Statistica Sinica **29** (2019), 1343-1366
doi:https://doi.org/10.5705/ss.202017.0122

CRITERIA FOR MULTIPLE SURROGATES

Peng Luo[1], Zheng Cai[2] and Zhi Geng[2]

[1]*Shenzhen University and* [2]*Peking University*

Abstract: An observed surrogate endpoint is often used to predict a treatment effect on an unobserved true endpoint when it is difficult or expensive to measure the true endpoint. Although several criteria have been proposed for identifying surrogate endpoints, they all suffer from the surrogate paradox: a treatment has a positive effect on the surrogate and the surrogate has a positive effect on the endpoint; however the treatment has a negative effect on the endpoint. To avoid this paradox, criteria have been proposed for a single surrogate that blocks the path from the treatment to the endpoint. This requires that there is a single path from the treatment to the endpoint and that the surrogate can block this path. However, in many applications, a treatment may affect an endpoint through several paths. Therefore, we use stochastic orders of random vectors to derive criteria for multiple surrogates that avoid the surrogate paradox and can be used to predict the sign of the treatment effect on the unobserved true endpoint. Furthermore under the conditional independence of the treatment and the true endpoint, given the multiple surrogates, we propose sufficient conditions for the sign-equivalence of the treatment effects on the surrogates and on the true endpoint. Lastly, we illustrate how these criteria can be applied to several commonly used models.

Key words and phrases: Average causal effect, prentice criteria, stochastic order, surrogate paradox.

1. Introduction

When a true endpoint of interest is difficult or expensive to measure within a reasonable length of time, researchers often measure a surrogate variable instead. Then, the causal effect of a treatment or an intervention on the unmeasured endpoint is predicted using the causal effect of the treatment on the measured surrogate. Several criteria for identifying reasonable surrogates have been proposed. Prentice (1989) provided a statistical surrogate criterion that requires conditional independence of the treatment and the endpoint, given the surrogate. Frangakis and Rubin (2002) proposed the principal surrogate criterion, which requires the property of causal necessity; in other words, if there is no treatment effect on the surrogate, then this implies that there is no treatment

effect on the endpoint. Lauritzen (2004) depicted a strong surrogate criterion using a causal diagram. This criterion requires that the surrogate blocks the path from the treatment to the endpoint. Gilbert and Hudgens (2008) proposed the average causal necessity and the average causal sufficiency for a reasonable surrogate. Joffe and Greene (2009) summarized related statistical approaches and discussed the relationships between these approaches. However, these criteria all suffer from the surrogate paradox, proposed by Chen, Geng and Jia (2007). According to this paradox, a treatment has a positive effect on the surrogate and the surrogate has a positive effect on the endpoint; however, the treatment has a negative effect on the endpoint. Chen, Geng and Jia (2007) and Ju and Geng (2010) proposed criteria for consistent surrogates that avoid the surrogate paradox based on knowledge of the causation between the surrogate and the endpoint. Wu, He and Geng (2011) proposed sufficient conditions to predict the sign of a treatment effect on the unmeasured endpoint using the sign of the treatment effect on the measured surrogate, based on knowledge of the association between the surrogate and the endpoint. These conditions can be checked empirically using observed data if the endpoint is observed in a validation study. Vanderweele (2013) extended the results of Chen, Geng and Jia (2007) and Ju and Geng (2010) to cases where the treatment has a direct effect on the endpoint. These criteria apply to a single surrogate only. However, in many applications, a treatment or an intervention affects the endpoint through several paths, which means a single surrogate cannot block these paths. For example, a drug may reduce the likelihood of death due to AIDS through two paths: by decreasing HIV-1 RNA, and by increasing the CD4 count. In this case, a single surrogate may not satisfy any criteria of the statistical, principal, and strong surrogates, because both HIV-1 RNA and the CD4 count should be used as surrogates for death due to AIDS. That is, the surrogate paradox may be avoided by using two or more surrogates if it cannot be avoided when using a single surrogate. Joffe (2013) also suggested that it is meaningful to generalize the criteria for a single surrogate to multiple surrogates.

In this paper, we propose criteria for multiple surrogates based on stochastic orders of random vectors. We provide the conditions to avoid the surrogate paradox and use the signs of the treatment effects on multiple surrogates to predict the sign of the treatment effect on the true endpoint. We further propose sufficient conditions for the sign-equivalence of the treatment effects on the surrogates and on the true endpoint under conditional independence of the treatment and the true endpoint, given the multiple surrogates. This can be viewed as

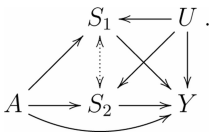

Figure 1. A causal diagram for two surrogates.

a multiple-surrogates version of Prentice's criterion. Furthermore, we illustrate how these criteria can be applied to several commonly used models. The conditions required by the proposed criteria can all be tested if there is a validation trial in which the endpoint is observed. In addition, some of the conditions can be tested if the endpoint has been observed in the control group of a previous trial with the same placebo.

2. Notation and Definitions

Let A denote a randomized treatment, and Y denote the true endpoint of interest. Let $S_1, \ldots, S_p$ denote p potential surrogates. If A has more than two levels, we can compare them in a pairwise manner. Without loss of generality, we suppose A has only two levels, with $A = 1$ for an active drug, and $A = 0$ for a placebo. We suppose that the surrogates $S_1, \ldots, S_p$ are not randomized. Thus, there may be some confounders U that affect both the surrogates $S_1, \ldots, S_p$ and the true endpoint Y. For simplicity, we omit the observed covariate vector Z; as such, our results can be treated as conditional on the observed Z. We depict the causal diagram for $p = 2$ surrogates in Figure 1. The double-headed arrow between S_1 and S_2 means that they are correlated.

Based on the notation of the potential outcome model (Rubin (1974)), let $Y(a, \mathbf{s})$ denote the potential outcome of the true endpoint under the treatment $A = a$ and the surrogate vector $\mathbf{S} = \mathbf{s}$, and let $\mathbf{S}(a) = (S_1(a), \ldots, S_p(a))$ and $Y(a) = Y(a, \mathbf{S}(a))$ denote the potential outcomes of the surrogate vector and the true endpoint, respectively under treatment $A = a$. In addition, we assume that the potential outcomes $S_1(a), \ldots, S_p(a)$ and $Y(a)$ are equal to the observed variables of $S_1, \ldots, S_p$ and Y, respectively, in the treatment group with $A = a$. This assumption is the so-called "consistency assumption" (Angrist, Imbens and Rubin (1996)). Next, we define the average causal effect (ACE).

Definition 1 (ACE)**.** *The ACE of A on Y is defined as*

$$ACE_{A\to Y} = E(Y(A = 1)) - E(Y(A = 0)).$$

In the above definition, $E(Y(a))$ is the average of the potential outcomes

that would be obtained if the treatment $A = a$ were received by all individuals, including those actually receiving the other treatment $A \neq a$.

Next, we define the distribution causal effect (DCE), which is a finer causal measurement than the ACE, for a continuous response Y.

Definition 2 (DCE)**.** *The DCE of A on Y for a specific threshold y is defined as*

$$DCE_{A\to(Y>y)} = P(Y(A=1) > y) - P(Y(A=0) > y).$$

We define $ACE_{A\to S_i}$ and $DCE_{A\to(S_i>s)}$ for a surrogate S_i in a similar way.

To assess the treatment effect on a random vector of multiple surrogates, we introduce three stochastic orders of random vectors and their related inequalities (Rubin (1974)). We denote a value using a lowercase letter, a random variable using an uppercase letter, and their vectors using bold letters. For a random vector $\mathbf{X} = (X_1, \ldots, X_n)$, let $F_{\mathbf{X}}(t_1, \ldots, t_n) = P(X_1 \leq t_1, \ldots, X_n \leq t_n)$ denote the distribution function and $\overline{F}_{\mathbf{X}}(t_1, \ldots, t_n) = P(X_1 > t_1, \ldots, X_n > t_n)$ denote the survival function. For two vector values $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, let $\mathbf{x} \preceq \mathbf{y}$ or $\mathbf{y} \succeq \mathbf{x}$ denote $x_i \leq y_i$ for $i = 1, \ldots, n$. A set $W \subseteq \mathbf{R}^n$ is called an upper (lower) set if $\mathbf{y} \in W$ whenever $\mathbf{y} \succeq (\preceq)\mathbf{x}$ and $\mathbf{x} \in W$. For example, $W_1 = \{(x_1, x_2), \forall x_1 \geq 2 \text{ or } x_2 \geq 2\}$, $W_2 = \{(x_1, x_2), x_1 \geq 1 \ \& \ x_2 \geq 1\}$, and $W_3 = \{(x_1, x_2), \forall x_1 + x_2 \geq c\}$ are three upper sets for $n = 2$.

