Qualitative implication and equivalence relations between treatment effects on surrogates and endpoints

Zhenguo Wu, Ping He, Chuan Ju and Zhi Geng

Peking University, Beijing, China

Address for correspondence: Zhi Geng, School of Mathematical Sciences, Peking University, Beijing 100871, China

E-mail: zgeng@math.pku.edu.cn

Summary. By the criterion for surrogates proposed by Prentice (1989), a null treatment effect on a surrogate implies a null treatment effect on a true endpoint. In this paper, we show that there may exist simultaneously a positive treatment effect on a surrogate satisfying Prentice's criterion but a negative treatment effect on a true endpoint while the surrogate has a positive association with the endpoint or a positive causal effect on the endpoint. With or without requiring Prentice's criterion, we discuss what conditions are required to qualitatively assess the causal effect of treatment on an unobserved endpoint in terms of an observed surrogate. Rather than a correlation between a surrogate and an endpoint, we require stricter measurements of association for the qualitative assessment. Further we show that these conditions can be satisfied by generalized linear models and Cox's proportional hazard models.

Keywords: Causal effect; Prentice's criterion; Surrogate endpoint

1. Introduction

In a clinical trail where measuring a true endpoint is prohibitively expensive or infeasible, an easily and early available surrogate is often used to reduce the trial cost and duration.

Criteria for validating surrogates are very important for the design of clinical trials. In recent years, there have been a number of papers questioning the validity of surrogates, and many examples of surrogates causing problems in clinical trials have been found, such as CD4 counts for survival time in AIDS studies, bone mass for fracture in osteoporosis studies (Fleming and DeMets, 1996; Baker, 2006; Manns et al., 2006; Alonso and Molenberghs, 2008).

Prentice (1989) proposed a criterion for surrogates which ensures that a test of null hypothesis of no treatment effect on a surrogate is also a valid test of the corresponding null hypothesis of no treatment effect on a true endpoint. A critical condition of the criterion is that a true endpoint is conditionally independent of a treatment given a surrogate, hereafter called Prentice's criterion. Berger (2004) discussed whether Prentice's criterion can validate surrogate endpoints. Frangakis and Rubin (2002) and Rubin (2004) proposed a principal surrogate criterion based on principal stratification. Lauritzen (2004) proposed a strong surrogate criterion in terms of a causal network. Chen et al. (2007) showed that for a principal or strong surrogate, it is possible that a treatment has a positive effect on the surrogate which in turn has a positive effect on a true endpoint, but the treatment may have a negative effect on the true endpoint, which is called the surrogate paradox. Many approaches were proposed to quantitatively evaluate the treatment effect on a true endpoint under assumptions that the ratio of treatment effects on the true endpoint and on the surrogate (called the relative effect) or the density of the true endpoint conditional on the surrogate has been estimated from previous studies or validation trials and that these estimates can be applied to future studies (Burzykowski et al., 2005). In real applications, it may be problematic to extrapolate these estimates from the previous studies to future studies. Chen et al. (2007) and Ju and Geng (2009) discussed conditions for qualitatively assessing a treatment effect on a true endpoint with a surrogate based on the causal network presented by Lauritzen (2004) without satisfying Prentice's criterion. These conditions contain unobserved confounders, and thus they need to be verified by prior knowledge but cannot be tested even if observations of the true endpoint are available. VanderWeele and Robins (2009) discussed transitive relations of non-negative (or say non-strictly positive) effects among variables in a known causal network.

In this paper, we discuss Prentice's criterion and conditions for surrogates via the coun-

terfactual model (Rubin, 1974; Holland, 1986). We first show that a surrogate satisfying Prentice's criterion cannot be used to avoid phenomena of the surrogate paradox or to determine the sign of treatment effect on a true endpoint. Then we discuss what conditions are required to qualitatively assess a treatment effect on a true endpoint with or without requiring Prentice's criterion, and we show the implication and equivalence relations between the sign (positive, non-negative or null) of treatment effect on a surrogate and the sign of treatment effect on a true endpoint. With requiring Prentice's criterion, we propose some additional conditions and models to qualitatively assess a treatment effect on a true endpoint in terms of a surrogate. Intuitively, a condition required for a surrogate is that there exists a correlation between the surrogate and the endpoint. We shall illustrate that the correlation as an additional condition besides Prentice's criterion is not sufficient, and we shall show that stricter measurements of association between a surrogate and a true endpoint are necessary for the qualitative assessment. Generally, Prentice's criterion ensures only an implication relation that a null treatment effect on a surrogate implies a null treatment effect on a true endpoint, but it ensures neither that the converse holds nor that a positive treatment effect on a surrogate implies a positive treatment effect on a true endpoint even if there is a strong association between the surrogate and the true endpoint. We shall show that an equivalence relation between the signs of treatment effects on an endpoint and on a surrogate holds if the surrogate satisfies Prentice's criterion, has the stricter association with the endpoint and follows a distribution from the one-parameter exponential family. Unlike Chen et al. (2007) and Ju and Geng (2009), these conditions and models for surrogates can be tested if there are validation trials in which the true endpoint is observed. We neither need to consider latent confounders which affect both surrogates and endpoints, nor need to assume that the estimates are available from previous studies and are applicable to future studies. Further we show that these additional conditions can be satisfied by generalized linear models and Cox's proportional hazard models. A treatment effect on an unobserved true endpoint can be assessed qualitatively by using an observed surrogate which satisfies the conditions and models. The sign of treatment effect on the true endpoint may be determined by the sign of treatment effect on the surrogate based on the prior knowledge on conditional independence

and models.

In Section 2, we introduce definitions and notation, and then we give an example to illustrate that a phenomenon of the surrogate paradox can occur for a surrogate which satisfies Prentice's criterion. In Section 3, we present conditions for qualitatively assessing a treatment effect on a true endpoint in terms of a surrogate. In Section 4, we discuss conditions for surrogates under some commonly used models, such as the generalized linear models and Cox's proportional hazard models. We show an application to a breast screening study to illustrate our results in Section 5. Finally discussions are given in Section 6. All proofs of theorems are presented in Appendix.

