$ is in general not identifiable since $S^{(T)}$ and $S^{(C)}$ can not be observed simultaneously. To circumvent this difficulty, we assume that

$$Y^{(T)} \perp S^{(C)} | S^{(T)} \quad \text{and} \quad Y^{(C)} \perp S^{(T)} | S^{(C)}. \quad (2)$$

That is, given the surrogate marker value in one group, the surrogate marker value in the other group becomes noninformative to the potential response in the current group. Under assumption (2),

$$\Delta_s(s) = E(Y^{(T)} | S^{(T)} = s) - E(Y^{(C)} | S^{(C)} = s) \quad \text{and} \\ \Delta_s = \int_{-\infty}^{\infty} E(Y^{(T)} | S^{(T)} = s) dF_C(s) - \int_{-\infty}^{\infty} E(Y^{(C)} | S^{(C)} = s) dF_C(s), \quad (3)$$

where the first term in (3) can be interpreted as the expected outcome under treatment if the treatment has no effect on the surrogate marker and the residual treatment effect $\Delta_s(\cdot)$ can be used to measure the surrogacy of S . Note that Δ_s given in (3) only depends on observed quantities and can be viewed as a measure for the direct treatment effect beyond the surrogate marker even without assumption (2). Further discussions on the interpretation of Δ_s can be found in Section 7.

Thus, the proportion of treatment effect explained by the surrogate marker, which we denote by R_S , can be expressed using a contrast between Δ_s and Δ :

$$R_S = \{\Delta - \Delta_s\} / \Delta = 1 - \Delta_s / \Delta. \quad (4)$$

This definition is not new and has been proposed by Wang & Taylor [9]. In addition, Taylor *et al.* [13] provides a detailed description and assessment of the causal interpretation that is possible using the proportion of treatment effect explained by a surrogate marker as a measure of surrogacy. In this paper, we focus on nonparametrically estimating this proportion with a continuous surrogate marker. Informally, we use R_S to measure the extent to which the treatment effect on the surrogate marker captures information about the treatment effect on Y by comparing the total treatment effect with the hypothetical treatment effect when there is no effect of treatment on S i.e. when $F_T(s) = F_C(s)$. In the next section we describe previously proposed methods to estimate this quantity when a single continuous surrogate marker is available which have generally focused on model-based estimators, and

then describe our proposed estimation procedure in this setting. The following section then extends this methodology to a setting with multiple surrogate markers.

3. Estimation

3.1. Single Surrogate Marker

We first focus on a setting where a single continuous surrogate marker, S , is available and our goal is to estimate the proportion of treatment effect explained by S , R_S . Let the treatment effect measure of interest be the expected difference in outcome under treatment compared to control, $\Delta = E(Y^T) - E(Y^C) = E(Y^T|G=T) - E(Y^C|G=C)$ assuming that we are in a randomized treatment setting. The observed data consist of n_T independent identically distributed (i.i.d) copies of (Y^T, S^T) , $\{(Y_{Ti}, S_{Ti}), i = 1, \dots, n_T\}$, from the treatment group $G=T$ and n_C i.i.d copies of (Y^C, S^C) , $\{(Y_{Ci}, S_{Ci}), i = 1, \dots, n_C\}$, from the treatment group $G=C$. One can then estimate the treatment effect as

$$\hat{\Delta} = n_T^{-1} \sum_{i=1}^{n_T} Y_{Ti} - n_C^{-1} \sum_{i=1}^{n_C} Y_{Ci}.$$

Since $R_S = 1 - \beta_2 / \hat{\Delta}$, we now focus on estimating β_2 , the residual treatment effect. As expressed in (3), β_2 aims to capture the expected difference in outcome if the distribution of the surrogate marker in the treatment group was the same as the distribution of the surrogate marker in the control group. One model-based approach to estimate β_2 , proposed by Wang & Taylor [9], is to specify models for $E(Y^g|S^g)$, $g = T, C$ such as:

$$E(Y^C|S^C) = \beta_0 + \beta_1 S^C \quad \text{and} \quad E(Y^T|S^T) = (\beta_0 + \beta_2) + \beta_1 S^T \quad (5)$$

It can be shown that if this model is correctly specified, $\beta_2 = \beta_2$. Thus, reasonable estimates for β_2 and R_S could be $\hat{\beta}_2$ and $1 - \hat{\beta}_2 / \hat{\Delta}$, respectively. This approach is equivalent to that proposed by Freedman *et al.* [8] where an estimate for the proportion of treatment effect explained by a surrogate is obtained by fitting the following two regression models:

$$E(Y|G) = \gamma_0 + \gamma_1 I(G=T) \quad \text{and} \quad E(Y|G, S) = \gamma_{0S} + \gamma_{1S} I(G=T) + \gamma_{2S} S \quad (6)$$

and estimating R_S as $\hat{R}_S^F = 1 - \hat{\gamma}_{1S} / \hat{\gamma}_1$ with $\hat{\gamma}_{1S}$ and $\hat{\gamma}_1$ being the estimators of γ_{1S} and γ_1 , respectively. Here we have used F to indicate that this is the estimate based on Freedman's proposed approach. When all the specified models are linear as in (5) and (6),

$$\hat{R}_S^F = 1 - \hat{\gamma}_{1S} / \hat{\gamma}_1 = 1 - \hat{\beta}_2 / \hat{\Delta}.$$

Alternatively, to allow for additional flexibility, Wang & Taylor [9] suggest that one could include an interaction term when specifying the models used to estimate β_2 :

$$E(S^{(C)}) = \alpha_0 \quad \text{and} \quad E(S^{(T)}) - E(S^{(C)}) = \alpha_1, \\ E(Y^{(C)}|S^{(C)}) = \beta_0 + \beta_1 S^{(C)} \quad \text{and} \quad E(Y^{(T)}|S^{(T)}) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) S^{(T)}. \quad (7)$$

It can be shown that when these models hold, $\hat{s} = \beta_2 + \beta_3 \alpha_0$. Thus, reasonable estimates for \hat{s} , $\hat{\Delta}$, and \hat{R}_S using this approach would be

$$\hat{\Delta}_S^M = \hat{\beta}_2 + \hat{\beta}_3 \hat{\alpha}_0, \quad \hat{\Delta} = \hat{\beta}_2 + \hat{\beta}_1 \hat{\alpha}_1 + \hat{\beta}_3 (\hat{\alpha}_0 + \hat{\alpha}_1) \quad \text{and} \quad \hat{R}_S^M = 1 - \hat{\Delta}_S^M / \hat{\Delta},$$

where we have used $\hat{\Delta}_S^M$ to denote this as the flexible model-based estimator.

While the latter approach is more flexible in terms of model specification given the interaction term, both model-based approaches rely on correct model specification to obtain a consistent estimate. In particular, the limitations of using \hat{R}_S^F as an estimate of the proportion of treatment effect explained by a surrogate have been widely noted and include the reliance of the estimate on the correct specification of both models in (6), which is often not practical, and the difficulty with interpreting the estimate when one or both models do not hold [10, 12, 9].

To develop a more robust approach, we instead propose an estimation procedure to estimate R_S that does not require any model specification and instead nonparametrically estimates \hat{s} . Specifically, we propose to estimate $\mu_T(s) = E(Y^{(T)}|S^{(T)} = s)$ nonparametrically using kernel smoothing and we denote the resulting estimator as $\hat{\mu}_T(s)$. That is,

$$\hat{\mu}_T(s) = \frac{\sum_{i=1}^{n_T} K_h(S_{Ti} - s) Y_{Ti}}{\sum_{i=1}^{n_T} K_h(S_{Ti} - s)}$$

where S_{Ti} is the observed $S^{(T)}$ for person i , Y_{Ti} is the observed $Y^{(T)}$ for person i , $K(\cdot)$ is a smooth symmetric density function with finite support, $K_h(\cdot) = K(\cdot/h)/h$ and h is a specified bandwidth. As in most nonparametric functional estimation procedures, the choice of the smoothing parameter h is critical. To eliminate the impact of the bias of the conditional mean function on the resulting estimator, we require the standard undersmoothing

assumption of $h = O\left(n_T^{-\delta}\right)$ with $\delta \in (1/4, 1/2)$. To obtain an appropriate h we first use the bandwidth selection procedure given by Scott [15] to obtain h_{opt} ; and then we let