Definition 3. *For two random vectors $\mathbf{X}$ and $\mathbf{Y}$, we say:*

1. *$\mathbf{X}$ is smaller than $\mathbf{Y}$ in the usual stochastic order (denoted by $\mathbf{X} \leq_{st} \mathbf{Y}$) if $P(\mathbf{X} \in W) \leq P(\mathbf{Y} \in W)$ for all upper sets $W \subseteq \mathbf{R}^n$;*
2. *$\mathbf{X}$ is smaller than $\mathbf{Y}$ in the upper orthant order (denoted by $\mathbf{X} \leq_{uo} \mathbf{Y}$) if $\overline{F}_{\mathbf{X}}(\mathbf{t}) \leq \overline{F}_{\mathbf{Y}}(\mathbf{t})$, for all $\mathbf{t}$;*
3. *$\mathbf{X}$ is smaller than $\mathbf{Y}$ in the lower orthant order (denoted by $\mathbf{X} \leq_{lo} \mathbf{Y}$) if $F_{\mathbf{X}}(\mathbf{t}) \geq F_{\mathbf{Y}}(\mathbf{t})$, for all $\mathbf{t}$.*

For the case of a single response, the three stochastic orders are equivalent and are extensions of the stochastic orders on a single response variable to a response vector. From the above definitions, it is clear that $\mathbf{X} \leq_{st} \mathbf{Y}$ implies both $\mathbf{X} \leq_{uo} \mathbf{Y}$ and $\mathbf{X} \leq_{lo} \mathbf{Y}$; however, the reverse is not true. Corresponding to the three definitions above, we present three inequalities in the Appendix that we use in the proofs of the theorems.

The effects of a binary treatment A on a vector $\mathbf{S} = (S_1, \ldots, S_p)$ of p multiple surrogates are defined by the three stochastic orders between the potential

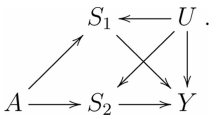

Figure 2. A causal diagram where S_1 and S_2 block the pathways from A to Y.

outcome vectors $\mathbf{S}(1)$ and $\mathbf{S}(0)$ of the surrogates. $\mathbf{S}(A=0) \leq_{st} \mathbf{S}(A=1)$ means that the treatment $A=1$ versus the placebo $A=0$ has a non-negative effect on the surrogate vector in the usual stochastic order. $\mathbf{S}(A=0) \leq_{uo} \mathbf{S}(A=1)$ means that the treatment $A=1$ results in more surrogate outcomes in any upper orthant set than the placebo does, and $\mathbf{S}(A=0) \leq_{lo} \mathbf{S}(A=1)$ means the treatment $A=1$ moves more surrogate outcomes out of any lower orthant set than the placebo does.

We define the strict order and the equal order below.

Definition 4. *For two random vectors* $\mathbf{X}$ *and* $\mathbf{Y}$*, we say:*

1. $\mathbf{X}$ *is strictly smaller than* $\mathbf{Y}$ *in the usual stochastic order (denoted by* $\mathbf{X} <_{st} \mathbf{Y}$*) if* $P(\mathbf{X} \in W) < P(\mathbf{Y} \in W)$ *for all upper sets* $W \subseteq \mathbf{R}^n$*, except for* W *with* $P(\mathbf{X} \in W) = P(\mathbf{Y} \in W) = 0$ *or 1.*
2. $\mathbf{X}$ *is equal to* $\mathbf{Y}$ *in the usual stochastic order (denoted by* $\mathbf{X} =_{st} \mathbf{Y}$*) if* $P(\mathbf{X} \in W) = P(\mathbf{Y} \in W)$ *for all upper sets* $W \subseteq \mathbf{R}^n$*.*

By this definition, $\mathbf{S}(A=0) <_{st} \mathbf{S}(A=1)$ means that the treatment $A=1$ versus the placebo $A=0$ has a positive effect on the surrogate vector, and $\mathbf{S}(A=0) =_{st} \mathbf{S}(A=1)$ means that the treatment $A=1$ versus the placebo $A=0$ has an equal effect on the surrogate vector.

Now, we present a numerical example to illustrate the surrogate paradox, even when two surrogates block all pathways from the treatment to the endpoint. Consider the causal diagram shown in Figure 2, where all variables are binary. Let A be the treatment variable, Y be the true endpoint, S_1 and S_2 be the two surrogate variables blocking all the pathways, and U be an unobserved confounder affecting S_1, S_2, and Y. The probabilities are given as follows: $P(A=1) = 0.767$ and $P(U=1) = 0.639$. The conditional probabilities $P(S_1=1|A,U)$ and $P(S_2=1|A,U)$ are given in Table 1, and $P(Y=1|S_1,S_2,U)$ is given in Table 2.

From the given causal diagram and probabilities, we obtain the following treatment effects:

$$ACE_{A\to S_1} = P(S_1=1 \mid \mathrm{do}(A=1)) - P(S_1=1 \mid \mathrm{do}(A=0)) = 0.018,$$

Table 1. Conditional probabilities of S_1 and S_2, given A and U.

$P(S_1=1\|A,U)$	$U=1$	$U=0$	$P(S_2=1\|A,U)$	$U=1$	$U=0$
$A=1$	0.927	0.028	$A=1$	0.580	0.727
$A=0$	0.397	0.920	$A=0$	0.083	0.824

Table 2. Conditional probabilities of Y given, S_1, S_2, and U.

$P(Y=1\|S_1,S_2,U=1)$	$S_2=1$	$S_2=0$	$P(Y=1\|S_1,S_2,U=0)$	$S_2=1$	$S_2=0$
$S_1=1$	0.410	0.298	$S_1=1$	0.730	0.808
$S_1=0$	0.254	0.222	$S_2=0$	0.469	0.002

$$ACE_{A\to S_2} = P(S_2=1 \mid \mathrm{do}(A=1)) - P(S_2=1 \mid \mathrm{do}(A=0)) = 0.283,$$
$$ACE_{A\to Y} = P(Y=1 \mid \mathrm{do}(A=1)) - P(Y=1 \mid \mathrm{do}(A=0)) = -0.069,$$

and the following causal effects of a surrogate on the endpoint Y, conditional on the other surrogate:

$$P(Y=1 \mid \mathrm{do}(S_1=1), S_2=1) - P(Y=1 \mid \mathrm{do}(S_1=0), S_2=1) = 0.206,$$
$$P(Y=1 \mid \mathrm{do}(S_1=1), S_2=0) - P(Y=1 \mid \mathrm{do}(S_1=0), S_2=0) = 0.228,$$
$$P(Y=1 \mid S_1=1, \mathrm{do}(S_2=1)) - P(Y=1 \mid S_1=1, \mathrm{do}(S_2=0)) = 0.085,$$
$$P(Y=1 \mid S_1=0, \mathrm{do}(S_2=1)) - P(Y=1 \mid S_1=0, \mathrm{do}(S_2=0)) = 0.331.$$

We find that treatment A has positive effects on both surrogates S_1 and S_2, and that each surrogate has a positive effect on the endpoint Y, conditional on the other surrogate. However, treatment A has a negative effect on the endpoint Y.