2. Definitions and Notation

Let T denote a treatment, Y a true endpoint, and S a candidate surrogate for the true endpoint. We only compare two levels t' and t'' of a treatment T every time. For a treatment T with more than two levels, we can compare its levels pairwise, and thus without loss of generality we assume that T is binary with values 0 or 1. In this paper, we consider randomized trials.

Prentice's criterion for surrogates requires that an endpoint Y is independent of a treatment T conditionally on the surrogate S, denoted as $Y \perp\!\!\!\perp T|S$. Let $T \perp\!\!\!\perp S$ denote the independence of T and S. Under the Prentice's criterion, it can be shown that $T \perp\!\!\!\perp S$ implies $T \perp\!\!\!\perp Y$. Another condition of Prentice's criterion for surrogates is that there exists an association or a correlation between a surrogate S and a true endpoint Y, denoted as $S \not\perp\!\!\!\perp Y$. We first show an example to illustrate that the sign of treatment effect on an unobserved true endpoint cannot be predicted by the sign of the observed effect on a surrogate which satisfies Prentice's criterion, even if there exists a strong positive association between the surrogate and the true endpoint.

Example 1. Consider a trial with a treatment T, a candidate surrogate S and a true endpoint Y. The joint distribution of T, S and Y is given in Table 1. From the joint distribution, we can obtain that Prentice's criterion $Y \perp\!\!\!\perp T|S$ is satisfied. Although the correlation between S and Y is quite strong ($r(S, Y) = 0.762 > 0$), we have a positive correlation between T and S

($r(T,S) = 0.056$) but a negative correlation between T and Y ($r(T,Y) = -0.039$). For a randomized trial, the correlations $r(T,S)$ and $r(T,Y)$ can also be used as causal measurements. Thus it shows a phenomenon of the surrogate paradox.

Table 1: The joint distribution of T, S and Y satisfying Prentice's criterion

	$S=0$		$S=1$		$S=2$	
	$T=0$	$T=1$	$T=0$	$T=1$	$T=0$	$T=1$
$Y=0$	0.144	0.270	0.032	0.012	0.004	0.012
$Y=1$	0.016	0.030	0.128	0.048	0.076	0.228

Below we define causal measurements via the counterfactual model (Rubin, 1974; Holland, 1986) and corresponding association measurements. Let Y_t and S_t denote the potential outcomes of the true endpoint and the surrogate under a treatment $T = t$ respectively. Assume that the observed outcomes of Y and S are equal to the potential outcomes Y_t and S_t respectively when the treatment $T = t$ is really assigned, that is, $Y = Y_t$ and $S = S_t$ for the really treatment $T = t$. First we introduce an average causal effect (ACE) and a prima facie average causal effect (FACE) as measurements of causation and association respectively. Let X and Z be variables and Z_x a potential outcome of Z under $X = x$.

Definition 1. (ACE and FACE) For $x > x'$, an average causal effect of X on Z is defined as

$$ACE(X \to Z|X = x, X = x') = E(Z_x) - E(Z_{x'}),$$

and the corresponding prima facie average causal effect of X on Z is defined as

$$FACE(X \to Z|X = x, X = x') = E(Z|x) - E(Z|x').$$

$E(Z_x)$ is the average of the potential outcomes which would be obtained if the treatment $X = x$ were assigned to all individuals including those taking other treatments $x' \neq x$. Without randomization or any untestable assumption, $E(Z_x)$ and $ACE(X \to Z|X = x, X = x')$ are unidentifiable. Unlike $E(Z_x)$, $E(Z|x)$ is the average of outcomes of individuals only in the treatment group of $X = x$, which is identifiable by the mean of observed outcomes Z in the treatment group.

Theorem 1. If $FACE(X \to Z|X = x, X = x') \geq 0$ for all $x > x'$, then the correlation between X and Z is non-negative (i.e., $r(X, Z) \geq 0$). Further if $FACE(X \to Z|X = x, X = x') > 0$ with a positive probability for some $x > x'$, then the correlation between X and Z is positive (i.e., $r(X, Z) > 0$). The converse also holds when X is a binary variable.

$FACE(X \to Z|X = x, X = x')$ for all $x > x'$ are pairwise measurements of association, and a correlation is a single summarized measurement of association. Thus all FACEs together describe a stricter association than a correlation.

Next we introduce a distributional causal effect (DCE) and a prima facie distributional causal effect (FDCE).

Definition 2. (DCE and FDCE) For $x > x'$, a distributional causal effect of X on Z for a specific threshold z is defined as the difference of cumulative distributions of the potential outcomes

$$DCE[X \to (Z > z)|X = x, X = x'] = P(Z_x > z) - P(Z_{x'} > z),$$

and a prima facie distributional causal effect of X on Z is defined as

$$FDCE[X \to (Z > z)|X = x, X = x'] = P(Z > z|x) - P(Z > z|x').$$

For simplicity, when a conditional variable is binary, the condition part in the above notation is omitted. For example, $ACE(T \to Y|T = 1, T = 0)$ is simplified as $ACE(T \to Y)$.

Since T is randomized, we have $T \perp\!\!\!\perp Y_t$ and get that $P(Y_t > y) = P(Y_t > y|T = t) = P(Y > y|T = t)$ and $E(Y_t) = E(Y_t|T = t) = E(Y|T = t)$. Thus we obtain that $ACE(T \to Y) = FACE(T \to Y)$ and $DCE[T \to (Y > y)] = FDCE[T \to (Y > y)]$. Similarly we have that $ACE(T \to S) = FACE(T \to S)$ and $DCE[T \to (S > s)] = FDCE[T \to (S > s)]$. Generally we neither have $ACE(S \to Y) = FACE(S \to Y)$ nor $DCE[S \to (Y > y)] = FDCE[S \to (Y > y)]$ since S is not randomized and there may be some unobserved confounder which affects both S and Y.

