8-1-2015

Surrogate Markers for Time-Varying Treatments and Outcomes

Jesse Y. Hsu
University of Pennsylvania

Edward H. Kennedy

Jason A. Roy

Alisa J. Stephens-Shields
University of Pennsylvania

Dylan S. Small
University of Pennsylvania

See next page for additional authors

Recommended Citation

Hsu, J. Y., Kennedy, E. H., Roy, J. A., Stephens-Shields, A. J., Small, D. S., & Joffe, M. M. (2015). Surrogate Markers for Time-Varying Treatments and Outcomes. *Clinical Trials, 12* (4), 309-316. http://dx.doi.org/10.1177/1740774515583500

Surrogate Markers for Time-Varying Treatments and Outcomes

Abstract
BACKGROUND: A surrogate marker is a variable commonly used in clinical trials to guide treatment decisions when the outcome of ultimate interest is not available. A good surrogate marker is one where the treatment effect on the surrogate is a strong predictor of the effect of treatment on the outcome. We review the situation when there is one treatment delivered at baseline, one surrogate measured at one later time point, and one ultimate outcome of interest and discuss new issues arising when variables are time-varying.

METHODS: Most of the literature on surrogate markers has only considered simple settings with one treatment, one surrogate, and one outcome of interest at a fixed time point. However, more complicated time-varying settings are common in practice. In this article, we describe the unique challenges in two settings, time-varying treatments and time-varying surrogates, while relating the ideas back to the causal-effects and causal-association paradigms.

CONCLUSION: In addition to discussing and extending popular notions of surrogacy to time-varying settings, we give examples illustrating that one can be misled by not taking into account time-varying information about the surrogate or treatment. We hope this article has provided some motivation for future work on estimation and inference in such settings.

Keywords
causal inference, observational studies, randomization, surrogacy, time-varying

Disciplines
Business | Business Analytics | Health Services Research | Statistics and Probability

Author(s)
Jesse Y. Hsu, Edward H. Kennedy, Jason A. Roy, Alisa J. Stephens-Shields, Dylan S. Small, and Marshall M. Joffe

HHS Public Access

Author manuscript

Clin Trials. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

Clin Trials. 2015 August ; 12(4): 309–316. doi:10.1177/1740774515583500.

Surrogate markers for time-varying treatments and outcomes

Jesse Y Hsu[1], **Edward H Kennedy**[1], **Jason A Roy**[1], **Alisa J Stephens-Shields**[1], **Dylan S Small**[2], and **Marshall M Joffe**[1]

[1]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

Abstract

Background—A surrogate marker is a variable commonly used in clinical trials to guide treatment decisions when the outcome of ultimate interest is not available. A good surrogate marker is one where the treatment effect on the surrogate is a strong predictor of the effect of treatment on the outcome. We review the situation when there is one treatment delivered at baseline, one surrogate measured at one later time point and one ultimate outcome of interest, and discuss new issues arising when variables are time-varying.

Methods—Most of the literature on surrogate markers has only considered simple settings with one treatment, one surrogate, and one outcome of interest at a fixed time point. However, more complicated time-varying settings are common in practice. In this paper, we describe the unique challenges in two settings, time-varying treatments and time-varying surrogates, while relating the ideas back to the causal-effects and causal-association paradigms.

Conclusions—In addition to discussing and extending popular notions of surrogacy to time-varying settings, we give examples illustrating that one can be misled by not taking into account time-varying information about the surrogate or treatment. We hope this paper has provided some motivation for future work on estimation and inference in such settings.

Keywords

Causal inference; observational studies; randomization; surrogacy; time-varying

Introduction: surrogate markers

A surrogate marker is a variable used to guide treatment decisions when the outcome of ultimate interest is not available. In clinical trials, commonly used surrogate markers include laboratory values, body weight, and blood pressure. What makes a good surrogate marker is different than what makes a good biomarker. A good biomarker is a variable that is a strong predictor of the outcome of interest. A good surrogate marker is one where the treatment

Corresponding author: Jesse Y Hsu; 618 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA; hsu9@mail.med.upenn.edu; Phone: (215) 573-8980.

effect on the surrogate is a strong predictor of the effect of treatment on the outcome. It is possible that a variable is a good biomarker, but a poor surrogate marker.