Next, we illustrate that the surrogate paradox cannot be avoided, even if the surrogate vector $\mathbf{S}(1)$ is larger than $\mathbf{S}(0)$ in the usual stochastic order, in which case, treatment A has a stronger positive effect on the surrogate vector $\mathbf{S}$. From the given causal diagram and probabilities, we obtain the distribution of $(S_1(0), S_2(0))$ and $(S_1(1), S_2(1))$, as shown in Table 3. We have $(S_1(1), S_2(1)) >_{st} (S_1(0), S_2(0))$ because

$$P((S_1(1), S_2(1)) \in W) > P((S_1(0), S_2(0)) \in W),$$

for all upper sets $W = \{(1,1)\}$, $\{(1,1),(1,0)\}$, $\{(1,1),(0,1)\}$, and $\{(1,1),(1,0),(0,1)\}$. $(S_1(1), S_2(1)) >_{st} (S_1(0), S_2(0))$ implies both $ACE_{A\to S_1} > 0$ and $ACE_{A\to S_2} > 0$. As shown above, the surrogate paradox still occurs, even if $(S_1(1), S_2(1)) >_{st} (S_1(0), S_2(0))$. In the next section, we propose criteria that avoid the surrogate paradox.

3. Criteria for Multiple Surrogates Without Models

Table 3. Distributions of $(S_1(0), S_2(0))$ and $(S_1(1), S_2(1))$.

$P(S_1(0), S_2(0))$	$S_1(0) = 1$	$S_1(0) = 0$	$P(S_1(1), S_2(1))$	$S_1(1) = 1$	$S_1(1) = 0$
$S_2(0) = 1$	0.294	0.057	$S_2(1) = 1$	0.351	0.282
$S_2(0) = 0$	0.291	0.358	$S_2(1) = 0$	0.252	0.115

In this section, we discuss the conditions for using multiple surrogates to avoid the surrogate paradox. Based on knowledge of the association between the endpoint and the surrogates, we first discuss the conditions for the implication relationships between the signs of the treatment effects on the observed surrogates and on the unobserved endpoint. Then, we give the conditions for the equivalence relationships between the signs.

3.1. Implication relationships between the signs of the treatment effects on the surrogates and on the endpoint

We first consider the case of knowledge on the expectation of the endpoint Y, conditional on the surrogates $\mathbf{S}$ and the treatment A. For simplicity, let $f(\mathbf{s}, a)$ denote $E(Y \mid \mathbf{s}, a)$. We say that a function $f(\mathbf{s}, a)$ increases in a vector $\mathbf{s} = (s_1, \ldots, s_p)$ if it increases in every element s_i, for $i = 1, \ldots, p$.

Theorem 1. *Suppose we have know the following about conditional expectation* $f(\mathbf{s}, a)$*:*

(1) *Either* $f(\mathbf{s}, 1)$ *or* $f(\mathbf{s}, 0)$ *is a nonconstant increasing function of* $\mathbf{s}$*, and*

(2) $f(\mathbf{s}, 1) \geq f(\mathbf{s}, 0)$*, for all* $\mathbf{s}$*.*

Then, $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$ *implies* $ACE_{A \to Y} \geq 0$*, and* $\mathbf{S}(1) >_{st} \mathbf{S}(0)$ *implies* $ACE_{A \to Y} > 0$*.*

Condition (1) means that at least one of $f(\mathbf{s}, 1)$ and $f(\mathbf{s}, 0)$ is increasing in $\mathbf{s}$. The increase of $f(\mathbf{s}, 0)$ in Condition (1) can be checked if the same placebo was used in previous trials. If Prentice's conditional independence criterion $Y \perp\!\!\!\perp A \mid \mathbf{S}$ holds, then $f(\mathbf{s}, 1) = f(\mathbf{s}, 0)$. Thus, Prentice's conditional independence criterion implies Condition (2) in Theorem 1. Therefore, in order to avoid the surrogate paradox when using Prentice's statistical surrogates, we need to check the monotonicity of $f(\mathbf{s}, 0)$ in $\mathbf{s}$ for the placebo that will be used in the new trial.

Replacing the endpoint Y in the above theorem with the indicator function $I_{\{Y>y\}}$, we obtain the following result on $DCE_{A \to (Y>y)}$. Define $g_y(\mathbf{s}, a) = P(Y > y \mid \mathbf{s}, A = a)$, and assume that $g_y(\mathbf{s}, a)$ is not a constant function of $\mathbf{s}$.

Corollary 1. *Suppose we know the following about the conditional distribution* $g_y(\mathbf{s}, a)$*:*

(1) *Either* $g_y(\mathbf{s}, 1)$ *for any* y *or* $g_y(\mathbf{s}, 0)$ *for any* y *is a nonconstant, non-negative increasing function of* $\mathbf{s}$*, and*

(2) $g_y(\mathbf{s}, 1) \geq g_y(\mathbf{s}, 0)$*, for all* $\mathbf{s}$ *and* y*.*

Then, $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$ *implies* $DCE_{A\to(Y>y)} \geq 0$*, and* $\mathbf{S}(1) >_{st} \mathbf{S}(0)$ *implies* $DCE_{A\to(Y>y)} > 0$*.*

The above two results mean that the ACE and DCE of treatment $A = 1$ versus placebo $A = 0$ on the endpoint Y are non-negative (or positive) if treatment $A = 1$ versus placebo $A = 0$ has a non-negative (or positive) effect on the surrogate vector $\mathbf{S}$ in the usual stochastic order.

Next, we discuss the conditions under which the stochastic order of the surrogate vectors $\mathbf{S}(1)$ and $\mathbf{S}(0)$ can be checked simply by comparing the expectations of the surrogate elements in the treated group with those in the control group. For two multivariate normal vectors $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{X} \leq_{st} \mathbf{Y}$ if and only if $\mu_{\mathbf{X}} \preceq \mu_{\mathbf{Y}}$ and $\mathbf{\Sigma_X} = \mathbf{\Sigma_Y}$ (Rubin (1974, p. 279)), where $\mu_{\mathbf{X}}$ and $\mathbf{\Sigma_X}$ are the mean vector and the covariance matrix, respectively, of $\mathbf{X}$, and $\mu_{\mathbf{Y}}$ and $\mathbf{\Sigma_Y}$ are those of $\mathbf{Y}$. Thus, for a multivariate normal distribution, the stochastic order of the two vectors can be checked by pairwise comparing the elements of the two expectations, yielding the following result.

Corollary 2. *Suppose we know:*

(1) *Either* $f(\mathbf{s}, 1)$ *or* $f(\mathbf{s}, 0)$ *is a nonconstant increasing function of* $\mathbf{s}$*,*

(2) $f(\mathbf{s}, 1) \geq f(\mathbf{s}, 0)$ *for all* $\mathbf{s}$*, and*

(3) *the conditional distribution of the surrogate vector* $\mathbf{S}$*, given* $A = a$*, is the multivariate normal distribution* $N(\mu_a, \mathbf{\Sigma})$ *for* $a = 0$ *and* 1*.*

Then, $ACE_{A\to Y} \geq 0$ *if* $ACE_{A\to S_i} \geq 0$*, for* $i = 1, \ldots, p$*; and* $ACE_{A\to Y} > 0$ *if* $ACE_{A\to S_i} > 0$*, for* $i = 1, \ldots, p$*.*

For the normal distributions, the sign of the ACE of treatment T on the endpoint Y can be predicted by checking the signs of the ACE of treatment T on each surrogate S_i, for $i = 1, \ldots, p$. Actually, Condition (3) in Corollary 2 can be relaxed to the general distributions by replacing (3) with (3′) letting $\mathbf{S}(1)$ and $\mathbf{S}(0)$ have the densities $f(\mathbf{S}(1) - \mu_1 \mid \theta)$ and $f(\mathbf{S}(0) - \mu_0 \mid \theta)$ respectively.

Condition (3′) means that $\mathbf{S}(1)$ and $\mathbf{S}(0)$ have the same density function $f()$ and the same parameters θ, except for the different location parameters μ_1 and μ_0.

Furthermore for the case of two binary surrogates, we have the following result.