We say that X has a non-negative (non-positive or null) DCE on Z if for any two levels $x > x'$, $DCE[X \to (Z > z)|X = x, X = x'] \geq (\leq \text{or} =)\ 0$ for all z. If, in addition, there exist a threshold z and a pair of levels $x > x'$ such that $DCE[X \to (Z > z)|X = x, X = x'] > (<)0$, then we say that X has a positive (negative) DCE on Z. Similarly we define a positive FDCE.

Notice that DCE depends on a specific threshold z but the sign (negative, positive or null) of DCE does not. A positive DCE and a non-negative DCE imply a positive ACE and a non-negative ACE respectively. However, the converse is not true. Thus DCE describes a stricter causal measurement than ACE.

As shown in the following example, however, Prentice's criterion cannot ensure that a positive DCE of T on S implies a positive DCE of T on Y even if ACE and DCE of S on Y are positive.

Example 2. Consider a randomized trial of treatment T. Generally, however, a surrogate S cannot be randomized in the trial, and thus there may be an unobserved confounder U which affects both the surrogate S and the true endpoint Y and then confounds the causal effect of S on Y. Consider a joint distribution $P(T = t, S = s, Y = y, U = u)$ given in Table 2. From the marginal distribution of T, S and Y on the bottom line 'Sum', we can see that Prentice's criterion $Y \perp\!\!\!\perp T|S$ is satisfied. Since T is randomized, we can easily obtain $ACE(T \to S) = 0.14 > 0$ and $ACE(T \to Y) = -0.004 < 0$ directly from the corresponding conditional distributions. $ACE(S \to Y)$ can be obtained by the following standardization

$$ACE(S \to Y) = \sum_{u=0}^{1} P(U = u)[P(Y = 1|S = 1, u) - P(Y = 1|S = 0, u)] = 0.04 > 0,$$

which also is the DCE for a binary Y. Thus a phenomenon of the surrogate paradox occurs. That is, the treatment T has a positive effect on the surrogate S, which in turn has a positive effect on the true endpoint Y, but the treatment T has a negative effect on the endpoint Y. This example also explains that we cannot simply depict Prentice's criterion as a causal network $T \to S \to Y$ since it may lead to a phenomenon of the surrogate paradox.

Table 2: A phenomenon of the surrogate paradox under Prentice's criterion

	$S = 1$				$S = 0$			
	$Y = 1$		$Y = 0$		$Y = 1$		$Y = 0$	
	$T = 1$	$T = 0$	$T = 1$	$T = 0$	$T = 1$	$T = 0$	$T = 1$	$T = 0$
$U = 1$	0.012	0.006	0.048	0.024	0.016	0.028	0.024	0.042
$U = 0$	0.040	0.020	0.040	0.020	0.128	0.144	0.192	0.216
Sum	0.052	0.026	0.088	0.044	0.144	0.172	0.216	0.258

In Section 3, we shall discuss additional conditions besides Prentice's criterion for qualitatively assessing the sign of treatment effect on an unobserved endpoint in terms of an observed surrogate without assumptions on models. In Section 4, we shall show that these additional conditions are satisfied by commonly used generalized linear models and Cox's proportional hazard models.

3. Criteria for surrogates without model assumptions

In this section, without model assumptions, we discuss what conditions are required to determine the sign of treatment effect on an unobserved endpoint Y by the sign of treatment effect on an observed surrogate S. Let $f(s,t)$ denote the expectation $E(Y|S=s,T=t)$ of Y conditional on $S=s$ and $T=t$. First we discuss the case without requiring Prentice's criterion.

Theorem 2. Suppose that

1. $f(s,1)$ or $f(s,0)$ monotonically increases in s (i.e., $f(s',1) \geq f(s'',1)$ for all $s' > s''$ or $f(s',0) \geq f(s'',0)$ for all $s' > s''$), and
2. $f(s,1) \geq f(s,0)$ for all s.

Then T has a non-negative ACE on Y if T has a non-negative DCE on S.

The supposition 1 requires that the monotonicity property of the expectation of Y in s holds for only one of treatment groups, and the supposition 2 describes the monotonicity property in t. If the true endpoint is observed, the suppositions can be checked no matter whether there is an unobserved confounder which affects both S and Y. Otherwise, we have to judge them based on prior or domain knowledge as follows. In an application, we can consider that the suppositions hold if S and T are risk factors or preventive factors to an endpoint Y. For example, smoking $T=1$ increases averagely the amount S of tar deposited in a person's lung (i.e., supposition 2 holds), and a larger amount of tar also increases averagely the probability of lung cancer for a smoking (but maybe not for nonsmoking) population (i.e., supposition 1 holds).

Theorem 2 means that the ACE sign of a treatment T on an unobserved endpoint Y can be determined by the DCE sign of T on an observed surrogate S if the suppositions are

satisfied. Since a surrogate and a treatment can be redefined by $S^* = -S$ and $T^* = 1 - T$, we can apply the theorem to determine the ACE sign of a treatment on an endpoint through the redefinition of the treatment or/and the surrogate. When the supposition 1 does not hold, we can try to define $S^* = -S$. When the supposition 2 does not hold, we can try to define $T^* = 1 - T$. Then we check whether the suppositions hold for S^* or/and T^*. For example, if the supposition 1 can be satisfied by defining $S^* = -S$ and the supposition 2 holds for T, then the theorem is also true for S^* and T.

Notice that Theorem 2 shows an implication relation from a DCE sign of T on S to an ACE sign of T on Y. The following example explains why a DCE sign of T on S instead of an ACE sign is necessary.

Example 3. Suppose that $P(S = 0|T = 1) = 0.1$, $P(S = 1|T = 1) = 0.8$, $P(S = 2|T = 1) = 0.1$, $P(S = 0|T = 0) = 0.4$, $P(S = 1|T = 0) = 0.3$, $P(S = 2|T = 0) = 0.3$ and $f(s,t) = s^3$. Then the suppositions 1 and 2 are satisfied. Instead of a non-negative DCE of T on S, we have $ACE(T \to S) = 0.1 > 0$, but $ACE(T \to Y) = E[f(S,T)|T = 1] - E[f(S,T)|T = 0] = (0.8 + 0.1 \times 8) - (0.3 + 0.3 \times 8) = -1.1 < 0$.