$h = h_{opt} n_T^{-c_0}$ for some $c_0 \in (1/20, 3/10)$ to ensure the desired rate for h . In all numerical examples, we chose $c_0 = 0.25$. We then estimate \hat{s} as

$$\hat{\Delta}_S = n_C^{-1} \sum_{i=1}^{n_C} \hat{\mu}_T(S_{Ci}) - n_C^{-1} \sum_{i=1}^{n_C} Y_{Ci} \\ = \sum_{i=1}^{n_C} \frac{\hat{\mu}_T(S_{Ci}) - Y_{Ci}}{n_C}, \quad (8)$$

where S_{Ci} is the observed $S^{(C)}$ for person i and Y_{Ci} is the observed $Y^{(C)}$ for person i . Note that the second term of (8) simply estimates the mean response among those with $G = C$ while the first term attempts to estimate the hypothetical mean response that would have been observed if the distribution of S among those with $G = T$ was the same as that of those with $G = C$. In this way, $\hat{\Delta}_S$ is an estimate of Δ_S , the residual treatment effect after removing the treatment effect attributable to S . Lastly, we estimate R_S as

$$\hat{R}_S = 1 - \hat{\Delta}_S / \hat{\Delta}.$$

Recall from above that it was assumed that the supports of $S^{(T)}$ and $S^{(C)}$ are the same. Since $\mu_T(s)$ is only identifiable on the support of $S^{(T)}$ without additional parametric assumption, this assumption is necessary to estimate $\int \mu_T(s) dF_C(s)$. Otherwise, if the supports of $S^{(T)}$ and $S^{(C)}$ are not dramatically disparate, one may alternatively consider a modified version of this estimation method that uses appropriate spline or local linear smoothing methods [16, 17], which allow mild extrapolation of $\mu_T(s)$ beyond the support of $S^{(T)}$.

3.2. Multiple Surrogate Markers

While work in the area of surrogate marker research has alluded to the need for a valid estimate of the proportion of treatment effect explained by a set of multiple surrogate markers, limited work has been done to propose and investigate robust estimates in this setting [12, 14]. One potential approach would be an extension of Freedman's estimate obtained by fitting the following two models:

$$E(Y|G) = \gamma_0 + \gamma_1 I(G=T) \quad \text{and} \quad E(Y|G, \mathbf{S}) = \gamma_{0S} + \gamma_{1S} I(G=T) + \gamma'_{2S} \mathbf{S} \quad (9)$$

where \mathbf{S} is the vector of candidate surrogate markers, and estimating the surrogacy by

$$\hat{R}_S^F = 1 - \frac{\hat{\gamma}_{1S}}{\hat{\gamma}_1}$$

where, again, $\hat{\gamma}_1$ and $\hat{\gamma}_{1S}$ are estimators for the corresponding regression coefficients.

Alternatively, one may generalize the definition of R_S for a single marker, described above, to multiple markers in a straightforward manner:

$$R_{\mathbf{S}} = 1 - \frac{\Delta_{\mathbf{S}}}{\Delta}, \quad (10)$$

where

$$\Delta_{\mathbf{S}} = \int E(Y^{(T)} - Y^{(C)} | \mathbf{S}^{(T)} = \mathbf{S}^{(C)} = \mathbf{s}) dF_C(\mathbf{s}) = \int E(Y^{(T)} | \mathbf{S}^{(T)} = \mathbf{s}) dF_C(\mathbf{s}) - E(Y^{(C)})$$

under the assumption that

$$Y^{(T)} \perp \mathbf{S}^{(C)} | \mathbf{S}^{(T)} \quad \text{and} \quad Y^{(C)} \perp \mathbf{S}^{(T)} | \mathbf{S}^{(C)}$$

where $F_C(\mathbf{s})$ is the marginal cumulative distribution function of $\mathbf{S}^{(C)}$.

If the model (9) is correctly specified, it is not difficult to show that Freedman's estimate is a consistent estimator of R_S . However, since (9) may not be correctly specified, we consider alternative approaches. To estimate Δ_S , one could consider a flexible model-based approach where models for $E(Y|G, \mathbf{S})$ and $E(S_j|G)$ for each S_j in $\mathbf{S} = \{S_1, \dots, S_p\}$ (where p is the number of surrogate markers) are specified and obtain an estimate of Δ_S assuming these models are correct. Without loss of generality, consider the case where there are three surrogate markers, $\mathbf{S} = \{S_1, S_2, S_3\}$ and one specifies the following linear models:

$$E\left(Y^{(C)} | \mathbf{S}^{(C)}\right) = \beta_0 + \beta_1 S_1^{(C)} + \beta_2 S_2^{(C)} + \beta_3 S_3^{(C)}, \quad (11)$$

$$E\left(Y^{(T)} | \mathbf{S}^{(T)}\right) = (\beta_0 + \beta_4) + (\beta_1 + \beta_5) S_1^{(T)} + (\beta_2 + \beta_6) S_2^{(T)} + (\beta_3 + \beta_7) S_3^{(T)}, \quad (12)$$

and

$$E\left(S_j^{(C)}\right) = \alpha_j, \quad j=1, 2, 3. \quad (13)$$

It can be shown that when these models hold

$$\Delta_S = \beta_4 + \beta_5 \alpha_1 + \beta_6 \alpha_2 + \beta_7 \alpha_3. \quad (14)$$

Thus, reasonable estimates for Δ_S and R_S here would be easily obtained by replacing the unknown regression coefficients in (14) by their consistent estimators. We denote the resulting estimators by $\hat{\Delta}_S^M$ and \hat{R}_S^M , similar to the single marker setting.

With this approach, one could now define a single “pseudo-marker”

$$W = (\beta_0 + \beta_4) + (\beta_1 + \beta_5) S_1 + (\beta_2 + \beta_6) S_2 + (\beta_3 + \beta_7) S_3,$$

using (12). If the surrogacy of \mathbf{S} is defined by (10), then the surrogacy of \mathbf{S} is equivalent to that of W when $W = \mu_T(\mathbf{S}) = E(Y^{(T)} | \mathbf{S}^{(T)} = \mathbf{S})$. Here, the surrogacy of the one-dimensional “marker” W is defined by (3) and (4). The formal justification of this claim is given in Appendix A of the Supplementary Materials. Therefore, we can claim that W has the full surrogacy of all three markers combined, i.e., $R_W = R_S$, which can be estimated based on (14). However, the validity of this approach, no matter how flexible the relevant regression model is, still depends on the correct specification of the model. Under a misspecified model, R_W is different from R_S and, more importantly, the model-based estimator \hat{R}_S^M is not a valid estimator of either of them.

An alternative way to estimate β_0 is to employ a nonparametric procedure. However, when the dimension of \mathbf{S} (the number of potential surrogate markers) is greater than two, a completely nonparametric procedure is infeasible due to the curse of dimensionality [18]. Instead, we propose to use a two-stage procedure combining the aforementioned model-based approach and the nonparametric estimation procedure proposed in Section 3.1. Specifically, our proposed two-stage procedure is based on a dimension reduction approach where we focus on the conditional distribution of $Y^{(T)}|\mathbf{S}^{(T)}$ first. That is, we employ a *working* semiparametric model such as

$$E\left(Y^{(T)}|\mathbf{S}^{(T)}=\mathbf{s}\right)=g\left(\beta' \mathbf{s}\right) \quad (15)$$

to reduce the dimension of \mathbf{S} at the first stage, where $g(\cdot)$ is a monotone increasing function given a priori. Even when the working model is misspecified, the resulting estimator $\hat{\beta}$ based on $\{(Y_{Ti}, \mathbf{S}_{Ti}), i = 1, \dots, n_T\}$ converges to a deterministic limit β_0 in probability as $n_T \rightarrow \infty$ under general regularity conditions. Other commonly used regression models such as the generalized transformation model can also be employed to estimate the conditional expectation $E(Y^{(T)}|\mathbf{S}^{(T)})$ either directly or indirectly. Without loss of generality, we consider $Q=\beta'_0\mathbf{S}$ as the new surrogate marker of interest and in the second stage, apply the approach proposed in the single marker setting to estimate its surrogacy. To be specific, we nonparametrically estimate $E(Y^{(T)}|Q_T=q)$ where $Q_T=\beta'_0\mathbf{S}^{(T)}$ as $\hat{\mu}_T(q)$ based on $\left\{\left(Y_{Ti}, \hat{Q}_{Ti}=\hat{\beta}' \mathbf{S}_{Ti}\right), i=1, \dots, n_T\right\}$ and then estimate β_0 as