Corollary 3. *Suppose we know:*

(1) *Either* $f(\mathbf{s},1)$ *or* $f(\mathbf{s},0)$ *is a nonconstant increasing function of* $\mathbf{s}$,

(2) $f(\mathbf{s},1) \geq f(\mathbf{s},0)$ *for all* $\mathbf{s}$, *and*

(3) $\mathbf{S}$ *has only two binary surrogates* S_1 *and* S_2, *and the odds ratios between* S_1 *and* S_2, *conditional on treatment* A, *do not depend on* A; *that is,* $OR_{S_1,S_2|A=1} = OR_{S_1,S_2|A=0}$, *where* $OR_{S_1,S_2|A=a}$ *is defined as* $P(S_1=0,S_2=0 \mid a)P(S_1=1,S_2=1 \mid a)/\{P(S_1=0,S_2=1 \mid a)P(S_1=1,S_2=0 \mid a)\}$.

Then, $ACE_{A\to Y} \geq 0$ *if* $ACE_{A\to S_i} \geq 0$, *for* $i=1,2$; *and* $ACE_{A\to Y} > 0$ *if* $ACE_{A\to S_i} > 0$, *for* $i=1,2$.

Corollaries 2 and 3 mean that for the cases of a normal surrogate vector and a vector of two binary surrogates, respectively, the stochastic order of two surrogate vectors $\mathbf{S}(1)$ and $\mathbf{S}(0)$ can be checked by pairwise comparison of the expectations of the two corresponding variables in the two vectors if the associations ($\mathbf{\Sigma}_a$ or $OR_{S_1,S_2|a}$) between the surrogates are the same in the two treatment groups of $a=1$ and 0. This makes it easy to check the stochastic order of two vectors $\mathbf{S}(1)$ and $\mathbf{S}(0)$.

Similarly, replacing Conditions (1) and (2) in Corollaries 2 and 3 with Conditions (1) and (2) in Corollary 1, we also have that $DCE_{A\to(Y>y)} \geq 0$ if $ACE_{A\to S_i} \geq 0$, for every $i=1,2,\ldots,p$, and that $DCE_{A\to(Y>y)} > 0$ if $ACE_{A\to S_i} > 0$, for every $i=1,2,\ldots,p$.

If we replace the usual stochastic order $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$ with a weaker order $\mathbf{S}(1) \geq_{uo} \mathbf{S}(0)$ or $\mathbf{S}(1) \leq_{lo} \mathbf{S}(0)$, then, in order to obtain a similar result, we require that the expectation function $f(\mathbf{s},a)$ or the distribution function $g_y(\mathbf{s},a)$ can be factorized to a multiplication form.

Theorem 2. *Suppose we know the following about the conditional expectation* $f(\mathbf{s},a)$*:*

(1) *Either* $f(\mathbf{s},1)$ *or* $f(\mathbf{s},0)$ *can be factorized as* $\prod_i f_i(s_i)$, *where each* $f_i(s_i)$ *is a univariate, non-negative increasing (decreasing) function of* s_i, *and*

(2) $f(\mathbf{s},1) \geq f(\mathbf{s},0)$ *for all* $\mathbf{s}$.

Then, the ACE of A on Y is non-negative if $\mathbf{S}(1) \geq_{uo} (\leq_{lo})\mathbf{S}(0)$.

In the above theorem, we use "increasing (decreasing) ... $\mathbf{S}(1) \geq_{uo} (\leq_{lo})$ $\mathbf{S}(0)$" to denote that "increasing ... $\mathbf{S}(1) \geq_{uo} \mathbf{S}(0)$" may be replaced with "decreasing ... $\mathbf{S}(1) \leq_{lo} \mathbf{S}(0)$." Replacing the expectation function $f(\mathbf{s}, a)$ with the distribution function $g_y(\mathbf{s}, a)$ in the above theorem, we have the following corollary on the DCE.

Corollary 4. *Suppose we know the following about the conditional distribution* $g_y(\mathbf{s}, a)$*:*

(1) *Either* $g_y(\mathbf{s}, 1)$ *for any* y *or* $g_y(\mathbf{s}, 0)$ *for any* y *can be factorized as* $\prod_i g_{yi}(s_i)$*, where each* $g_{yi}(s_i)$ *is a non-negative increasing (decreasing) function of* s_i*, and*

(2) $g_y(\mathbf{s}, 1) \geq g_y(\mathbf{s}, 0)$ *for all* $\mathbf{s}$ *and* y.

Then, $\mathbf{S}(1) \geq_{uo} (\leq_{lo})\mathbf{S}(0)$ *implies* $DCE_{A\to(Y>y)} \geq 0$.

3.2. Equivalence relationships between the signs of the treatment effects on the surrogates and on the endpoint

In the previous subsection, we present only the implication relationships from the signs of the treatment effects on the multiple surrogates to the sign of the treatment effect on the endpoint. However, there may be cases where the treatment has a positive effect on the endpoint; however this cannot be predicted by the signs of the treatment effects on these surrogates. In this subsection, we discuss the conditions necessary for the equivalence relationships between the signs of the treatment effects on the surrogates and on the endpoint.

Theorem 3. *Suppose that*

(1) *Prentice's criterion* $Y \perp\!\!\!\perp A \mid \mathbf{S}$ *holds,*

(2) $E(Y \mid s_1, \ldots, s_p)$ *is a strictly increasing function of* $s_1, \ldots, s_p$*,*

(3) *one of the following three conditions holds:*

- **(a)** *surrogates* $S_1, \ldots, S_p$*, given* A*, are conditionally mutually independent,*
- **(b)** *the conditional distribution of the surrogate vector* $\mathbf{S}$*, given* $A = a$*, is the multivariate normal distribution* $N(\mu_a, \mathbf{\Sigma})$*, for* $a = 0, 1$*,*

(c) **S** *has only two binary surrogates* S_1 *and* S_2*, and the odds ratios between* S_1 *and* S_2*, conditional on treatment* A*, do not depend on* A*; that is,* $OR_{S_1,S_2|A=1} = OR_{S_1,S_2|A=0}$*,*

(4) *all signs of* $DCE_{A\to(S_i>y)}$ *for* $i = 1, \ldots, p$ *are the same (null, positive, or negative).*

Then, $ACE_{A\to Y}$ *has the same sign (null, positive, or negative) as* $DCE_{A\to(S_i>y)}$ *for* $i = 1, 2, \ldots, p$*.*

To predict the sign of the ACE on the endpoint Y in the above theorem, we use the signs of the DCEs on the surrogates **S**. For the normal and binary surrogates ((b) and (c), respectively, in Condition (3)), the signs of the DCE and the ACE on **S** are the same. However, their signs may be different for (a), although the requirement of normal and binary surrogates can be relaxed to being surrogates from a one-parametric exponential family.

The conditions in Theorem 3 ensure not only that positive treatment effects on all surrogates imply a positive treatment effect on the endpoint, but also that a positive treatment effect on the endpoint implies positive treatment effects on the surrogates.

Similarly, by replacing "$E(Y \mid \mathbf{s})$" in Condition (2) in Theorem 3 with "$P(Y > y \mid \mathbf{s})$ for any y," the sign of $DCE_{A\to(Y>y)}$ is the same as that of $DCE_{A\to(S_i>y)}$ if the four conditions are satisfied.

4. Criteria for Multiple Surrogates with Models

In this section, we discuss the conditions necessary to predict the sign of the treatment effect on the unobserved endpoint when the endpoint Y follows the generalized additive model, Cox's proportional hazard model, or the hazard additive model.

4.1. Generalized additive model

First, we consider the following generalized additive model for the expectation $E(Y \mid \mathbf{s}, a)$:

$$g(E(Y \mid \mathbf{s}, a)) = f_1(s_1) + \cdots + f_p(s_p) + f_0(a), \tag{4.1}$$

where g is a known link function, and each f_i may be a parametric, non-parametric, or semi-parametric function. This is an extension of the generalized linear model, and includes the linear model, logistic model, and probit model.

Corollary 5. *For the generalized additive model, suppose that*

(1) *g is a strictly increasing function, and*

(2) *each f_i is an increasing function, for $i = 0, 1, \ldots, p$.*

Then, the ACE of A on Y is non-negative if $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$. Furthermore, if $\mathbf{S}(1) >_{st} \mathbf{S}(0)$ and f_i $(1 \leq i \leq p)$ is not constant in s_i, for some i, then the ACE of A on Y is positive.

For the generalized additive model, the conditions for the equivalence relationship between the effect signs of A on $\mathbf{S}$ and Y are given below.