There is a large literature on surrogate markers.[1–6] Most of this literature focuses on a surrogate marker measured at one time point and an ultimate outcome of interest measured at one (typically later) time point. More complicated settings with time-varying variables are common in clinical trials; the time varying variables could be treatments, clinical outcomes, or surrogate markers. For motivation, consider studies of an antiretroviral drug (ART), Zidovudine, to reduce mortality for patients with HIV. A patient's CD4 count is a known surrogate marker for mortality.[7] In time-varying settings, another treatment option such as Lamivudine can be added during the course of treatments (multiple treatments), or a repeated measurement of CD4 count can be treated as another surrogate marker (multiple surrogates). The goal of this paper is to discuss new issues that arise when thinking about surrogate markers in the context of time-varying variables.

We first review the situation when there is one treatment, one surrogate and one ultimate outcome of interest in the Review Section. Then, we discuss new issues arising when there are time-varying situations.

Review: one treatment, one surrogate, and one outcome of interest

We review surrogate markers in simple settings, following the arguments of Joffe and Greene.[8] A surrogate is a variable for which knowing the effect of treatment on the surrogate allows prediction of the effect of treatment on the outcome. There are two major frameworks for evaluating surrogates: (1) the causal-effects paradigm, and (2) the causal-association paradigm.

First, to talk about causal inference we need to introduce notation. In what follows, we use both potential outcomes or counterfactual language[9,10] and graphical models language.[11] Let A denote the (randomized) treatment, S the surrogate, and Y the ultimate outcome of interest. Define U as a common cause of S and Y, possibly unmeasured. Let Y^a and S^a be the outcome and surrogate, respectively, that would be seen under treatment level of a. Causal effects are contrasts of potential outcomes for different treatment levels for the same subjects. For example, for a binary treatment, $a = 0, 1$, the effect of treatment for a particular subject would be $Y^1 - Y^0$, the difference between the outcome a subject would get had he/she received the treatment versus what the subject would get had he/she received the control. Similarly we also let $Y^{a,s}$ denote the potential outcome that would have been observed under treatment level a and surrogate level s. Thus the potential outcome Y^a under treatment level a can also be expressed as Y^{a,S^a}.

In the causal-effects paradigm, the quality of a surrogate is typically assessed with "proportion explained" or "proportion mediated" measures, i.e., how much of the effect of treatment is explained or mediated by the surrogate. Thus a variable is considered a good surrogate if knowing the causal effect of a treatment on the surrogate along with the causal effect of the surrogate on the outcome allows good prediction of the causal effect of the treatment on the outcome. We may learn about the effect of the surrogate on the outcome from prior experiments or external information, and conduct a surrogate experiment to

obtain the effect of a new treatment on the surrogate, from which we would make a decision about the effect of new treatment on the outcome.

Freedman et al.[12] proposed to measure the proportion of effect explained using the framework proposed by Prentice,[13] who defined a surrogate as a variable S for which a test of the null hypothesis of no relationship to treatment A is also a valid test of no relationship to the outcome Y. Operationally, Prentice suggested the following criterion: S is a surrogate if it is correlated with the outcome Y and if, once conditioned upon, it renders the treatment A and outcome Y independent. The proportion explained approach for a general endpoint outcome works as follows. Consider two generalized linear models: one that models Y given A directly, $g\{E(Y \mid A)\} = \beta_0 + A\beta_A$, and another that models Y given A and S without interaction between A and S, $g\{E(Y \mid A, S)\} = \gamma_0 + A\gamma_A + S\gamma_S$. Then we can express a proportion of the total effect of A explained by S by $1 - \gamma_A/\beta_A$.[12] Freedman et al. consider S to be a surrogate if the proportion of effect explained is greater than zero. Although this "proportion explained" approach does not utilize explicitly causal ideas, it does seem to implicitly require no unmeasured confounding of the effect of S on Y (Figure 1a), which unfortunately cannot be ensured by randomization of treatment A.