$$\hat{\Delta}_S=\sum_{i=1}^{n_C} \frac{\hat{\mu}_T\left(\hat{Q}_{Ci}\right)-Y_{Ci}}{n_C},$$

where $\hat{Q}_{Ci}=\hat{\beta}' \mathbf{S}_{Ci}$. We estimate the proportion of treatment effect captured by \mathbf{S} via Q , R_S as $\hat{R}_S=1-\hat{\Delta}_S/\hat{\Delta}$. Regardless of the correct model specification of (15), $\hat{\Delta}_S$ will always be a consistent estimator of Δ_Q where

$$\Delta_Q=\int E\left(Y^{(T)}|Q_T=q\right) dF_{Q_C}(q)-E\left(Y^{(C)}\right)$$

and $F_{Q_C}(\cdot)$ is the cumulative distribution function of $Q_C=\beta'_0\mathbf{S}^{(C)}$ and \hat{R}_S will always be a consistent estimator of $R_Q=1-\Delta_Q/\Delta$, the surrogacy of $Q=\beta'_0\mathbf{S}$. While the estimator \hat{R}_S is constructed to estimate the full surrogacy of \mathbf{S} , it may not approximate the latter well if the working model (15) fails to characterize the dependence of $Y^{(T)}$ on $\mathbf{S}^{(T)}$. Therefore, in practice one may want to fit a more flexible regression model than (15), e.g. (with a slight abuse of notation) one could assume a working regression model

$$E\left(Y^{(T)}|\mathbf{S}^{(T)}\right)=g\left\{\beta' \mathbf{Z}\left(\mathbf{S}^{(T)}\right)\right\},$$

where $\mathbf{Z}(\mathbf{S})$ is a known transformation of the marker \mathbf{S} including, for example, interactions and nonlinear transformations of the components of \mathbf{S} . In the end, we may use $\hat{Q} = \hat{\beta}' \mathbf{Z}(\mathbf{S})$ as the new surrogate marker for the treatment effect and \hat{R}_S as the estimator for its surrogacy. Note that even though we may not capture the full surrogacy of \mathbf{S} , \hat{Q} and \hat{R}_S could still serve as a good surrogate marker and an accurate assessment of its surrogacy, respectively.

4. Inference and Variance Estimation

It can be shown that under suitable regularity conditions the proposed estimates \hat{R}_S and \hat{R}_Q are consistent estimators of R_S and R_Q , respectively. Furthermore, $n^{\frac{1}{2}} \{ \hat{R}_S - R_S \}$ and $n^{\frac{1}{2}} \{ \hat{R}_Q - R_Q \}$ converge weakly to respective normal distributions. The theoretical justification is provided in Appendix B (single marker setting) and Appendix C (multiple marker setting) of the Supplementary Materials. The variances associated with these asymptotic distributions are difficult to estimate empirically. Therefore, we propose to estimate the variability of our proposed estimators and construct confidence intervals using a perturbation-resampling method, which has been successfully used in many applications [19, 20, 21, 22, 23]. This perturbation-resampling method is similar to the wild bootstrap [24, 25, 26].

Specifically, let $\{ \mathbf{V}^{(b)} = (V_{T_1}^{(b)}, \dots, V_{T_{n_T}}^{(b)}, V_{C_1}^{(b)}, \dots, V_{C_{n_C}}^{(b)})^T, b=1, \dots, D \}$ be $n \times D$ independent copies of a positive random variables V from a known distribution with unit mean and unit variance, such as the standard exponential distribution. In the single marker setting, let

$$\begin{aligned} \hat{\mu}_T^{(b)}(s) &= \frac{\sum_{i=1}^{n_T} V_{T_i}^{(b)} K_h(S_{T_i} - s) Y_{T_i}}{\sum_{i=1}^{n_T} V_{T_i}^{(b)} K_h(S_{T_i} - s)}, \\ \hat{\Delta}_S^{(b)} &= \frac{\sum_{i=1}^{n_C} V_{C_i}^{(b)} \{ \hat{\mu}_T^{(b)}(S_{C_i}) - Y_{C_i} \}}{\sum_{i=1}^{n_C} V_{C_i}^{(b)}}, \\ \hat{\Delta}^{(b)} &= \frac{\sum_{i=1}^{n_T} V_{T_i}^{(b)} Y_{T_i}}{\sum_{i=1}^{n_T} V_{T_i}^{(b)}} - \frac{\sum_{i=1}^{n_C} V_{C_i}^{(b)} Y_{C_i}}{\sum_{i=1}^{n_C} V_{C_i}^{(b)}}, \end{aligned}$$

and

$$\hat{R}_S^{(b)} = 1 - \hat{\Delta}_S^{(b)} / \hat{\Delta}^{(b)}.$$

Then one can estimate the distribution of

$$\begin{pmatrix} \hat{\Delta}_S - \Delta_S \\ \hat{\Delta} - \Delta \end{pmatrix}$$

by the empirical distribution of

$$\begin{pmatrix} \hat{\Delta}_S^{(b)} - \hat{\Delta}_S \\ \hat{\Delta}^{(b)} - \hat{\Delta} \end{pmatrix}, \quad b=1, \dots, D.$$

For example, the variance of former can be approximated by $\hat{\Sigma}$, the empirical variance of the latter conditional on the observed data. To construct a $100(1 - \alpha)\%$ confidence interval for R_S , one can calculate the $100(\alpha/2)^{\text{th}}$ and $100(1 - \alpha/2)^{\text{th}}$ empirical percentile of $\hat{R}_S^{(b)}$ or estimate the variance of $\hat{R}_S - R_S$ by the empirical variance of $\hat{R}_S^{(b)} - \hat{R}_S$ and construct the corresponding Wald-type confidence interval. An alternative is to employ Fieller's method for making inference on the ratio of two parameters [27, 28] and obtain the $100(1 - \alpha)\%$ confidence interval for R_S as

$$\left\{ 1 - r: \frac{(\hat{\Delta}_S - r\hat{\Delta})^2}{\hat{\sigma}_{11} - 2r\hat{\sigma}_{12} + r^2\hat{\sigma}_{22}} \leq c_\alpha \right\},$$

where $\hat{\Sigma} = (\hat{\sigma}_{ij})_{1 \leq i, j \leq 2}$ and c_α is the $(1 - \alpha)$ th percentile of

$$\left\{ \frac{\left\{ \hat{\Delta}_S^{(b)} - (1 - \hat{R}_S) \hat{\Delta}^{(b)} \right\}^2}{\hat{\sigma}_{11} - 2(1 - \hat{R}_S) \hat{\sigma}_{12} + (1 - \hat{R}_S)^2 \hat{\sigma}_{22}}, b=1, \dots, C \right\}.$$

In the multiple marker setting, let $\hat{Q}_{T_i}^{(b)}$ and $\hat{Q}_{C_i}^{(b)}$ be the perturbed estimates of \hat{Q}_{T_i} and \hat{Q}_{C_i} estimated using weights $\mathbf{V}^{(b)}$, respectively. Specifically, if $\hat{\beta}$ is obtained via solving the estimating equation

$$n_T^{-1} \sum_{i=1}^{n_T} \mathbf{S}_{T_i} \left\{ Y_{T_i} - g(\beta' \mathbf{S}_{T_i}) \right\} = 0$$

then $\hat{\beta}^{(b)}$ is the root of the perturbed estimating equation

$$n_T^{-1} \sum_{i=1}^{n_T} V_{T_i}^{(b)} \mathbf{S}_{T_i} \left\{ Y_{T_i} - g(\beta' \mathbf{S}_{T_i}) \right\} = 0$$

and $\hat{Q}_{g_i}^{(b)} = \mathbf{S}_{g_i}' \hat{\beta}^{(b)}$, for $g = T$ or C . Furthermore, the perturbed counterparts of $\hat{\mu}_T(\cdot)$ and $\hat{\Delta}_S$ are

$$\hat{\mu}_T^{(b)}(q) = \frac{\sum_{i=1}^n V_{T_i}^{(b)} K_h(\hat{Q}_{T_i}^{(b)} - q) Y_{T_i}}{\sum_{i=1}^n V_{T_i}^{(b)} K_h(\hat{Q}_{T_i}^{(b)} - q)} \quad \text{and} \quad \hat{\Delta}_S^{(b)} = \frac{\sum_{i=1}^n V_{C_i}^{(b)} \left\{ \hat{\mu}_T^{(b)}(\hat{Q}_{C_i}^{(b)}) - Y_{C_i} \right\}}{\sum_{i=1}^n V_{C_i}^{(b)}}$$

respectively, and $\hat{R}_S^{(b)}$ can be defined as $1 - \hat{\Delta}_S^{(b)} / \hat{\Delta}^{(b)}$. Again the distribution of

$$\begin{pmatrix} \hat{\Delta}_S^{(b)} - \Delta_Q \\ \hat{\Delta}^{(b)} - \Delta \end{pmatrix}$$

can be approximated by the empirical distribution of

$$\begin{pmatrix} \hat{\Delta}_S^{(b)} - \hat{\Delta}_S \\ \hat{\Delta}^{(b)} - \hat{\Delta} \end{pmatrix}, b=1, \dots, D,$$

when the sample size is large. Therefore, we may make inference for R_Q similarly and obtain confidence intervals as described in the single marker setting. The theoretical justification for the perturbation-resampling procedure is provided in Appendix D of the Supplementary Materials.