For Theorem 2, an expectation of Y may depend on a treatment $T = t$ conditionally on $S = s$, that is, $f(s,1) \neq f(s,0)$. This means that there exists a direct effect from T to Y, and thus Prentice's criterion is not satisfied.

Next we discuss the condition for determining the sign of ACE of T on Y in terms of a surrogate which satisfies Prentice's criterion or a weaker condition of the conditional expectation independence (i.e., $E(Y|T = t, S = s) = E(Y|S = s)$, denoted as $Y \perp\!\!\!\perp_E T|S$).

Theorem 3. Suppose that

1. $Y \perp\!\!\!\perp T|S$ or weakly $Y \perp\!\!\!\perp_E T|S$, and

2. $FACE(S \to Y|S = s', S = s'') \geq 0$ for all $s' > s''$.

Then we have $ACE(T \to Y) \geq 0$ if $DCE[T \to (S > s)] \geq 0$ for all s.

By Theorem 1, the supposition 2 implies a non-negative correlation between S and Y. Theorem 3 means that under Prentice's criterion or the conditional expectation independence, the ACE sign of T on Y can be determined by the DCE sign of T on S. When S is binary,

the supposition 2 always holds for S or $S^* = -S$. For example, if the supposition 2 can be satisfied by defining $S^* = -S$ and $DCE[T \to (S^* > s)] \geq 0$ for all s, then we can also determine $ACE(T \to Y) \geq 0$. When the condition $DCE[T \to (S > s)] \geq 0$ in the theorem does not hold, we can try to define $T^* = 1 - T$ and check whether the condition holds for T^*. If it holds for T^*, Theorem 3 is also true for T^*.

If we strengthen the suppositions to that $f(s,t)$ is strictly monotonic in s and $FACE(S \to Y|S = s', S = s'') > 0$ for all $s' > s''$ in Theorems 2 and 3 respectively, then we can get that a positive DCE of T on S implies a positive $ACE(T \to Y)$.

Below we give an example to illustrate that the supposition 2 of positive FACEs in Theorem 3 cannot be relaxed to a positive correlation between S and Y.

Example 4. Consider the distribution given in Table 3, which satisfies Prentice's criterion. We have a quite strong positive correlation between S and Y ($r(S,Y) = 0.706$), but the supposition 2 of Theorem 3 does not hold ($FACE(S \to Y|S = 2, S = 1) = -0.05$). Although $DCE[T \to (S > 1)] = 0.1 > 0$ and $DCE[T \to (S > 0)] = 0$, we have $ACE(T \to Y) = -0.005 < 0$. For correlations, we have $r(T,S) = 0.057$ but $r(T,Y) = -0.005$.

Table 3: The joint distribution of T, S and Y satisfying Prentice's criterion

	$S=0$		$S=1$		$S=2$	
	$T=0$	$T=1$	$T=0$	$T=1$	$T=0$	$T=1$
$Y=0$	0.216	0.324	0.008	0.006	0.012	0.027
$Y=1$	0.024	0.036	0.072	0.054	0.068	0.153

Corresponding to Theorems 2 and 3 about ACE on endpoints, we give below Theorems 4 and 5 about DCE on endpoints. Define $g_y(s,t) = P(Y > y|S = s, T = t)$.

Theorem 4. Suppose that

1. For each y, $g_y(s,1)$ or $g_y(s,0)$ monotonically increases in s (i.e., for each y, either $[g_y(s',1) \geq g_y(s'',1)$ for all $s' > s'']$ or $[g_y(s',0) \geq g_y(s'',0)$ for all $s' > s'']$), and

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

Then T has a non-negative DCE on Y if T has a non-negative DCE on S.

The example of smoking, tar and lung cancer given after Theorem 2 can also be used to interpret the suppositions in Theorem 4 at a distribution level instead of an expectation level. The smoking and the tar are both risk factors to lung cancer at a probability distribution level. Under Prentice's criterion of conditional independence, the above result can be simplified as follows.

Theorem 5. Suppose that

1. Prentice criterion is satisfied, i.e., $Y \perp\!\!\!\perp T|S$, and
2. $FDCE[S \to (Y > y)|S = s', S = s''] \geq 0$ for all $s' > s''$ and all y.

Then we have $DCE[T \to (Y > y)] \geq 0$ for all y if $DCE[T \to (S > s)] \geq 0$ for all s.

Theorems 4 and 5 mean that under the monotonicity property of $g_y()$ or Prentice's criterion, a DCE sign of T on an unobserved endpoint Y can be determined by a DCE sign of T on an observed surrogate S. Similarly, we can apply Theorems 4 and 5 to determine the DCE sign of a treatment on an endpoint Y by redefining a treatment and/or a surrogate. If we strengthen the suppositions in Theorems 4 and 5 to that $g_y()$ is strictly monotonic in s for some y and that S has a positive FDCE on Y, respectively, then a positive DCE T on S implies a positive DCE of T on Y.

In Theorems 3 and 5, we gave under some suppositions that the sign of DCE of T on S is sufficient to determine the signs of ACE and DCE of T on Y, but it is not necessary. That is, there may be a non-negative ACE or DCE of T on Y but $DCE[T \to (S > s)] < 0$ for some s. Below we shall show that it becomes to be necessary and thus there is an equivalence relation between the treatment-effect signs if we further suppose that S is from the one-parameter exponential family conditional on T.

Definition 3. We say that X is from the one-parameter exponential family if its density function has the form $p(x;\theta) = C(\theta)\exp\{Q(\theta)\cdot x\}h(x)$, where Q is strictly monotonic.

Poisson, binomial and inverse binomial distributions and the normal distribution with an unknown constant variance belong to the one-parameter exponential family. First for ACE of T on Y, we give the following stricter result than Theorem 3.