An alternative approach in the causal-effects paradigm is based on more explicitly causal ideas from the mediation literature. The total effect of treatment on the outcome can be pieced together from its direct effect and its indirect effect through the surrogate, which we now define. Informally, Figure 1b shows the graphical presentation for direct and indirect effects. The formal definition involves potential outcome Y^{a,S^a}, which is the outcome that would be observed at treatment level a and surrogate S^a. The total effect of treatment, $Y^{a=1,S^1} - Y^{a=0,S^0}$, can be decomposed as the sum of a natural direct and indirect effect. The subject-level natural direct effect (NDE) of treatment when the surrogate is fixed at its level under treatment level $a=0$ is $Y^{a=1,S^0} - Y^{a=0,S^0}$, a contrast that holds the surrogate constant at the value it would have obtained had treatment been set to zero while changing the treatment from 1 to 0, and the natural indirect effect (NIE) through the surrogate when treatment is fixed at level $a=1$ is $Y^{a=1,S^1} - Y^{a=1,S^0}$, a contrast holding the treatment constant and changing the surrogate from the value it would have obtained under treatment versus control;[14] the sum of this natural direct effect and natural indirect effect is the total effect of treatment, $Y^{a=1,S^1} - Y^{a=0,S^0}$. Another decomposition of the total effect is the natural direct effect when the surrogate is fixed at its level under treatment level $a=1$, $Y^{a=1,S^1} - Y^{a=0,S^0}$, plus the natural indirect effect when the treatment is fixed at level $a=0$, $Y^{a=0,S^1} - Y^{a=0,S^0}$. The average direct and indirect effects can be estimated using, for example, a structural model approach under the assumption that the initial treatment is randomized or ignorable conditional on baseline covariates, and that the surrogate is sequentially ignorable given the initial treatment and baseline covariates.[8,15]

The causal-effects paradigm typically assumes that the surrogate is on the causal pathway from the treatment to the outcome. Understanding the way in which the surrogate fits into the causal process may be helpful for generalizing whether a variable that is a good surrogate in one setting will be a good surrogate in another. However, a drawback of the causal effects paradigm is that many surrogates are not causal intermediates (i.e., not on the causal pathway), but instead are proxies for causal intermediates. Figure 1c depicts a setting

where S^* is the causal intermediate but the proxy S is observed; we call such an S a proxy surrogate. An example of a proxy surrogate is hemoglobin A1C in diabetes; hemoglobin A1C is a proxy for blood sugar and not a causal intermediate -- if there were a way to change the amount of glycosylated hemoglobin (what hemoglobin A1C measures) without changing levels of blood sugar, it would have little or no effect on health outcomes for diabetic patients.

In the causal-association paradigm, evaluation of a surrogate is based on examination of the association, across studies or population subgroups, between the effect of a treatment on the surrogate and the effect of a treatment on the clinical outcome. A good surrogate is a variable for which the effect of a treatment on the surrogate is highly associated with the effect of the treatment on the outcome. One approach in this paradigm is based on meta-analysis.[2] The meta-analytic approach examines the relationship across studies between the effect of the randomized treatment on the surrogate and the effect of the randomized treatment on the clinical outcome. Denote the effect of treatment on surrogate in study j as θ_j and the effect of treatment on outcome in that study as ϕ_j. Ideally, for a good surrogate, we would find (1) there is a monotonic relationship between θ_j and ϕ_j; (2) when θ_j is 0, ϕ_j is also 0; and (3) θ_j should predict ϕ_j well; i.e., in a regression of ϕ_j on θ_j, there should be little variability around the regression line. This approach can be applied not only across studies but also across specific subgroups defined by baseline covariates within a study. An alternative approach in the causal-association paradigm is principal stratification.[16] In principal stratification, we focus on the association of individual-level effects of A on S and on Y; i.e., the association of $S^1 - S^0$ with $Y^1 - Y^0$. A variable S is called a principal surrogate if the effect of treatment on the outcome is 0 in any individual for whom A does not affect S.[16] Because S^1 and S^0 are not simultaneously observable in any individual, the causal effect of treatment on the surrogate is not observable without further assumptions and so assessment of whether S is a principal surrogate requires further assumptions. An advantage of the causal-association paradigm is that it deals naturally with proxy surrogates as well as causal surrogates. In general, the causal-effects and causal-association paradigms and their corresponding approaches can have different advantages in different settings. Which approach is used should depend on the specific study goals and on the nature of the putative surrogate.[6]