5. Simulation Studies

5.1. Single Surrogate Marker

We examined several simulation settings to assess the performance of our proposed estimator, \hat{R}_S , and proposed variance estimation procedure in the single surrogate marker setting. In addition, we compared the performance of our proposed estimator to both model-based estimators described in Section 3.1, \hat{R}_S^F and \hat{R}_S^M . When the models that are required to be specified to obtain \hat{R}_S^F and \hat{R}_S^M are correct, we would expect the proposed estimate and both model-based to be unbiased though the proposed estimate may not be as efficient. However, when these models are not correctly specified, we would expect the model-based estimators to potentially yield biased estimates of R while \hat{R}_S should remain unbiased. Therefore, we examine two main simulation settings with varying values of R_S : one where the models specified in the estimation procedure for \hat{R}_S^M are correct and one where these models are not correctly specified. In both simulation settings the model specification required for \hat{R}_S^F does not hold because we allow for an interaction between the treatment and the surrogate. Throughout all simulations we use a normal density kernel for our proposed estimate. For each simulation setting we present results when $n_T = n_C = 200$ and $n_T = n_C = 1000$ to assess performance and sensitivity to sample size. For all three estimates (\hat{R}_S, \hat{R}_S^M

and \hat{R}_S^F), we estimate variance using our proposed perturbation approach and construct confidence intervals using both the quantiles of the perturbed estimates and Fieller's method.

In the first simulation setting, Setting (i), we generated data such that the models specified in (7) are correct. That is,

$$\begin{aligned} S^{(T)} &\sim N(\alpha_0 + 1, 2) \\ S^{(C)} &\sim N(\alpha_0, 1) \\ Y^{(T)} | S^{(T)} = s &\sim N(3 + 6s, 3) \\ Y^{(C)} | S^{(C)} = s &\sim N(2 + 5s, 3) \end{aligned}$$

where $\alpha_0 = 23, 5$, and $-1/3$ correspond to $R_S = 0.2, 0.5$ and 0.9 , respectively. That is, when $\alpha_0 = 5$, the treatment effect on the surrogate marker explains approximately 50% of the overall treatment effect. The top portion of Table 1 shows the performance of our proposed estimator \hat{R}_S and both model-based estimators, \hat{R}_S^M and \hat{R}_S^F in this setting when $R_S = 0.2, 0.5$, and 0.9 . The proposed estimation procedure performs well with very small bias, an estimated standard deviation using the perturbation approach close to the empirical standard deviation and a coverage level close to the nominal level of 95%. As expected, both \hat{R}_S and \hat{R}_S^M have very small bias and the efficiency loss demonstrated by the proposed estimation approach when the model specified by the flexible model-based approach is correct, is fairly mild. However, \hat{R}_S^F has poor performance with inadequate coverage and large bias compared to the other two estimators as was expected given that this estimation approach incorrectly assumes that there is no interaction between the treatment and the surrogate marker. Performance in general is better when $n_T = n_C = 1000$ compared to when $n_T = n_C = 200$ as expected, with the flexible model-based approach being slightly superior to the proposed method when the smaller sample size is used.

In the second simulation setting, Setting (ii), we generated data such that the models specified in (7) are not correct. Specifically,

$$\begin{aligned} S^{(T)} &\sim \exp\{N(1.7, 0.2)\} \\ S^{(C)} &\sim \exp\{N(1.62, 0.1)\} \\ Y^{(T)} | S^{(T)} = s &\sim \alpha_0 + \alpha_1 s^2 + \alpha_2 \exp(s/5) + \exp\{N(0, 0.3)\} \\ Y^{(C)} | S^{(C)} = s &\sim 0.8s^2 + 0.2 \exp(s/5) + \exp\{N(0, 0.3)\} \end{aligned}$$

where $(\alpha_0, \alpha_1, \alpha_2) = (20, 1, 0.5)$ corresponds to $R_S \approx 0.2$, $(\alpha_0, \alpha_1, \alpha_2) = (0.5, 1, 0.5)$ corresponds to $R_S \approx 0.5$, and $(\alpha_0, \alpha_1, \alpha_2) = (0, 0.82, 0.22)$ corresponds to $R_S \approx 0.9$. The bottom portion of Table 1 shows the performance of our proposed estimator \hat{R}_S and both model-based estimators, \hat{R}_S^M and \hat{R}_S^F in this setting. As expected when the models are misspecified, both model-based estimators have rather large bias and poor coverage with higher bias and poorer coverage as R_S increases, while the proposed estimate is unbiased and has coverage levels close to the nominal level of 95%. For all estimators, the perturbation approach produces standard deviation estimates that are close to the empirical

estimates. As expected, when $n_T = n_C = 200$, the estimates from the proposed method have larger biases than those in the larger sample size setting but still outperform the two model-based approaches.

5.2. Multiple Surrogate Markers

We also examined several simulation settings to assess the performance of our proposed estimator, \hat{R}_S , and proposed variance estimation procedure in the multiple surrogate marker setting, and compared the performance of our proposed estimator to the extension of

Freedman's estimator, \hat{R}_S^F , and the model-based estimator, \hat{R}_S^M , described in Section 3.2.

Since we assumed a working linear regression model for $E(Y^{(T)} | \mathbf{S}^{(T)} = \mathbf{s}) = \beta_0' \mathbf{s}$, both the proposed and model-based estimators aim to estimate R_Q , the surrogacy of $Q = \beta_0' \mathbf{S}$. On the other hand, it can be shown that Freedman's method aims to estimate R_F , the surrogacy of

$\gamma_{2S}' \mathbf{S}$, where γ_{2S} is the limit of the estimated regression coefficient of model (9) as the sample size goes to infinity. As in the previous section, when the specified models are correct, we would expect the proposed estimate and both model-based estimates to be unbiased though we expect some efficiency loss with our proposed estimator. However, when these models are not correctly specified, we would expect the proposed estimator and

its associated inference are still valid for estimating R_Q , while neither \hat{R}_S^M nor \hat{R}_S^F are consistent for their respective quantities (R_Q and R_F respectively). Therefore, we examine two main simulation settings with varying values of R_S : one where the specified models are correct and one where these models are misspecified. In both simulation settings the model

specification required for \hat{R}_S^F does not hold because we allow for an interaction between the treatment and the surrogate. We will also compare R_Q and R_F to the true surrogacy of \mathbf{S} , R_S . Throughout all simulations we use a normal density kernel for our proposed estimate and for each simulation setting we present results when $n_T = n_C = 200$ and $n_T = n_C = 1000$ to assess performance and sensitivity to sample size, as in the single marker settings. For all three

estimates (\hat{R}_S , \hat{R}_S^M and \hat{R}_S^F), we estimate variance using our proposed perturbation approach and construct confidence intervals using both the quantiles of the perturbed estimates and Fieller's method.