Theorem 6. Suppose that

1. $Y \perp\!\!\!\perp T|S$ or weakly $Y \perp\!\!\!\perp_E T|S$,

2. $FACE(S \to Y|S = s', S = s'') > 0$ for all $s' > s''$, and

3. S is from the one-parameter exponential family conditional on T.

Then $ACE(T \to S)$, $DCE(T \to S)$ and $ACE(T \to Y)$ have the same sign (null, positive, or negative).

Next for DCE of T on Y, we give the following stricter result than Theorem 5.

Theorem 7. Suppose that

1. $Y \perp\!\!\!\perp T|S$,

2. S has a positive FDCE on Y, and

3. S is from the one-parameter exponential family conditional on T.

Then $ACE(T \to S)$, $DCE(T \to S)$ and $DCE(T \to Y)$ have the same sign (null, positive, or negative).

According to Theorems 6 and 7, Prentice's criterion and additional suppositions can ensure equivalence relations between the signs of treatment effects on surrogates and on endpoints. Without these additional suppositions, only Prentice's criterion cannot ensure these equivalence relations unless S is binary. Especially, as shown by Buyse and Molenberghs (1998), Prentice's criterion ensures that an independence of T and S implies an independence of T and Y, but a dependence of T and S does not imply a dependence of T and Y. Thus, rejecting a null hypothesis $T \perp\!\!\!\perp S$ does not mean rejecting a null hypothesis $T \perp\!\!\!\perp Y$.

In Theorems 6 and 7, we also obtain from the supposition of the one-parameter exponential family that ACE and DCE of T on S have the same sign. When S has a normal distribution conditional on T, it is from the one-parameter exponential family if its variance is the same for all $T = t$.

4. Criteria for surrogates with model assumptions

In this section, we first consider generalized linear models (McCullagh and Nelder, 1989) and Cox's proportional hazard models (Cox, 1972), and we discuss the condition for qual-

itatively assessing the causal effect on an unobserved endpoint with a surrogate. Then we extend the results to hybrid models of them. With model assumptions, we can obtain more informative results on strict signs of ACE and DCE on a true endpoint.

4.1 Generalized linear models

For generalized linear models and variables from the one-parameter exponential families, we show that a surrogate satisfying Prentice's criterion $Y \perp\!\!\!\perp T|S$ (or weakly $Y \perp\!\!\!\perp_E T|S$) can be used to qualitatively assess or to predict the signs of ACE and DCE on an unobserved true endpoint Y.

We consider the following extension of generalized linear models:

$$\begin{aligned} h_1[E(Y|S=s,T=t)] &= a_1(s)+b_1(t)+c_1, \\ h_2[E(S|T=t)] &= a_2(t)+c_2, \end{aligned}$$

where $h_1()$ and $h_2()$ are strictly monotonic increasing link functions. The following corollaries show the conditions for the equivalence relations between the ACE (DCE) signs on S and on Y.

Corollary 1. Suppose that

1. $Y \perp\!\!\!\perp T|S$ or weakly $Y \perp\!\!\!\perp_E T|S$ (i.e., $b_1(t)=0$),
2. $a_1(s)$ strictly monotonically increases in s, and
3. S is from the one-parameter exponential family conditional on T.

Then the following statements are equivalent:

1. $a_2(t)$ is strictly monotonically increasing (decreasing, zero) in t,
2. T has a positive (negative, null) ACE on S, and
3. T has a positive (negative, null) ACE on Y.

Corollary 2. Suppose that

1. $Y \perp\!\!\!\perp T|S$,

2. $a_1(s)$ strictly monotonically increases in s, and

3. S given T and Y given S are from the one-parameter exponential families.

Then the following statements are equivalent:

1. $a_2(t)$ is strictly monotonically increasing (decreasing, zero) in t,

2. T has a positive (negative, null) DCE on S, and

3. T has a positive (negative, null) DCE on Y.

Ordinarily, for a generalized linear model, the functions $a_1(s)$ and $a_2(t)$ are defined as $a_1 s$ and $a_2 t$ respectively, and thus the supposition and condition of monotonicity always hold for T (or $T^* = 1 - T$) and S (or $S^* = -S$). Thus for the generalized linear model and the one-parameter exponential families, a surrogate satisfying Prentice's criterion can be used to qualitatively assess ACE and DCE on an unobserved true endpoint. The monotonicity of $a_2(t)$ and $a_1(s)$ means that increasing t increases the expectation of S and that increasing s increases the expectation of Y respectively, and thus the causal effects DCE and ACE of T on Y are positive without phenomena of the surrogate paradox occurring. If $a_1(s)$ monotonically decreases, then Corollaries 1 and 2 are also true by defining $S^* = -S$.

4.2 Proportional hazard model

In this subsection, we consider the following proportional hazard model under Prentice's criterion

$$\begin{aligned} \lambda(y|S = s, T = t) &= \lambda(y|S = s) = \lambda_0(y)[a_1(s) + c_1], \\ \lambda^*(s|T = t) &= \lambda_0^*(s)[a_2(t) + c_2], \end{aligned}$$

where $\lambda_0()$ and $\lambda_0^*()$ are baseline hazards. For the model, we have the following equivalence relation between the signs of DCEs on S and on Y.

Corollary 3. Suppose that $a_1(s)$ strictly monotonically decreases in s. Then the following statements are equivalent:

1. $a_2(t)$ is strictly monotonically decreasing (increasing, zero) in t,

2. T has a positive (negative, null) DCE on S, and

3. T has a positive (negative, null) DCE on Y.

4.3 Hybrid models

The results discussed in the previous subsections on the generalized linear model and the proportional hazard model can be extended to a hybrid model for which one equation of S or Y is a generalized linear model and the other is a proportional model, such as

$$\begin{aligned} h_1[E(Y|S=s,T=t)] &= h_1[E(Y|S=s)] = a_1(s)+c_1, \\ \lambda^*(s|T=t) &= \lambda_0^*(s)[a_2(t)+c_2], \end{aligned}$$

or

$$\begin{aligned} \lambda(y|S=s,T=t) &= \lambda(y|S=s) = \lambda_0(y)[a_1(s)+c_1], \\ h_2[E(S|T=t)] &= a_2(t)+c_2. \end{aligned}$$

Similar to the previous corollaries, a surrogate satisfying Prentice's criterion can also be used to qualitatively assess the ACE and DCE on an unobserved endpoint.