Corollary 6. *For the generalized additive model, suppose that*

(1) *Prentice's criterion $Y \perp\!\!\!\perp A \mid \mathbf{S}$ holds,*

(2) *g is a strictly increasing function,*

(3) *each f_i $(i = 0, 1, \ldots, p)$ is an increasing function, and there is some f_i $(1 \leq i \leq p)$ that is not constant in s_i,*

(4) *one of the following three conditions holds:*

- **(a)** *the surrogates $S_1, \ldots, S_p$, given A, are conditionally and mutually independent,*
- **(b)** *the conditional distribution of the surrogate vector $\mathbf{S}$, given $A = a$, is the multivariate normal distribution $N(\mu_a, \Sigma)$, for $a = 0, 1$,*
- **(c)** *$\mathbf{S}$ has only two binary surrogates S_1 and S_2, and $OR_{S_1,S_2|A=1} = OR_{S_1,S_2|A=0}$,*

(5) *the signs of $DCE_{A\to(S_i>y)}$ for $i = 1, \ldots, p$ are the same (null, positive, or negative).*

Then, $ACE_{A\to Y}$, $DCE_{A\to(S_i>y)}$, and $ACE_{A\to S_i}$, for all $i = 1, \ldots, p$, have the same sign (null, positive, or negative).

For the following model, the stochastic order $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$ in Corollary 5 may be replaced by a weaker order $\mathbf{S}(1) \leq_{lo} \mathbf{S}(0)$ or $\mathbf{S}(1) \geq_{uo} \mathbf{S}(0)$:

$$E(Y|\mathbf{s}, a) = \sum_{j=1}^{m} \prod_{i=1}^{p} g_{ij}(s_i, a). \tag{4.2}$$

This model (4.2) includes the linear model and the log additive model.

Corollary 7. *For the model given in* (4.2)*, suppose that*

(1) *g_{ij} is a non-negative strictly increasing (decreasing) function of s_i for every i and j, and*

(2) *$g_{ij}(s_i, a = 1) \geq g_{ij}(s_i, a = 0)$.*

Then, the ACE of A on Y is non-negative if $\mathbf{S}(1) \geq_{uo} (\leq_{lo}) \mathbf{S}(0)$.

4.2. Hazard models

In this subsection, we consider two hazard models: the Cox proportional hazard model:

$$\lambda(y \mid \mathbf{s}, a) = \lambda_0(y) \exp\left\{ g_0(a) + \sum_{i=1}^{p} g_i(s_i) \right\},$$

and the hazard additive model:

$$\lambda(y \mid \mathbf{s}, a) = \lambda_0(y) + g_0(a) + \sum_{i=1}^{p} g_i(s_i),$$

where g_i denotes a known function and $\lambda_0(y)$ is a baseline hazard function.

Corollary 8. *For the Cox proportional hazard model, suppose that*

(1) *$g_i(s_i)$ is a univariate decreasing function, for $i = 1, \ldots, p$, and*

(2) *$g_0(1) \leq g_0(0)$.*

Then, the DCE of A on Y is non-negative if $S(1) \geq_{st} S(0)$.

Corollary 9. *For the hazard additive model, suppose that*

(1) *$g_i(s_i)$ denotes all univariate decreasing (increasing) functions, and*

(2) *$g_0(1) \leq g_0(0)$.*

Then, the DCE of A on Y is non-negative if $S(1) \geq_{uo} (\leq_{lo}) S(0)$.

For both hazard models, the conditions for the equivalence relationship between the effect signs of A on $\mathbf{S}$ and Y are given below.

Corollary 10. *For both hazard models, suppose that*

(1) *$Y \perp\!\!\!\perp A \mid \mathbf{S}$, which is equivalent to $g_0(1) = g_0(0)$,*

(2) *$g_i(s_i)$ denotes all univariate decreasing functions, at least one of which is not constant,*

(3) *one of the following conditions holds:*

- **(a)** *the surrogates $S_1, \ldots, S_p$, given A, are conditionally mutually independent,*
- **(b)** *the conditional distribution of the surrogate vector $\mathbf{S}$, given $A = a$, is the multivariate normal distribution $N(\mu_a, \Sigma)$, for $a = 0, 1$,*
- **(c)** $\mathbf{S}$ *has only two binary surrogates S_1 and S_2, and $OR_{S_1,S_2|A=1} = OR_{S_1,S_2|A=0}$,*

(4) *the signs of $DCE_{A\to(S_i>y)}$, for $i = 1, \ldots, p$, are the same (null, positive, or negative).*

Then $DCE_{A\to(Y>y)}$, $ACE_{A\to Y}$, $DCE_{A\to(S_i>y)}$, and $ACE_{A\to S_i}$, for $i = 1, \ldots, p$, have the same sign (null, positive, or negative).

5. Simulation Studies

In this section, we illustrate our results and the sensitivity of their conditions using simulation studies with two surrogates. For two normal surrogates conditional on a treatment, as shown in Corollary 2, we have $ACE_{A\to Y} > 0$ if $ACE_{A\to S_i} > 0$, for $i = 1, 2$, and $E(Y|S_1 = s_1, S_2 = s_2, A = a)$ is an increasing function in s_1, s_2, and a. Below, we show the percentages of correctly predicting $ACE_{A\to Y} > 0$ in 100,000 repetitions. To check the sensitivity of each condition in Corollary 2, we also show the percentages for scenarios in which one of the conditions in Corollary 2 is not satisfied. We consider the following six scenarios for the conditions in Corollary 2:

Scenario 1: All conditions are satisfied,

Scenario 2: Only $ACE_{A\to S_1} > 0$ is not satisfied,

Scenario 3: Only $ACE_{A\to S_2} > 0$ is not satisfied,

Scenario 4: $E(Y \mid S_1 = s_1, S_2 = s_2, A = a)$ is not an increasing function in a,

Scenario 5: $E(Y \mid S_1 = s_1, S_2 = s_2, A = a)$ is not an increasing function in s_1,

Scenario 6: $E(Y \mid S_1 = s_1, S_2 = s_2, A = a)$ is not an increasing function in s_2.

In our simulations, we used a binary randomized treatment A, two normal surrogates S_1 and S_2, a continuous endpoint Y, and an unobserved confounder U that follows a normal distribution. The data-generation mechanism is as follows:

1. $A \sim Bernoulli(0.5)$,
2. the unobserved confounder $U \sim N(0,1)$,
3. $\begin{array}{l} S_1 = \alpha_1 A + \alpha_2 U + \epsilon_1, \\ S_2 = \beta_1 A + \beta_2 U + \epsilon_2, \end{array} \quad \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$,
4. $Y = \gamma_1 A + \gamma_2 S_1 + \gamma_3 S_2 + \gamma_4 U + \epsilon$ and $\epsilon \sim N(0,1)$.

The covariance matrices of (S_1, S_2), conditional on $A = 1$ and $A = 0$, are the same, as required in Corollary 2. For each sample, the parameters α_1, α_2, β_1, β_2, γ_1, γ_2, γ_3, γ_4, and ρ are randomly generated from the following uniform distributions:

1. $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3 \sim U[-3, 3]$,
2. $\gamma_4, \rho \sim U[-1, 1]$.

If S_1 and S_2 block all pathways from A to Y, then there is no direct effect of A on Y, that is, $\gamma_1 = 0$. Because $\gamma_1 \sim U[-3, 3]$ and $P(\gamma_1 = 0) = 0$, there may be direct effect from A to Y in our simulations.

We replicated 100,000 simulation runs with sample sizes 100, 200, 300, and 400. For each simulation, we checked the conditions in Corollary 2 to determine which of the above six scenarios occurs. We rejected a simulation if it did not belong to any of the six scenarios. For each scenario of the accepted simulations, we calculated the percentage of the simulations with a positive $ACE_{A \to Y}$.