In the third simulation setting, Setting (iii), we generated data such that the models specified in (12) and (13) are correct. That is,

$$\begin{aligned} S_1^{(T)} &\sim N(6, 4), & S_1^{(C)} &\sim N(5, 1) \\ S_2^{(T)} &\sim \text{Binom}(0.6), & S_2^{(C)} &\sim \text{Binom}(0.4) \\ S_3^{(T)} &\sim \text{Exp}(1), & S_3^{(C)} &\sim \text{Exp}(2) \\ Y^{(T)} | \mathbf{S}^{(T)} = (s_1, s_2, s_3)' &\sim N(\alpha_0 + 2s_1 + \alpha_1 s_2 + \alpha_2 s_3, 9) \\ Y^{(C)} | \mathbf{S}^{(C)} = (s_1, s_2, s_3)' &\sim N(2 + 1.92s_1 + 3.5s_2 + 2.3s_3, 9) \end{aligned}$$

where $(\alpha_0, \alpha_1, \alpha_2) = (17.5, 4, 2.5)$ corresponds to $R_S = 0.2$, $(\alpha_0, \alpha_1, \alpha_2) = (5.35, 4, 2.5)$ corresponds to $R_S = 0.5$, and $(\alpha_0, \alpha_1, \alpha_2) = (1.99, 3.6, 2.3)$ corresponds to $R_S = 0.9$. The top

portion of Table 2 shows the performance of our proposed estimator \hat{R}_S and both model-based estimators, \hat{R}_S^M and \hat{R}_S^F when $R_S = 0.2, 0.5,$ and 0.9 . The proposed estimation procedure performs well with very small bias, an estimated standard deviation using the perturbation approach close to the empirical standard deviation and a coverage level close to the nominal level of 95%. As expected, both \hat{R}_S and \hat{R}_S^M have very small bias and the efficiency loss demonstrated by the proposed estimation approach when the model specified by the flexible model-based approach is correct, is fairly mild. However, the 95% confidence interval based on \hat{R}_S^F has inadequate coverage due to relatively large bias compared to the other two estimators. It is expected given that this estimation approach assumes there is no interaction between the treatment and the surrogate marker. As in the single marker setting, performances of all methods are in general better when $n_T = n_C = 1000$ compared to when $n_T = n_C = 200$, with the flexible model-based approach being slightly superior to the proposed method especially when the smaller sample size is used.

In the last simulation setting, Setting (iv), we generated data such that the models specified in (12) and (13) are not correct. Specifically,

$$\begin{aligned}
 S_1^{(T)} &\sim N(6, 4), & S_1^{(C)} &\sim N(5, 1) \\
 S_2^{(T)} &\sim \text{Binom}(0.6), & S_2^{(C)} &\sim \text{Binom}(0.4) \\
 S_3^{(T)} &\sim \text{Exp}(1), & S_3^{(C)} &\sim \text{Exp}(2) \\
 Y^{(T)} | \mathbf{S}^{(T)} = (s_1, s_2, s_3)' &\sim \alpha_0 + \alpha_1 s_1^2 s_2 + \alpha_2 s_1 s_2 s_3 + \exp(\alpha_3 s_1 s_2) + \exp\{N(0, 0.09)\} \\
 Y^{(C)} | \mathbf{S}^{(C)} = (s_1, s_2, s_3)' &\sim \alpha_4 + \alpha_5 s_1^2 s_2 + \alpha_6 s_1 s_2 s_3 + \exp(\alpha_7 s_1 s_2) + \exp\{N(0, 0.09)\}
 \end{aligned}$$

where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7) = (47, 0.7, 0.4, 0.2, 0.5, 0.7, 0.39, 0.2)$ corresponds to $R_S \approx 0.2$, $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7) = (9.3, 0.7, 0.4, 0.2, 0.5, 0.5, 0.3, 0.1)$ corresponds to $R_S \approx 0.5$, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7) = (1.2, 0.52, 0.32, 0.1, 0.5, 0.5, 0.3, 0.1)$ corresponds to $R_S \approx 0.9$. The lower portion of Table 2 shows the performance of $\hat{R}_S, \hat{R}_S^M,$ and \hat{R}_S^F in this setting when $R_S \approx 0.2, 0.5,$ and 0.9 . As noted above, since we assume a working linear regression model for $E(Y^{(T)} | \mathbf{S}^{(T)} = \mathbf{s}) = \beta_0' \mathbf{s}$, both the proposed and flexible model-based estimators aim to estimate R_Q , the surrogacy of $Q = \beta_0' \mathbf{S}$, while Freedman's method aims to estimate R^F , the surrogacy of $\gamma_{2S}' \mathbf{S}$. Therefore, in Table 2 we provide both the true R_S (0.2, 0.5 or 0.9) and (R_Q, R^F) and calculate bias, MSE, and coverage with respect to the quantity each estimator aims to estimate. As discussed in Section 3.2, when the working model (15) in our proposed procedure is not correct, the proposed estimator, \hat{R}_S , will always be a consistent estimator of R_Q but will only approximate the true R_S . In contrast, when the models in the flexible model-based procedure, (11) and (12), are not correct, the model-based estimator \hat{R}_S^M is not a consistent estimate of either R_Q or R_S . In this simulation setting, none of the specified models are correct [hatwide] i.e. the working model for the proposed procedure is not correct and the models specified by the flexible model-based procedure and Freedman's approach are also not correct. The proposed estimation

procedure outperforms both model-based estimators in terms of bias and MSE and has better coverage for R_Q . When $R_S = 0.90$ and the working model is misspecified, the coverage of our proposed estimator is slightly lower than the nominal level which may be due to the potential inadequacy of the normal approximation, particularly when R_S is close to 1.

6. Example

To illustrate our proposed estimation procedure we use data from a study of a 6-month group-mediated cognitive behavioral (GMCB) intervention for peripheral artery disease (PAD) participants. Previous results from this study described in McDermott *et al.* [29] showed that the GMCB intervention, which promoted home-based walking exercise, improved distance covered in a 6-minute walk, 12 months after completing the intervention, compared to a control group. The intervention consisted of weekly visits to an exercise facility and incorporated group support and self-regulatory skills ($n_T = 81$) while the control group condition involved weekly on-site group meetings at a medical center where participants received health educational lectures on topics not related to exercise ($n_C = 85$). The primary outcome was the distance the participant completed after 6 minutes of walking up and down a 100-foot hallway at 12 months after randomization. There were three potential surrogate outcomes of interest which were obtained from the Walking Impairment Questionnaire (WIQ), a PAD-specific measure of self-reported limitations with 3 domains: walking distance, walking speed, and stair climbing (all on a scale from 0-100) [30]. Given that measurement of the primary outcome, the distance covered in 6 minutes, requires special supervision and attendance by the patients, we were interested in the questionnaire measures as potential surrogates as they would require fewer resources to collect.

For illustration, we estimate the proportion of treatment effect explained by each of the three potential surrogates alone (walking distance, walking speed and stair climbing from the WIQ), and then estimate the proportion of treatment effect explained by all three surrogates together using our proposed procedure and the model-based procedures. The overall treatment effect, defined as the difference in the change in distance covered in six minutes in the intervention group (average gain of 25.6 meters) compared to the control group (average loss of 7.4 meters), was 33.9 meters. Table 3 shows the resulting estimates of R_S for S equal to each of the three surrogate measures using the proposed estimator, \hat{R}_S , the flexible model-based estimator, \hat{R}_S^M , and Freedman's estimator, \hat{R}_S^F , and corresponding standard deviation estimates and 95% confidence intervals. The estimated proportion of treatment effect explained varied substantially depending on the estimation procedure used. While none of the measures appear to explain a substantial proportion of the treatment effect, self-reported walking speed from the WIQ appears to capture the largest proportion of the treatment effect. Specifically, walking speed explains about 48% of the treatment effect using our proposed estimation procedure, 37% using the flexible model-based procedure and 17% using Friedman's approach. Walking distance from the WIQ also appears to explain a reasonable proportion of the treatment effect while stair climbing explains little of the treatment effect. The last portion of Table 3 shows the estimated quantities for all three surrogate measures together. With the flexible model-based procedure, one could use the

estimated model parameters to construct a single “pseudo-marker” as described in Section 3.2 which would be:

$$W=9.80+0.33(\text{Walking Distance})+0.95(\text{Walking Speed})+0.44(\text{Stair Climbing}).$$

Using our proposed procedure, 48% of the treatment effect is explained by the three measures together while the flexible model-based procedure and Freedman's approach estimate the proportion of treatment effect explained as 38% and 18%, respectively. It is important to note that all procedures produce rather wide confidence intervals in this illustration, we discuss this further below.