5. Application to frequency of breast screening

In this section, we illustrate our conditions for surrogates using the data on breast screening frequency, breast cancer mortality and tumour size analyzed by Day and Duffy (1996). To compare the screening effects of a frequency of 3 years with a frequency of 1 year on breast cancer survival, Day and Duffy (1996) checked the validation of Prentice's criterion and considered that tumour size is the most important surrogate for 10-year death. Let $T=0$ denote a screening frequency of 1 year and $T=1$ a frequency of 3 years, S denote the tumour size at diagnosis, and Y denote the true endpoint of cancer death ($Y=1$) or not ($Y=0$). In Table 4, the first four columns give the data of their Table 2. From the fourth column, we know that the supposition 2 of Theorem 3 is satisfied, that is, $FACE(S \to Y|s',s'') > 0$ for all $s' > s''$. From the second and third columns, we have that $DCE[T \to (S > s)] > 0$ for all s. Thus we can determine that ACE and DCE of T on Y are positive, which can be

confirmed by the $ACE(T \to Y) = DCE[T \to (Y > 0)] = 0.20 - 0.13 > 0$ from the last line of Table 4. Suppose that we have the prior knowledge that the larger tumour at diagnosis is more dangerous. Then we can use the tumour size as a surrogate for the cancer death to qualitatively assessing the effect of different screening frequencies on survival so that the costs and the time duration for the study can be reduced greatly.

Now suppose that Prentice's criterion does not hold, and our artificial distribution $P(Y = 1|s,t)$ is given in the last two columns of Table 4. By the distribution, we have that the supposition 1 of Theorem 4 holds, which means that the larger tumour is more dangerous. Also the supposition 2 of Theorem 4 holds, which means that more frequent screening may have other positive effect on survival except the pathway through the tumour size. If we have the prior knowledge, then, by the observed positive treatment effect on the turmour size, we can determine that the arm of more frequent screening has a positive DCE on survival.

Table 4: Distributions of diagnosed tumour sizes and 10-year death-rates*

Size (mm)	Distributions of tumour size $P(s\|t)$		10-year death-rate	$P(Y=1\|s,t)$	
S	3 years ($T=1$)	1 year ($T=0$)	$p(Y=1\|s)$	$T=1$	$T=0$
$1-9$	0.20	0.43	0.02	0.030	0.015
$10-14$	0.25	0.21	0.05	0.060	0.038
$15-19$	0.21	0.14	0.15	0.200	0.075
$20-29$	0.20	0.13	0.35	0.400	0.273
$30-49$	0.09	0.07	0.50	0.550	0.436
$50+$	0.05	0.02	0.70	0.750	0.575
$P(Y=1\|t)$	0.20	0.13		0.230	0.103

* The first four columns are from Table 2 of Day and Duffy (1996).

6. Discussions

In this paper, we showed that Prentice's criterion cannot ensure that the sign of treatment effect on an unobserved true endpoint can be predicted by the sign of treatment effect on an observed surrogate although it can ensure the validation of tests on null hypothesis of no treatment effect on an endpoint. We presented some conditions and model assumptions

to predict or determine the sign of treatment effect on an endpoint with or without requiring Prentice's criterion. These conditions and models can be tested if the true endpoint is observed in previous validation trials. We showed in Theorems 3 and 5 that in addition to Prentice's criterion, non-negative association measurements FACE and FDCE rather than a correlation between a surrogate and an endpoint are required for the qualitative assessment. We remove the assumption of extrapolation that estimates from the previous validation trials are applicable to future trials. The cases with a direct effect of a treatment on a true endpoint were also discussed in Theorems 2 and 4 where Prentice's criterion is not satisfied.

Suppose that a surrogate and an endpoint have a generalized linear model or a proportional hazard model. Then the surrogate satisfying Prentice's criterion can be used not only to ensure the validation of null hypothesis of no treatment effect on an endpoint, but also to predict or determine the sign of treatment effect on the endpoint.

The criteria proposed by Chen et al. (2007) and Ju and Geng (2009) are based on a causal network presented by Laurizten (2004), which do not require Prentice's criterion. Their conditions for surrogates contain unobserved confounders between a surrogate and an endpoint, and thus they are untestable even if the endpoint is observed. These criteria require prior knowledge or assumptions on monotonicity of causal effects of a surrogate on an endpoint. Comparing with these criteria, we can see that the conditions proposed in this paper can be tested if there are validation trials in which an endpoint is observed. If the endpoint is never observed, to predict or determine the sign of treatment effect on the endpoint, we need prior knowledge or assumptions on monotonicity of associations between the surrogate and the endpoint.

The results on surrogates proposed in this paper can also be applied to an intermediate variable analysis and instrumental variable methods. These results qualitatively describe the implication and equivalence relations of causal effects among variables in a causal pathway.

Acknowledgements

This research was supported by NSFC (10771007, 10721403, 10931002, 30973976) and MOE-Microsoft Key Laboratory of Statistics and Information Technology of Peking Univer-

sity.