From the data-generating process, we can get

$$E(Y|A, S_1, S_2) = \gamma_1 A + \gamma_2 S_1 + \gamma_3 S_2 + \gamma_4(\alpha_2, \beta_2) \begin{pmatrix} \alpha_2^2 + 1, & \alpha_2\beta_2 + \rho \\ \alpha_2\beta_2 + \rho, & \beta_2^2 + 1 \end{pmatrix}^{-1} \begin{pmatrix} S_1 - \alpha_1 A \\ S_2 - \beta_1 A \end{pmatrix}.$$

Thus, the linear model $E(Y \mid A, S_1, S_2) = b_0 + b_1 A + b_2 S_1 + b_3 S_2$ holds. Although we could check the conditions in Corollary 2 using all of the parameters generated in each simulation, we use statistical tests on b_1, b_2, and b_3 in the simulations to mimic what would happen in practice, where these parameters are unknown. We judge that Conditions (1) and (2) of Corollary 2 are satisfied if b_1, b_2, and b_3 are all significantly positive at a significance level of 0.01.

In order to confirm whether $ACE_{A \to S_i} > 0 (i = 1, 2)$ and $ACE_{A \to Y} > 0$, we check $H_0^{(1)}$: $ACE_{A \to S_1} \leq 0$, $H_0^{(2)}$: $ACE_{A \to S_2} \leq 0$, and $H_0^{(3)}$: $ACE_{A \to Y} \leq 0$ using a t-test with a significance level of 0.01 in each case. Note that in our simulations,

Table 4. The percentages of a significant $ACE_{A\to Y} > 0$ under different conditions.

Sample size	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
100	100.00	69.86	68.64	93.14	59.15	58.38
200	100.00	68.99	71.26	91.39	62.84	63.37
300	100.00	70.62	69.91	91.12	65.80	66.22
400	100.00	69.91	69.69	91.46	65.03	64.27

we test $H_0^{(3)}$ to check whether a significant $ACE_{A\to Y} > 0$ can be obtained from the simulated data if Y is observed, although we cannot test $H_0^{(3)}$ because Y is, in general, not observed in real trials. We summarize the percentages of correctly predicting a significant $ACE_{A\to Y} > 0$ for the six scenarios in Table 4. We find that for scenario 1, where all conditions in Corollary 2 are satisfied significantly, a high percentage of the simulations predict a significant $ACE_{A\to Y}$. This might be because we required that all conditions for $ACE_{A\to Y} > 0$ be simultaneously significant in each simulation. For other scenarios, where one of the conditions is violated, the proportions of $ACE_{A\to Y} > 0$ are much smaller than 100%, which means we cannot predict a significant $ACE_{A\to Y} > 0$ correctly. This provides a strong evidence that the conditions of Corollary 2 are essential to predicting a positive treatment effect on the unobserved endpoint.

6. Application to Campaign Finance Reform Experiment Data

To illustrate the application of our criteria for multiple surrogates, we reanalyze the data set used by Druckman and Nelson in their experiment (Druckman and Nelson (2003)). This data set is used to study how citizens' conversations reflect the influence of the elite on public opinion. A total of 78 individuals participated in the experiment. Participants are assigned to two groups randomly. Participants in one group are asked to read an article on a proposed campaign finance reform that emphasizes its possible violation of free speech. The participants in the other group are asked to read an article emphasizing the potential of the reform to limit special interests. Then, the authors measured two surrogates: the participants' perceptions of the importance of free speech and special interests, and their beliefs about the impact of the proposed reform on these items. Finally, the authors measured the outcome variable, namely, the overall level of support for the proposed campaign finance reform.

The two surrogates and the outcome are measured using questionnaires that employ seven-point scales. We denote the random assignment by A, the two surrogates by M (*belief importance*) and W (*belief content*), and the outcome by

Y. High scores of M, W, and Y indicate an increase in the perceived importance, a more positive effect of the reform, and an increase in the support of the finance reform bill, respectively.

Because A is randomized, we can identify $ACE_{A\to M}$, $ACE_{A\to W}$, and $ACE_{A\to Y}$ using $E(M|A=1)-E(M|A=0)$, $E(W|A=1)-E(W|A=0)$, and $E(Y|A=1)-E(Y|A=0)$, respectively, which can be estimated using the sample means. We used the Shapiro–Wilk test to test whether the vector (M,W) follows a joint normal distribution, indicated by a p-value greater than 0.1. Thus, we treat M and W from a two-dimensional normal distribution. We test $H_0:\Sigma_0=\Sigma_1$ using the likelihood ratio test, where Σ_0 and Σ_1 are the covariance matrices of (M,W), given A, equal to zero and one, respectively, to check whether $(M(A=0),W(A=0))$ and $(M(A=1),W(A=1))$ share the same covariance matrix. The likelihood ratio statistic is 3.845 with three degrees of freedom and the p-value is $p=0.279$. Thus, we suppose that the covariance matrices of $(M(A=0),W(A=0))$ and $(M(A=1),W(A=1))$ are the same. The point estimates of the ACE of A on M and W are -0.6882 and 0.5882, respectively, and the corresponding 90 percent confidence intervals are $(-1.288,-0.088)$ and $(0.009,1.168)$, respectively. Thus, we have $ACE_{A\to M}<0$ and $ACE_{A\to W}>0$. Therefore, we let $M^*=-M$ so that A has a positive ACE on M^*. We fit the linear regression model of Y on A, and M^* and W. The coefficients of A, M^*, and W are 0.3094, 0.3028, and 0.3927, respectively. The coefficients of M^* and W are both significant ($p=0.006$ and 0.0007, respectively). As a result, $E(Y|A=a,M^*=m^*,W=w)$ is an increasing function of a, m^*, and w. Based on Corollary 2, we obtain the following implication relationship: $ACE_{A\to M^*}>0$ and $ACE_{A\to W}>0$ imply $ACE_{A\to Y}>0$. This is consistent with the fact that the point estimate of $ACE_{A\to Y}$ is 0.7487 and the 90 percent confidence interval is $(0.121,1.376)$.

Furthermore, the p-value of the hypothesis that the coefficient of A in the linear regression model is equal to zero is 0.35. Thus, we cannot reject the null hypothesis, and may suppose that Prentice's criterion $Y\perp\!\!\!\perp A\mid(M,W)$ is satisfied. We fit the linear regression model of Y on M^* and W. The coefficients of M^* and W are 0.3190 and 0.4063, respectively. The coefficients of M^* and W are both significant ($p=0.006$ and 0.0007, respectively). From Theorem 3, we have that the signs of $DCE_{A\to(M^*>y)}$, $DCE_{A\to(W>y)}$, $ACE_{A\to M^*}$, $ACE_{A\to W}$, and $ACE_{A\to Y}$ are the same.

7. Discussion

The statistical surrogate criterion, principal surrogate criterion, strong surrogate criterion, and consistent surrogate criterion apply to the case of a single surrogate only. In real applications, there are often multiple causal paths from the treatment or exposure to the endpoint. Thus, the conditions required in the above criteria cannot be satisfied and should be extended to cases with multiple surrogates. Previous works have noted that the statistical surrogate criterion, principal surrogate criterion, and strong surrogate criterion cannot avoid the surrogate paradox. Although the consistent surrogate criteria proposed by Chen, Geng and Jia (2007), Ju and Geng (2010), and Vanderweele (2013) can avoid the surrogate paradox, they can only be used for a single surrogate and the conditions in their criteria involve unobserved confounders between a surrogate and an endpoint. Thus, they are untestable, even if the endpoint is observed. In this paper, we have proposed criteria for multiple surrogates that do not involve unobserved confounders. Therefore, these conditions can be tested if there are validation trials in which the endpoint is observed. Furthermore, the monotonicity of $f(\mathbf{s}, 0)$ required in Condition (1) in our theorems and corollaries can be checked if the same control group was used in previous trials where the endpoint Y was observed. Note that the monotonicity is an important additional condition required for Prentice's statistical criterion to avoid the surrogate paradox.

We have proposed a testing approach for easily checking the stochastic orders when the surrogate vector is normal or has only two binary surrogates. This approach can be generalized to the case of a surrogate vector with a mixture of normal and two binary surrogates. For more general cases, nonparametric approaches, such as a goodness-of-fit test, should be considered to test the stochastic orders.