7. Discussion

We have proposed a nonparametric procedure to estimate the proportion of treatment effect explained by a single potential surrogate marker and have extended this procedure to a setting with multiple surrogate markers. Specifically, our procedure uses kernel smoothing to estimate the conditional mean of the primary outcome given the surrogate marker under treatment and applies this estimate to the control group to obtain an estimate of the residual treatment effect. In the multiple marker setting, we use a working model to obtain a single summary measure and again use kernel smoothing in an effort to obtain an estimate that is more robust to misspecification of the working model. We have compared our proposed approach to available model-based approaches which require specification of models describing the relationship between the surrogate and the primary outcome and demonstrated through simulations that the proposed procedure outperforms the model-based procedures when the specified models do not hold. In addition, we have proposed a variance estimation procedure based on perturbation-resampling and showed that the resulting variance estimates are close to empirical estimates.

As discussed in Molenberghs *et al.* [11], focusing on the proportion of treatment effect explained by a surrogate as the quantity of interest to capture surrogacy has some limitations. While our definition of the quantity in (4) improves upon the more common definition using coefficients from a linear regression model because it does not rely on correct model specification, the quantity will still tend to be unstable when the treatment effect is close to zero and confidence intervals for the quantity will tend to be very wide unless one has a very large sample size or the treatment effect is large, as we observed in our illustration with PAD participants. Therefore, use of the proportion of treatment effect explained quantity would not be advisable in a study where the treatment effect is small. While we have shown that our proposed estimators can greatly reduce bias by being robust to model misspecification, further research to develop more efficient estimators would be useful.

We note that causal interpretations of the estimated quantity should be approached with care since certain untestable assumptions are required to hold in order for this quantity to truly capture the proportion of treatment effect explained by the surrogate. However, independent of those assumptions, the defined surrogacy quantity could be used to quantify the ability of

the surrogate marker to replace the primary outcome in estimating the treatment effect. Specifically, one could measure the treatment effect based on S only using

$$\tilde{\Delta} = \int \mu(s) dF_T(s) - \int \mu(s) dF_C(s)$$

where $\mu(s)$ is a prediction of Y based on $S = s$ and $F_T(\cdot)$ is the marginal cumulative distribution function of S^T , the surrogate marker under treatment. If $\mu(s) = E(Y^T | S^T = s)$, then the difference that would be observed when $\tilde{\Delta}$ is used as a measure of the treatment effect instead of Δ is exactly (3) i.e. $\Delta - \tilde{\Delta}$. This equivalence does not rely on the assumptions in (2) and suggests that the defined surrogacy quantity has an appropriate interpretation in terms of the bias in replacing the treatment effect on the primary outcome with the expected treatment effect given the surrogate information.

A limitation of our proposed approach is the requirement that the supports of S^T and S^C are equivalent. However, in most practical applications where one aims to estimate the proportion of treatment effect explained, it is more likely that there will be substantial regions of overlap between S^T and S^C and less likely that these two regions of support will be substantially separated. In addition, due to the use of kernel smoothing, care should be taken in using the proposed estimator in small sample size settings; as shown in Section 5, performance may be adequate with sample sizes of 200 in each group but in certain settings, performance can be undesirable.

Finally, we have assumed that both the surrogate marker and primary outcome are fully observed and that the surrogate marker (at least one in the multiple surrogate case) is continuous. In simple settings, where the surrogate marker is discrete, alternatives that do not involve nonparametric smoothing can be used. When the primary outcome of interest is a time-to-event outcome such as time to death or time to diabetes diagnosis, the methods proposed here would require further (nontrivial) extension to handle missing surrogate marker measurements and censoring of the primary outcome itself. Further research in these areas is warranted.

An R package implementing the methods described here, called `Rsurrogate`, is available on CRAN.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Support for this research was provided by National Institutes of Health grants R21DK103118 and R01HL088589.

References

1. Lindström J, Ilanne-Parikka P, Peltonen M, Aunola S, Eriksson JG, Hemiö K, Hämäläinen H, Härkönen P, Keinänen-Kiukaanniemi S, Laakso M, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish diabetes prevention study. *The Lancet*. 2006; 368(9548):1673–1679. doi:10.1016/S0140-6736(06)69701-8.

2. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, Li H, Li H, Jiang Y, An Y, et al. The long-term effect of lifestyle interventions to prevent diabetes in the china da qing diabetes prevention study: a 20-year follow-up study. *The Lancet*. 2008; 371(9626):1783–1789. doi:10.1016/S0140-6736(08)60766-7.
3. Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: cardiovascular diseases. *Statistics in Medicine*. 1989; 8(4):415–425. doi: 10.1002/sim.4780080405. [PubMed: 2727465]
4. National Institute of Diabetes and Digestive and Kidney Diseases. [October 1, 2013] advances and emerging opportunities in diabetes research: a strategic planning report of the Diabetes Mellitus Interagency Coordinating Committee. <http://www2.niddk.nih.gov/AboutNIDDK/ReportsAndStrategicPlanning/DiabetesPlan/PlanPosting.htm>
5. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*. 1989; 8(4):431–440. doi: 10.1002/sim.4780080407. [PubMed: 2727467]
6. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics*. 2009; 65(2): 530–538. doi:10.1111/j.1541-0420.2008.01106.x. [PubMed: 18759836]
7. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58(1):21–29. doi:10.1111/j.0006-341X.2002.00021.x. [PubMed: 11890317]
8. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine*. 1992; 11(2):167–178. doi:10.1002/sim.4780110204. [PubMed: 1579756]
9. Wang Y, Taylor JM. A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics*. 2002; 58(4):803–812. doi: 10.1111/j.0006-341X.2002.00803.x. [PubMed: 12495134]
10. Li Z, Meredith MP, Hoseyni MS. A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in medicine*. 2001; 20(21):3175–3188. doi:10.1002/sim.984. [PubMed: 11746311]
11. Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Controlled Clinical Trials*. 2002; 23(6): 607–625. doi:10.1016/S0197-2456(02)00236-2. [PubMed: 12505240]
12. Lin D, Fleming T, De Gruttola V, et al. Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in medicine*. 1997; 16(13):1515–1527. doi:10.1002/(SICI)1097-0258(19970715)16:13<1515::AID-SIM572>3.0.CO;2-1. [PubMed: 9249922]
13. Taylor JM, Wang Y, Thiébaud R. Counterfactual links to the proportion of treatment effect explained by a surrogate marker. *Biometrics*. 2005; 61(4):1102–1111. doi:10.1111/j.1541-0420.2005.00380.x. [PubMed: 16401284]
14. Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. *Biometrics*. 2001; 57(1):81–87. doi:<http://dx.doi.org/10.1111/j.0006-341X.2001.00081.x>. [PubMed: 11252622]
15. Scott, D. *Multivariate density estimation*. John Wiley & Sons; 1992.
16. Craven P, Wahba G. Smoothing noisy data with spline functions. *Numerische Mathematik*. 1978; 31(4):377–403.
17. Stone CJ. Consistent nonparametric regression. *The annals of statistics*. 1977:595–620.
18. Robins J, Ritov Y. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine*. 1997; 16(3):285–319. doi:10.1002/(SICI)1097-0258(19970215)16:3<285::AID-SIM535>3.0.CO;2-#. [PubMed: 9004398]
19. Park Y, Wei LJ. Estimating subjectspecific survival functions under the accelerated failure time model. *Biometrika*. 2003; 90(3):717–723.
20. Cai T, Tian L, Wei LJ. Semiparametric box–cox power transformation models for censored survival observations. *Biometrika*. 2005; 92(3):619–632. doi:10.1093/biomet/92.3.619.
21. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*. 2007; 102(478):527–537.
22. Tian L, Cai T, Goetghebeur E, Wei LJ. Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika*. 2007; 94(2):297–311. doi:10.1093/biomet/asm036.
23. Cai T, Tian L, Uno H, Solomon S, Wei LJ. Calibrating parametric subject-specific risk estimation. *Journal of the American Statistical Association*. 2010; 97(2):389–404.

24. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*. 1986:1261–1295.
25. Hardle, W. *Applied Nonparametric Regression*. Vol. 27. Cambridge Univ Press; 1990.
26. Mammen E. Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory and Related Fields*. 1992; 93(4):439–455. doi: 10.1007/BF01192716.
27. Fieller EC. Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1954:175–185.
28. Fieller E. The biological standardization of insulin. *Supplement to the Journal of the Royal Statistical Society*. 1940:1–64.
29. McDermott MM, Guralnik JM, Criqui MH, Ferrucci L, Zhao L, Liu K, Domanchuk K, Spring B, Tian L, Kibbe M, et al. Home-based walking exercise in peripheral artery disease: 12-month follow-up of the goals randomized trial. *Journal of the American Heart Association*. 2014; 3(3):e000711. doi:10.1161/JAHA.113.000711. [PubMed: 24850615]
30. Regensteiner JG, Steiner JF, Panzer R, Hiatt WR. Evaluation of walking impairment by questionnaire in patients with peripheral arterial disease. *Journal of Vascular Medicine Biology*. 1990; 2:142–152.

Table 1

Performance of the proposed (\hat{R}_S) and model-based estimators (\hat{R}_S^M and \hat{R}_S^F) in terms of bias, empirical standard deviation (ESD), average standard deviation (ASD) obtained using the proposed perturbation-resampling approach, mean squared error (MSE) and coverage of the 95% confidence intervals obtained using both the quantiles of the perturbed estimates (Quantile) and Fieller's method (Fieller), in Setting (i) where there is a single surrogate marker and the models assumed by the flexible model-based estimator, \hat{R}_S^M , are correctly specified, and Setting (ii) where there is a single surrogate marker and the specified models are not correct with $n_T = n_S = 200$ and with $n_T = n_S = 1000$.

| Setting (i) Model Correctly Specified | | | | | | | | | |
|--|-------------------------------|---------------|---------------|-------------------------------|---------------|---------------|-------------------------------|---------------|---------------|
| $n_T = 200, n_C = 200$ | | | | | | | | | |
| | $R_S = 0.2$ | | | $R_S = 0.5$ | | | $R_S = 0.9$ | | |
| | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F |
| Bias | -0.0023 | -0.0005 | -0.0072 | -0.0066 | -0.0021 | -0.0189 | -0.0072 | 0.0009 | -0.0295 |
| ESD | 0.0260 | 0.0259 | 0.0251 | 0.0457 | 0.0445 | 0.0432 | 0.0559 | 0.0506 | 0.0488 |
| ASD | 0.0262 | 0.0258 | 0.0250 | 0.0466 | 0.0450 | 0.0434 | 0.0561 | 0.0513 | 0.0494 |
| MSE | 0.0007 | 0.0007 | 0.0007 | 0.0021 | 0.0020 | 0.0022 | 0.0032 | 0.0026 | 0.0033 |
| Coverage (Quantile) | 0.946 | 0.941 | 0.938 | 0.945 | 0.941 | 0.914 | 0.926 | 0.942 | 0.885 |
| Coverage (Fieller) | 0.951 | 0.947 | 0.942 | 0.946 | 0.947 | 0.915 | 0.934 | 0.944 | 0.885 |

| $n_T = 1000, n_C = 1000$ | | | | | | | | | |
|--|-------------------------------|---------------|---------------|-------------------------------|---------------|---------------|-------------------------------|---------------|---------------|
| | $R_S = 0.2$ | | | $R_S = 0.5$ | | | $R_S = 0.9$ | | |
| | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F |
| Bias | -0.0003 | 0.0003 | -0.0065 | -0.0014 | -0.0001 | -0.0169 | -0.0031 | -0.0007 | -0.031 |
| ESD | 0.0118 | 0.0115 | 0.0111 | 0.0207 | 0.0196 | 0.0188 | 0.0241 | 0.0214 | 0.0205 |
| ASD | 0.0117 | 0.0115 | 0.0111 | 0.0203 | 0.0196 | 0.0189 | 0.0238 | 0.0216 | 0.0208 |
| MSE | 0.0001 | 0.0001 | 0.0002 | 0.0004 | 0.0004 | 0.0006 | 0.0006 | 0.0005 | 0.0014 |
| Coverage (Quantile) | 0.948 | 0.947 | 0.899 | 0.939 | 0.946 | 0.840 | 0.936 | 0.934 | 0.676 |
| Coverage (Fieller) | 0.948 | 0.950 | 0.900 | 0.940 | 0.951 | 0.835 | 0.941 | 0.941 | 0.682 |

| Setting (ii) Model Misspecified | | | | | | | | | |
|--|-------------------------------|---------------|---------------|-------------------------------|---------------|---------------|-------------------------------|---------------|---------------|
| $n_T = 200, n_C = 200$ | | | | | | | | | |
| | $R_S = 0.2$ | | | $R_S = 0.5$ | | | $R_S = 0.9$ | | |
| | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F |
| Bias | -0.0022 | -0.0141 | -0.0241 | -0.0064 | -0.0364 | -0.0612 | -0.007 | -0.0629 | -0.0844 |
| ESD | 0.0250 | 0.0290 | 0.0275 | 0.0404 | 0.0521 | 0.0492 | 0.0161 | 0.0436 | 0.0404 |
| ASD | 0.0255 | 0.0292 | 0.0277 | 0.0417 | 0.0525 | 0.0497 | 0.0176 | 0.0441 | 0.0414 |

| Setting (ii) Model Misspecified | | | | | | | | | |
|---------------------------------|-------------|---------------|---------------|-------------|---------------|---------------|-------------|---------------|---------------|
| $n_T = 200, n_C = 200$ | | | | | | | | | |
| | $R_S = 0.2$ | | | $R_S = 0.5$ | | | $R_S = 0.9$ | | |
| | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F |
| MSE | 0.0006 | 0.0010 | 0.0013 | 0.0017 | 0.0040 | 0.0062 | 0.0003 | 0.0059 | 0.0088 |
| Coverage (Quantile) | 0.946 | 0.912 | 0.842 | 0.948 | 0.883 | 0.737 | 0.937 | 0.562 | 0.246 |
| Coverage (Fieller) | 0.947 | 0.908 | 0.832 | 0.947 | 0.873 | 0.715 | 0.933 | 0.537 | 0.224 |

| $n_T = 1000, n_C = 1000$ | | | | | | | | | |
|--------------------------|-------------|---------------|---------------|-------------|---------------|---------------|-------------|---------------|---------------|
| | $R_S = 0.2$ | | | $R_S = 0.5$ | | | $R_S = 0.9$ | | |
| | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F |
| Bias | -0.0008 | -0.0134 | -0.0234 | -0.0022 | -0.0335 | -0.0585 | -0.0018 | -0.0587 | -0.0805 |
| ESD | 0.0116 | 0.0133 | 0.0126 | 0.0185 | 0.0236 | 0.0223 | 0.0071 | 0.0192 | 0.0177 |
| ASD | 0.0116 | 0.0133 | 0.0126 | 0.0185 | 0.0236 | 0.0224 | 0.0072 | 0.0191 | 0.0177 |
| MSE | 0.0001 | 0.0004 | 0.0007 | 0.0003 | 0.0017 | 0.0039 | 0.0001 | 0.0038 | 0.0068 |
| Coverage (Quantile) | 0.950 | 0.826 | 0.528 | 0.944 | 0.681 | 0.214 | 0.936 | 0.079 | 0.005 |
| Coverage (Fieller) | 0.951 | 0.818 | 0.521 | 0.948 | 0.673 | 0.201 | 0.942 | 0.067 | 0.002 |

Table 2

Performance of the proposed (\hat{R}_S) and model-based estimators (\hat{R}_S^M and \hat{R}_S^F) in terms of bias, empirical standard deviation (ESD), average standard deviation (ASD) obtained using the proposed perturbation-resampling approach, mean squared error (MSE) and coverage of the 95% confidence intervals obtained using both the quantiles of the perturbed estimates (Quantile) and Fieller's method (Fieller), in Setting (iii) where there are three surrogate markers and the models assumed by the flexible model-based estimator, \hat{R}_S^M , are correctly specified, and Setting (iv) where there are three surrogate markers and the specified models are not correct with $n_T = n_S = 200$ and with $n_T = n_S = 1000$.