Appendix: Proofs of theorems and corollaries

Proof of Theorem 1. Let $F(x) = P(X \leq x)$ denote the distributional function of X. Then we have

$$\begin{aligned} Cov(X,Z) &= E(XZ) - E(X)E(Z) = E\{X[E(Z|X) - E(Z)]\} \\ &= E\{X[E(Z|X) - E(E(Z|X'))]\} = E\{E[XFACE(X \to Z|X, X')]\}. \end{aligned}$$

By definition of expectation and exchange of x and x', the above equation can be written as

$$\begin{aligned} & \int_{-\infty}^{+\infty}\int_{-\infty}^{x} xFACE(X \to Z|x,x')dF(x')dF(x) + \int_{-\infty}^{+\infty}\int_{x}^{+\infty} xFACE(X \to Z|x,x')dF(x')dF(x) \\ = & \int_{-\infty}^{+\infty}\int_{-\infty}^{x} xFACE(X \to Z|x,x')dF(x')dF(x) + \int_{-\infty}^{+\infty}\int_{x'}^{+\infty} x'FACE(X \to Z|x',x)dF(x)dF(x') \\ = & \int_{-\infty}^{+\infty}\int_{-\infty}^{x} xFACE(X \to Z|x,x')dF(x')dF(x) + \int_{-\infty}^{+\infty}\int_{-\infty}^{x} x'FACE(X \to Z|x',x)dF(x')dF(x) \\ = & \int_{-\infty}^{+\infty}\int_{-\infty}^{x} (x-x')FACE(X \to Z|x,x')dF(x')dF(x). \end{aligned}$$

Thus we can get $Cov(X,Z)$ non-negative (positive) from the conditions. When X is a binary variable with values $x_1 < x_2$, then the above equation can be written as

$$Cov(X,Z) = (x_2 - x_1)P(X = x_1)P(X = x_2)FACE(X \to Z|x_2, x_1).$$

Thus we can find that the converse of the theorem is also true. □

Before proving other results, we give four lemmas to be used in the proofs.

Lemma 1. Suppose that X is stochastically larger than Y (i.e.,$P(X > c) \geq P(Y > c)$ for all c). Let f be an increasing function, then $E[f(X)] \geq E[f(Y)]$. Further if X is strictly stochastically larger than Y (i.e., $P(X > c) > P(Y > c)$ for some c) and f is strictly increasing, then $E[f(X)] > E[f(Y)]$.

Proof. From the equation $EX = \int_0^{+\infty} P(X > x)dx - \int_{-\infty}^{0}[1 - P(X > x)]dx$ (Chung, 1974, page 49), we can get $E[f(X)] - E[f(Y)] = \int_{-\infty}^{+\infty}[P(f(X) > c) - P(f(Y) > c)]dc$. Then we can get the result from the monotonicity of f and the right continuity of cumulative distribution functions easily. □

Lemma 2. Suppose that X_1 and X_2 are from the one-parameter exponential families with the density functions in Definition 3. Let $p = \inf_{x\in R}\{p(x;\theta) > 0\}$ and $q = \sup_{x\in R}\{p(x;\theta) > 0\}$. If $E(X_1) > E(X_2)$, then X_1 is stochastically larger than X_2, and further we have $P(X_1 > c) > P(X_2 > c)$ for all $c \in (p, q)$.

Proof. The one-parameter exponential families exhibit monotonic increasing likelihood ratio in θ. From Lemma 3.4.2 (ii) of Lehmann and Romano (2005, Page 70), for any $\theta > \theta'$, we have $P_\theta(X > x) \geq P_{\theta'}(X > x)$ for all x. In addition, from Theorem 3.4.1 (ii) of Lehmann and Romano (2005, Page 66), we have that for $c \in (p, q)$, $P_\theta(X > c)$ strictly increases in θ. Thus from Lemma 1, $E(X)$ also strictly increases in θ. If $E(X_1) > E(X_2)$, we obtain $\theta_{X_1} > \theta_{X_2}$, and hence the result follows. □

Lemma 3. Suppose that T is randomized. Let $f(s,t) = E(Y|S = s, T = t)$. Then

$$ACE(T \to Y|T = 1, T = 0) = E[f(S_{T=1}, 1)] - E[f(S_{T=0}, 0)]$$

Proof. Since $T \perp\!\!\!\perp (S_{T=t}, Y_{T=t,S=s})$, we have

$$E(Y_{T=t,S=s} = y|S_{T=t} = s) = E(Y_{T=t,S=s} = y|S_{T=t} = s, T = t) = E(Y = y|S = s, T = t).$$

We use $F_t(s) = P(S_{T=t} \leq s)$ to denote the distributional function of $S_{T=t}$. Then we can get

$$\begin{aligned} E(Y_{T=t}) &= E(Y_{T=t,S=S_{T=t}}) = \int_{-\infty}^{+\infty} E(Y_{T=t,S=s}|S_{T=t} = s)dF_t(s) \\ &= \int_{-\infty}^{+\infty} E(Y|S = s, T = t)dF_t(s) = \int_{-\infty}^{+\infty} f(s,t)dF_t(s) = E[f(S_{T=t}, t)]. \end{aligned}$$

Thus $ACE(T \to Y|T = 1, T = 0) = E[f(S_{T=1}, 1)] - E[f(S_{T=0}, 0)]$. □

Lemma 4. Let $g_y(s,t) = P(Y > y|S = s, T = t)$. For randomized T, we have

$$DCE(T \to (Y > y)|T = 1, T = 0) = E[g_y(S_{T=1}, 1)] - E[g_y(S_{T=0}, 0)].$$

Proof. □

Proof of Theorem 2. Since T has a non-negative DCE on S, we can get that $S_{T=1}$ is stochastically larger than $S_{T=0}$. Then from the supposition and Lemma 1, we can get

$$E[f(S_{T=1}, 1)] \geq E[f(S_{T=0}, 1)](or E[f(S_{T=1}, 0)]) \geq E[f(S_{T=0}, 0)].$$

Then from Lemma 3, we have $ACE(T \to Y|T = 1, T = 0) \geq 0$. □

Proof of Theorem 3. Under the conditional expectation independence, we denote $f(s) = f(s,1) = f(s,0)$. Then we can get $f(s)$ increasing from Supposition 2. Since T has a non-negative DCE on S, we have that $S_{T=1}$ is stochastically larger than $S_{T=0}$. Then from Lemmas 1 and 3, we can get

$$ACE(T \to Y|T=1,T=0) = E[f(S_{T=1})] - E[f(S_{T=0})] \geq 0. \qquad \square$$

Proofs of Theorems 4 and 5. Similar to the proofs of Theorems 2 and 3, we can get the results from Lemmas 1 and 4. $\square$

In the following proofs, we only show the positive part, and the negative and null parts can be shown similarly.