In this paper, we have only discussed criteria for using multiple surrogates to predict the signs of treatment effects on the unobserved endpoint. In some real applications, if an additional validation sample with an observed endpoint Y from a previous clinical trial is also available, then we can try to quantitatively evaluate the treatment effects on the endpoint. Here, the validation sample from the previous trial can be used to provide information on $E(Y|s, a)$, such as its point estimate or its prior distribution. Then, we can use the data from the current trial with observed surrogates, but a missing endpoint, or can combine the data from the current and previous trials to obtain more efficient estimates of $p(s|a)$ and, thus, more efficient estimates of $E(Y|a)$ for $a = 0, 1$. Using the

estimates, we can quantitatively evaluate the treatment effects $ACE_{A\to Y}$ and $DCE_{A\to(Y>y)}$.

For multiple hypothesis testing on the treatment effects on multiple surrogates, the false discovery rate should be considered in future work.

Acknowledgment

The authors thank the associate editor and two reviewers for their valuable comments and suggestions. The authors also thank Dr Lan Liu for her valuable discussion. This research was supported by 863 Program of China (2015AA020507), 973 Program of China (2015CB856000), and NSFC (11331011, 11771028, 91630314).

Appendix: Proofs of theorems and corollaries

Proof of Theorem 1. First, we introduce the following lemma.

Lemma A1. *If* $\mathbf{X} \geq_{st} \mathbf{Y}$ *and* $\phi(\mathbf{x})$ *is any increasing function in* $\mathbb{R}^n$*, then the following inequality holds:*

$$E(\phi(\mathbf{X})) \geq E(\phi(\mathbf{Y})),$$

provided that the expectations exist. In addition, if $\mathbf{X} >_{st} \mathbf{Y}$ *and* $\phi(\mathbf{x})$ *is any increasing function in* $\mathbb{R}^n$ *satisfying* $\phi(\mathbf{X})$ *or* $\phi(\mathbf{Y})$ *is not constant a.s., then the following inequality holds:*

$$E(\phi(\mathbf{X})) > E(\phi(\mathbf{Y})),$$

provided that the expectations exist.

The proof of the first part of the lemma is available in Shaked and Shanthikumar (2007). Below, we prove the second part.

Proof of Lemma 1. Recall that

$$E(X) = \int_0^\infty P(X > c)dc - \int_0^\infty (1 - P(X > -c))\, dc.$$

We have

$$E(\phi(\mathbf{X})) - E(\phi(\mathbf{Y})) = \int_{-\infty}^{\infty} \{P(\phi(\mathbf{X}) > c) - P(\phi(\mathbf{Y}) > c)\}dc.$$

Let $W(c)$ denote $\{\mathbf{x} \mid \phi(\mathbf{x}) > c\}$. Note that $W(c)$ is an upper set because $\phi(\mathbf{x})$ is an increasing function implies that, for $\mathbf{x} \in W(c)$, any $\mathbf{y} \succeq \mathbf{x}$ also implies $\phi(\mathbf{y}) > c$, which leads to $\mathbf{y} \in W(c)$.

Because $\mathbf{X} >_{st} \mathbf{Y}$ implies $P(\phi(\mathbf{X}) > c) - P(\phi(\mathbf{Y}) > c) = P(\mathbf{X} \in W(c)) - P(\mathbf{Y} \in W(c)) \geq 0$, the above integral is non-negative. Let $D = \{c \mid P(\phi(\mathbf{X}) > c) - P(\phi(\mathbf{Y}) > c) > 0\}$. In order to show that the above integral is positive, we need only show that $m(D) > 0$, where $m(\cdot)$ is the Lebesgue measure. By the definition of $\mathbf{X} >_{st} \mathbf{Y}$, $P(\phi(\mathbf{X}) > c) - P(\phi(\mathbf{Y}) > c) = 0$ holds if and only if $P(\phi(\mathbf{X}) > c) = P(\phi(\mathbf{Y}) > c) = 0$ or 1. It is easy to show that $D \supseteq \{c \mid 0 < P(\phi(\mathbf{X}) > c) < 1 \text{ or } 0 < P(\phi(\mathbf{Y}) > c) < 1\}$.

For the case that $\phi(\mathbf{X})$ is not constant a.s., let $a = \inf\{t \mid P(\phi(\mathbf{X}) > t) < 1\}$ and $b = \sup\{t \mid P(\phi(\mathbf{X}) > t) > 0\}$. It is easy to see that $a \leq b$ because $P(\phi(\mathbf{X}) > t)$ is a right-continuous decreasing function in t. If $a = b$, then $P(\phi(\mathbf{X}) > t) = 1$ when $t < a$, and $P(\phi(\mathbf{X}) > t) = 0$ when $t > b$. Because $P(\phi(\mathbf{X}) > t)$ is right-continuous, we have $P(\phi(\mathbf{X}) = a) = 1$, which contradicts that $\phi(\mathbf{X})$ is not constant a.s.. Thus, we have proved that $a < b$. Then, for $\forall a < c < b$, we have $0 < P(\phi(\mathbf{X}) > c) < 1$; thus, $c \in D$. Therefore, $(a, b) \subset D$, and $m(D) \geq b - a > 0$.

For the case that $\phi(\mathbf{Y})$ is not constant a.s., the proof is similar. Therefore we obtain

$$E(\phi(\mathbf{X})) - E(\phi(\mathbf{Y})) = \int_D \{P(\phi(\mathbf{X}) > c) - P(\phi(\mathbf{Y}) > c)\} dc > 0.$$

This completes the proof.

Below, we only consider discrete surrogates $\mathbf{S}$. For the continuous case, the proof is similar. Because $f(\mathbf{s}, a)$ is an increasing function in $\mathbf{R}^p$ ($a = 0$ or 1), we have the following inequality ($a = 0$ or 1), from Lemma 1:

$$\sum_{\mathbf{s}} f(\mathbf{s}, a) P(\mathbf{S}(\mathbf{1}) = \mathbf{s}) \geq \sum_{\mathbf{s}} f(\mathbf{s}, a) P(\mathbf{S}(\mathbf{0}) = \mathbf{s}). \tag{A.1}$$

Because A is randomized, A is independent of the potential outcome; that is, $A \perp\!\!\!\perp (\mathbf{S}(a), Y(a))$. Thus, for the case of $a = 1$ in (3), we have

$$\begin{aligned}
E(Y(1)) &= \sum_{\mathbf{s}} E(Y(1) | \mathbf{S}(\mathbf{1}) = \mathbf{s}) P(\mathbf{S}(1) = \mathbf{s}) \\
&= \sum_{\mathbf{s}} E(Y(1) | \mathbf{S}(\mathbf{1}) = \mathbf{s}, A = 1) P(\mathbf{S}(1) = \mathbf{s}) \\
&= \sum_{\mathbf{s}} E(Y | \mathbf{S} = \mathbf{s}, A = 1) P(\mathbf{S}(1) = \mathbf{s}) \\
&= \sum_{\mathbf{s}} f(\mathbf{s}, 1) P(\mathbf{S}(1) = \mathbf{s}) \geq \sum_{\mathbf{s}} f(\mathbf{s}, 1) P(\mathbf{S}(0) = \mathbf{s}) \\
&\geq \sum_{\mathbf{s}} f(\mathbf{s}, 0) P(\mathbf{S}(0) = \mathbf{s}) = E(Y(0)).
\end{aligned}$$

The last inequality holds from Condition (2).

Similarly, for the case of $a = 0$ in (3), we have

$$\begin{aligned} E(Y(1)) &= \sum_{\mathbf{s}} E(Y(1)|\mathbf{S}(\mathbf{1}) = \mathbf{s})P(\mathbf{S}(1) = \mathbf{s}) \\ &= \sum_{\mathbf{s}} E(Y(1)|\mathbf{S}(\mathbf{1}) = \mathbf{s}, A = 1)P(\mathbf{S}(1) = \mathbf{s}) \\ &= \sum_{\mathbf{s}} E(Y|\mathbf{S} = \mathbf{s}, A = 1)P(\mathbf{S}(1) = \mathbf{s}) \\ &= \sum_{\mathbf{s}} f(\mathbf{s}, 1)P(\mathbf{S}(1) = \mathbf{s}) \geq \sum_{\mathbf{s}} f(\mathbf{s}, 0)P(\mathbf{S}(1) = \mathbf{s}) \\ &\geq \sum_{\mathbf{s}} f(\mathbf{s}, 0)P(\mathbf{S}(0) = \mathbf{s}) = E(Y(0)). \end{aligned}$$

In summary, we have proved that $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$ implies $ACE_{A\to Y} \geq 0$. The proof is similar if $\mathbf{S}(1) >_{st} \mathbf{S}(0)$ is correct instead of $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$, based on the second inequality of Lemma 1.