| Setting (iii) Model Correctly Specified | | | | | | | | | |
|---|---------------|---------------|--------------|---------------|---------------|--------------|---------------|---------------|-------------|
| $n_T = 200, n_C = 200$ | | | | | | | | | |
| $R_S = 0.20$ | | | $R_S = 0.50$ | | | $R_S = 0.90$ | | | |
| \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S |
| Bias | -0.0023 | -0.0003 | -0.0045 | -0.0049 | 0.000 | -0.0103 | 0.0041 | -0.0014 | -0.0014 |
| ESD | 0.0201 | 0.0189 | 0.0182 | 0.0423 | 0.0385 | 0.0367 | 0.0772 | 0.0745 | 0.0745 |
| ASD | 0.0200 | 0.0191 | 0.0183 | 0.0412 | 0.0389 | 0.0367 | 0.0781 | 0.0746 | 0.0746 |
| MSE | 0.0004 | 0.0004 | 0.0003 | 0.0018 | 0.0015 | 0.0015 | 0.006 | 0.0055 | 0.0055 |
| Coverage (Quantile) | 0.946 | 0.952 | 0.937 | 0.931 | 0.946 | 0.926 | 0.924 | 0.943 | 0.934 |
| Coverage (Fieller) | 0.950 | 0.951 | 0.941 | 0.936 | 0.949 | 0.932 | 0.929 | 0.942 | 0.938 |

| $n_T = 1000, n_C = 1000$ | | | | | | | | | |
|--------------------------|---------------|---------------|--------------|---------------|---------------|--------------|---------------|---------------|-------------|
| $R_S = 0.20$ | | | $R_S = 0.50$ | | | $R_S = 0.90$ | | | |
| \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F | \hat{R}_S |
| Bias | -0.0001 | 0.0004 | -0.0039 | -0.0001 | 0.0011 | -0.0096 | 0.0027 | -0.0036 | -0.0036 |
| ESD | 0.0091 | 0.0086 | 0.0081 | 0.0190 | 0.0175 | 0.0164 | 0.0364 | 0.0323 | 0.0323 |
| ASD | 0.0089 | 0.0085 | 0.0082 | 0.0184 | 0.0173 | 0.0163 | 0.0357 | 0.0322 | 0.0322 |
| MSE | 0.0001 | 0.0001 | 0.0001 | 0.0004 | 0.0003 | 0.0004 | 0.0013 | 0.0011 | 0.0011 |
| Coverage (Quantile) | 0.948 | 0.947 | 0.922 | 0.945 | 0.946 | 0.907 | 0.948 | 0.943 | 0.943 |
| Coverage (Fieller) | 0.950 | 0.948 | 0.923 | 0.948 | 0.947 | 0.908 | 0.948 | 0.950 | 0.945 |

| Setting (iv) Model Misspecified | | | | | | | | | | | |
|---------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $n_T = 200, n_C = 200$ | | | | | | | | | | | |
| $R_S = 0.20$ | | | $R_S = 0.50$ | | | $R_S = 0.90$ | | | | | |
| $R_Q = 0.206$ | | | $R_Q = 0.523$ | | | $R_Q = 0.924$ | | | $R_F = 0.546$ | $R_F = 0.926$ | |
| R_S | \hat{R}_S^M | \hat{R}_S^F | R_S | \hat{R}_S^M | \hat{R}_S^F | R_S | \hat{R}_S^M | \hat{R}_S^F | R_S | \hat{R}_S^M | \hat{R}_S^F |
| Bias | -0.0031 | 0.0190 | -0.0169 | -0.0095 | 0.0462 | -0.0138 | 0.0887 | -0.0850 | -0.0138 | 0.0887 | -0.0850 |
| ESD | 0.0259 | 0.0360 | 0.0283 | 0.0458 | 0.0742 | 0.0298 | 0.0952 | 0.0664 | 0.0298 | 0.0952 | 0.0664 |
| ASD | 0.0260 | 0.0351 | 0.0277 | 0.0463 | 0.0720 | 0.0313 | 0.0934 | 0.0655 | 0.0313 | 0.0934 | 0.0655 |
| MSE | 0.0007 | 0.0017 | 0.0011 | 0.0022 | 0.0076 | 0.0011 | 0.0169 | 0.0116 | 0.0011 | 0.0169 | 0.0116 |
| Coverage (Quantile) | 0.943 | 0.909 | 0.888 | 0.935 | 0.894 | 0.912 | 0.826 | 0.699 | 0.912 | 0.826 | 0.699 |
| Coverage(Fieller) | 0.941 | 0.924 | 0.884 | 0.936 | 0.907 | 0.944 | 0.847 | 0.704 | 0.944 | 0.847 | 0.704 |

| $n_T = 1000, n_C = 1000$ | | | | | | | | | | | |
|--------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $R_S = 0.20$ | | | $R_S = 0.50$ | | | $R_S = 0.90$ | | | | | |
| $R_Q = 0.206$ | | | $R_Q = 0.523$ | | | $R_Q = 0.924$ | | | $R_F = 0.546$ | $R_F = 0.926$ | |
| R_S | \hat{R}_S^M | \hat{R}_S^F | R_S | \hat{R}_S^M | \hat{R}_S^F | R_S | \hat{R}_S^M | \hat{R}_S^F | R_S | \hat{R}_S^M | \hat{R}_S^F |
| Bias | -0.0008 | 0.0186 | -0.0171 | -0.0024 | 0.0468 | -0.0035 | 0.0875 | -0.0859 | -0.0035 | 0.0875 | -0.0859 |
| ESD | 0.0114 | 0.0155 | 0.0121 | 0.0196 | 0.0316 | 0.0125 | 0.0411 | 0.0289 | 0.0125 | 0.0411 | 0.0289 |
| ASD | 0.0117 | 0.0159 | 0.0125 | 0.0203 | 0.0323 | 0.0126 | 0.0411 | 0.0289 | 0.0126 | 0.0411 | 0.0289 |
| MSE | 0.0001 | 0.0006 | 0.0004 | 0.0004 | 0.0032 | 0.0002 | 0.0094 | 0.0082 | 0.0002 | 0.0094 | 0.0082 |
| Coverage (Quantile) | 0.956 | 0.788 | 0.719 | 0.960 | 0.701 | 0.926 | 0.443 | 0.142 | 0.926 | 0.443 | 0.142 |
| Coverage(Fieller) | 0.957 | 0.808 | 0.714 | 0.960 | 0.715 | 0.939 | 0.464 | 0.134 | 0.939 | 0.464 | 0.134 |

Table 3

Estimates, standard deviations (SD), and 95% confidence intervals (CI) using both the quantile method (Quantile) and Fieller's method (Fieller) for estimates of the proportion of treatment effect explained by each of three potential surrogate markers alone (walking distance, walking speed and stair climbing obtained from the WIQ) and the proportion explained by all three together, where the outcome of interest is change in 6 minute walking distance from baseline to 12 months using the proposed estimator, \hat{R}_S , the flexible model-based estimator, \hat{R}_S^M , and Freedman's estimator, \hat{R}_S^F

| | \hat{R}_S | \hat{R}_S^M | \hat{R}_S^F |
|---|--------------|---------------|---------------|
| Potential surrogate: walking distance from the WIQ | | | |
| Estimate | 0.4403 | 0.1701 | 0.0753 |
| SE | 0.22 | 0.1771 | 0.0976 |
| 95% CI (Quantile) | (0.04,0.92) | (-0.07,0.57) | (-0.03,0.31) |
| 95% CI (Fieller) | (0.01,1.01) | (-0.11,0.55) | (-0.08,0.27) |
| Potential surrogate: walking speed from the WIQ | | | |
| Estimate | 0.4808 | 0.3722 | 0.1726 |
| SE | 0.2437 | 0.2839 | 0.1618 |
| 95% CI (Quantile) | (0.13,1.04) | (0.04,1.06) | (-0.02,0.53) |
| 95% CI (Fieller) | (0.08,1.05) | (-0.04,1.11) | (-0.06,0.52) |
| Potential surrogate: stair climbing from the WIQ | | | |
| Estimate | 0.122 | 0.1392 | 0.0695 |
| SE | 0.2099 | 0.1581 | 0.0913 |
| 95% CI (Quantile) | (-0.34,0.47) | (-0.08,0.52) | (-0.02,0.3) |
| 95% CI (Fieller) | (-0.43,0.46) | (-0.17,0.51) | (-0.11,0.28) |
| All three potential surrogates | | | |
| Estimate | 0.4835 | 0.3801 | 0.1771 |
| SE | 0.3236 | 0.3095 | 0.1785 |
| 95% CI (Quantile) | (-0.21,1.06) | (0.04,1.12) | (-0.02,0.61) |
| 95% CI (Fieller) | (-0.28,1.04) | (-0.06,1.07) | (-0.09,0.57) |