Proof of Theorem 6.

1. First, we show that $ACE(T \to S) > 0$ implies $DCE(T \to S) > 0$. $ACE(T \to S) > 0$ means $E(S_{T=1}) > E(S_{T=0})$. Thus by Lemma 2 and supposition 3, we have $DCE(T \to S) > 0$.

2. Next, we obtain from Theorem 3 and suppositions 1 and 2 that $DCE(T \to S) > 0$ implies $ACE(T \to Y) > 0$.

3. Finally, we show that $ACE(T \to Y) > 0$ implies $ACE(T \to S) > 0$. If $ACE(T \to S) \leq 0$, we can get $DCE(T \to S) \leq 0$ from Lemma 2 and supposition 3. Then we have $ACE(T \to Y) \leq 0$ from Theorem 3 with $T^* = 1 - T$. Thus $ACE(T \to Y) > 0$ implies $ACE(T \to S) > 0$. $\square$

Proof of Theorem 7. Similar to the proof of Theorem 6, we can use Lemma 2 and Theorem 5 instead of Theorem 3 to prove this theorem. $\square$

Proof of Corollary 1. First, since $h_2()$ is strictly monotonically increasing, we can get that statements 1 and 2 are equivalent. Next, we have that $FACE(S \to Y|S=s', S=s'') > 0$ for all $s' > s''$ since $E(Y|S=s) = h_1^{-1}(a_1(s) + c_1)$ strictly monotonically increases in s. Thus by Theorem 6, statements 2 and 3 are equivalent. $\square$

Proof of Corollary 2. Since the conditional distribution of Y given S is from the one-parameter exponential family, we have $FDCE[S \to (Y > y)|S=s', S=s''] > 0$ for all

$s' > s''$ and all y from Lemma 2 and supposition 2. Then we can get the result from Theorem 7 similar to the proof of Corollary 1. □

Proof of Corollary 3.

1. First, we show that $a_2(1) < a_2(0)$ implies a positive DCE of T on S. From the definition of the hazard function λ, we have $P(S > s|T = t) = exp\{-\int_0^s \lambda(u|T = t)du\} = [exp\{-\int_0^s \lambda_0(u)du\}]^{a_2(t)+c_2}$. Since $exp\{-\int_0^s \lambda_0(u)du\}$ is the baseline survival function and $a_2(1) < a_2(0)$, we can get a positive DCE of T on S.

2. Next, we have $FDCE[S \to (Y > y)|S = s', S = s''] \geq 0$ for all $s' > s''$ and all y from a strictly monotonically decreasing $a_1(s)$ similarly. Thus by Theorem 5, a positive DCE of T on S implies a positive DCE of T on Y.

3. Finally, we show that a positive DCE of T on Y implies $a_2(1) < a_2(0)$. If $a_2(1) \geq a_2(0)$, we can get a non-positive DCE of T on S similar to part 1 of this proof. Then we have a non-positive DCE of T on Y from Theorem 5 with $T^* = 1 - T$. Thus a positive DCE of T on Y implies $a_2(1) < a_2(0)$. □

References

Alonso, A. and Molenbergh, G. (2008) Surrogate end points: hopes and perils. *Expert Review of Pharmacoeconomics and Outcomes Research,* **8**, 255-259.

Baker, S. G. (2006) Surrogate endpoints: wishful thinking or reality? *J. Nat. Canc. Inst.,* **98**, 502-503.

Berger, V. W. (2004) Does the Prentice criterion validate surrogate endpoints? *Statist. Medicine,* **23**, 1571-1578.

Burzykowski, T. Molenberghs, G. and Buyse, M. (2005) *The Evaluation of Surrogate End-points.* New York: Springer.

Buyse, M. and Molenberghs, G. (1998) Criteria for the validation of surrogate end-points in randomized experiments. *Biometrics*, **54**, 1014-1029.

Chen, H., Geng, Z. and Jia, J. (2007) Criteria for surrogate endpoints. *J. R. Statist. Soc.* **B, 69**, 919-932.

Chung, K. L. (1974) *A course in probability theory.* 2nd edn. New York: Academic Press.

Cox, D. R. (1972) Regression models and life tables. *J. R. Statist. Soc.*, **B, 34**, 187-220.

Day, N. E. and Duffy, S. W. (1996) Trial design based on surrogate end pints - application to comparison of different breast screening frequencies. *J. R. Statist. Soc.* **A, 159**, 49-60.

Fleming, T. R. and DeMets, D. L. (1996) Surrogate end points in clinical trials: are we being misled? *Ann. Intern. Med.*, **125**, 605-613.

Frangakis, C. E. and Rubin, D. B. (2002) Principal Stratification in Causal Inference. *Biometrics*, **58**, 21-29.

Holland, P. W. (1986) Statistics and causal inference. *J. Amer. Statist. Asso.*, **81**, 945-960.

Ju, C. and Geng, Z. (2009) Criteria for surrogate endpoints based on causal distributions. To appear in *J. R. Statist. Soc.*, **B**.

Lauritzen, S. L. (2004) Discussion on causality. *Scand. J. Statist.*, **31**, 189-192.

Lehmann, E. L. and Romano, J. P. (2005) *Testing Statistical Hypotheses.* 3rd edn. New York: Springer.

Manns, B., Owen, W. F., Winkelmayer, W. C., Devereaux, P. J. and Tonelli, M. (2006) Surrogate markers in clinical studies: problems solved or created? *Am. J. Kidney Diseases*, 48, 159-166.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models,* 2nd edn. London: Chapman and Hall.

Prentice, R. L. (1989) Surrogate endpoints in clinical trials: definition and operational criteria. *Statist. Med.*, 8, 431-440.

Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, **66**, 688-701.

Rubin, D. B. (2004) Direct and indirect causal effects via potential outcomes. *Scand. J. Statist.*, **31**, 161-170.

VanderWeele, T. J. and Robins, J. M. (2009) Proerties of monotonic effects on directed acyclic graphs. *J. Mach. Learn. Res.*, **10**, 699-718.