Proof of Corollary 1. Replacing Y with $I_{Y>y}$ in the above proof, we obtain the result of Corollary 1.

Proof of Corollary 2. Combining $ACE(A \to S_i) \geq 0$, for $i = 1, 2, \ldots, p$, and Condition (3), we have $\mathbf{S}(1) \geq_{st} \mathbf{S}(0)$. Similarly, combining $ACE(A \to S_i) > 0$, for $i = 1, 2, \ldots, p$, and Condition (3), we have $\mathbf{S}(1) >_{st} \mathbf{S}(0)$. In addition, the first two conditions in Corollary 2 are the same as the first two conditions in Theorem 1. Thus, the conclusion of Corollary 2 is correct, based on Theorem 1.

Proof of Corollary 3. Let p_{ija} denote $P(S_1(a) = i, S_2(a) = j)$. To prove $S(1) \geq_{st} S(0)$, we need to show: (1) $p_{111} \geq p_{110}$; (2) $p_{111} + p_{101} \geq p_{110} + p_{100}$; (3) $p_{111} + p_{011} \geq p_{110} + p_{010}$; and (4) $p_{111} + p_{101} + p_{011} \geq p_{110} + p_{100} + p_{010}$. From $ACE_{A\to S_i} \geq 0$, (2) and (3) are correct. To prove (1), we might assume $p_{111} < p_{110}$ first. Then from (2) and (3), we easily obtain $p_{101} > p_{100}$ and $p_{011} > p_{010}$. Because $1 = p_{111} + p_{101} + p_{011} + p_{001} = p_{110} + p_{100} + p_{010} + p_{000}$, we have $p_{011} + p_{001} \leq p_{010} + p_{000}$ from (2), and thus $p_{011} > p_{010}$ implies $p_{001} > p_{000}$. Then, we have $p_{111}p_{001}/(p_{101}p_{011}) > p_{110}p_{000}/(p_{100}p_{010})$ because the left-hand side has a bigger numerator and a smaller denominator, which contradicts the condition $OR_{S_1,S_2|A=1} = OR_{S_1,S_2|A=0}$. Thus, (1) is proved. (4) is equivalent to $p_{001} \leq p_{000}$, and the proof is almost the same as that of (1).

Similarly, we can prove $ACE(A \to Y) > 0$ if $ACE(A \to S_1) > 0$ and $ACE(A \to S_2) > 0$.

Proof of Theorem 2. First, we introduce two lemmas given in Shaked and Shanthikumar (2007).

Lemma A2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two n-dimensional random vectors and* $\mathbf{X} \leq_{uo} \mathbf{Y}$. *For every collection* $\{g_1, g_2, \ldots, g_n\}$ *of univariate non-negative increasing functions, the following inequality holds:*

$$E\left[\prod_{i=1}^{n} g_i(X_i)\right] \leq E\left[\prod_{i=1}^{n} g_i(Y_i)\right].$$

Lemma A3. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two n-dimensional random vectors and* $\mathbf{X} \leq_{lo} \mathbf{Y}$. *For every collection* $\{h_1, h_2, \ldots, h_n\}$ *of univariate non-negative decreasing functions, the following inequality holds:*

$$E\left[\prod_{i=1}^{n} h_i(X_i)\right] \geq E\left[\prod_{i=1}^{n} h_i(Y_i)\right].$$

Similarly to the proof of Theorem 1, based on these lemmas and the conditions of Theorem 2, we obtain the conclusion of Theorem 2.

Proof of Corollary 4. Replacing Y with $I_{Y>y}$ in the above proof, we obtain the result of Corollary 4.

Proof of Theorem 3. First, combining Prentice's criterion and Condition (2), we know that $f(\mathbf{s}, 1) = f(\mathbf{s}, 0) = f(\mathbf{s})$ is strictly increasing. Using the same method as that used in the proofs of Corollaries 2 and 3, it is easy to complete the proof when either the second or the third condition of (3) is correct. Below, we prove the theorem under the first condition of (3).

When the surrogates are conditionally mutually independent, the expectation of the potential outcome can be transformed as follows:

$$\begin{aligned}
E(Y(a)) &= \sum_{\mathbf{s}} E(Y(a) \mid \mathbf{S}(\mathbf{a}) = \mathbf{s}) P(\mathbf{S}(a) = \mathbf{s}) \\
&= \sum_{\mathbf{s}} E(Y \mid \mathbf{S} = \mathbf{s}, A = a) P(\mathbf{S}(a) = \mathbf{s}) \\
&= \sum_{\mathbf{s}} E(Y \mid \mathbf{S} = \mathbf{s}, A = a) P(S_1(a) = s_1) \ldots P(S_p(a) = s_p) \\
&= \sum_{s_p} \cdots \sum_{s_1} f(s_1, \ldots, s_p) P(S_1(a) = s_1) \ldots P(S_p(a) = s_p) \\
&= E_{S_p(a)} \ldots E_{S_1(a)} \{f(S_1(a), \ldots, S_p(a))\}.
\end{aligned}$$

Thus, if the signs of $DCE(A \to (S_i > y))$, for $i = 1, 2, \ldots, p$, are all positive, then $S_i(1) >_{st} S_i(0)$, for all i. Because f is a strictly increasing function, from Lemma 1 we have

$$\begin{aligned} E(Y(1)) &= E_{S_p(1)} \ldots E_{S_1(1)}\{f(S_1(1), \ldots, S_p(1))\} \\ &> E_{S_p(1)} \ldots E_{S_1(0)}\{f(S_1(0), \ldots, S_p(1))\} \\ &> \ldots \\ &> E_{S_p(0)} \ldots E_{S_1(0)}\{f(S_1(0), \ldots, S_p(0))\} \\ &= E(Y(0)). \end{aligned}$$

Therefore, $ACE_{A \to Y}$ is also positive. When the signs of $DCE(A \to (S_i > y))$ are all null or negative, the corresponding results can be derived. This completes the proof.

References

Angrist, J. D., Imbens, J. D. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.

Chen, H., Geng, Z. and Jia, J. (2007). Criteria for surrogate end points. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 919–932.

Druckman, J. and Nelson, K. (2003). Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science* **47**, 729–745.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Gilbert, P. B. and Hudgens, M. G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics* **64**, 1146–1154.

Joffe, M. M. (2013). Discussion of "Surrogate measures and consistent surrogates". *Biometrics* **69**, 572–575.

Joffe, M. M. and Greene, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65**, 530–538.

Ju, C. and Geng, Z. (2010). Criteria for surrogate end points based on causal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 129–142.

Lauritzen, S. L. (2004). Discussion on causality. *Scandinavian Journal of Statistics* **31**, 189–192.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* **8**, 431–440.

Rubin, D. B. (1974). Estimating causal effects of a treatments in randomized and nonrandomized studies. *Ournal of Educational Psychology* **66**, 688–701.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.

Vanderweele, T. J. (2013). Surrogate measures and consistent surrogates. *Biometrics* **69**, 561–581.

Wu, Z., He, P. and Geng, Z. (2011). Sufficient conditions for concluding surrogacy based on observed data. *Statistics in Medicine* **30**, 2422–2434.

Room 406, Sicience building, Shenzhen University, Nanhai Ave 3688, Shenzhen, Guangdong, China, 518060.

E-mail: luopeng@szu.edu.cn

School of Mathematical Sciences, Peking University, Beijing 100871, China.

E-mail: zhengcai@pku.edu.cn

Center for Statistical Science, School of Mathematical Sciences, Peking University, Beijing 100871, China.

E-mail: zhigeng@pku.edu.cn

(Received March 2017; accepted November 2017)