Time-varying settings

To this point, nearly all of the literature on surrogate markers has considered relatively simple settings where the treatment, surrogate, and outcome are all scalars measured at one fixed time each. However, more complicated time-varying scenarios are common in practice; in fact, Prentice actually considered an example where the surrogate was measured over time and the outcome was time-to-event.[13] A few authors have considered surrogates for failure time outcomes (e.g., Qin et al.[17] and Gabriel and Gilbert),[18] but to the best of our knowledge no one has addressed the problem of evaluating surrogate markers in settings where the treatment or surrogate are measured repeatedly over time. We describe the unique challenges in these two settings, while relating the ideas back to the two paradigms discussed above.

Time-varying treatments

Before proceeding to time-varying treatments in both causal-effects and causal-association paradigms in next paragraphs, we first introduce notation and the concept of global and local contrasts for time-varying treatments. We consider two sequential treatments, A_1 and A_2 with one surrogate marker S and one clinical outcome Y (Figure 2). In the HIV example, treatments A_1 and A_2 could be Zidovudine and Lamivudine, surrogate S could be patient's CD4 count and clinical outcome Y could be time to death. The effect of treatments on the outcome is contrasted under different treatment plans or regimes; e.g., $A_1=a_1, a_1^*$ and $A_2=a_2, a_2^*$. There are global contrasts and local contrasts. Whether S is a good surrogate may depend on whether we are interested in global vs. local contrasts. A global contrast considers $Y^{a_1,a_2}-Y^{a_1^*,a_2^*}$ where $(a_1, a_2) \neq (a_1^*, a_2^*)$; i.e., we make contrasts over all different treatment plans. On the other hand, a local contrast considers different treatments varying at only one time point in time; e.g., $a_1 \neq a_1^*$ and $a_2=a_2^*$ for the effect of a blip of treatment at time 1 or $a_1=a_1^*$ and $a_2 \neq a_2^*$ for the effect of blip of treatment at time 2. The effect of a blip of treatment at time j is the causal effect of treatment vs. control at time j conditional on the observed past and setting all future treatment to be the control. For statistical inference, we could consider marginal structural models for global contrasts and structural nested models for local contrasts.[19]

In the causal-effects paradigm, we typically try to assess how much of the effect of treatment on the outcome is explained or mediated by the surrogate. For the proportion explained approach in the time-varying setting, we can update the two generalized linear models in the Review Section as $g\{E(Y|A_1, A_2)\} = \beta_0 + A_1\beta_{A_1} + A_2\beta_{A_2} + A_1A_2\beta_{A_1A_2}$ and $g\{E(Y|A_1, A_2, S)\} = \gamma_0 + A_1\gamma_{A_1} + A_2\gamma_{A_2} + A_1A_2\gamma_{A_1A_2} + S\gamma s$. Then, we can express a proportion of the total effect of A_1 and A_2 explained by S as $1-\{(a_1-a_1^*)\gamma_{A_1}+(a_2-a_2^*)\gamma_{A_2}+(a_1a_2-a_1^*a_2^*)\gamma_{A_1A_2}\}/\{(a_1-a_1^*)\beta_{A_1}+(a_2-a_2^*)\beta_{A_2}+(a_1a_2-a_1^*a_2^*)\beta_{A_1A_2}\}$.

From the mediation perspective we can discuss effects in terms of global and local contrasts. For a global contrast, the overall effect of treatment would be a contrast of what would have happened if varying A_1 and A_2, that is $Y^{a_1,a_2}-Y^{a_1^*,a_2^*}=Y^{a_1,a_2,S^{a_1,a_2}}-Y^{a_1^*,a_2^*,S^{a_1^*,a_2^*}}$, where $a_1 \neq a_1^*$ and $a_2 \neq a_2^*$, (all pathways from A_1 and A_2 to Y in Figure 2a), and the indirect effect of treatment involves all pathways except those from treatment directly into the outcome, that is, $Y^{a_1,a_2,S^{a_1,a_2}}-Y^{a_1,a_2,S^{a_1^*,a_2^*}}$, ($A_1 \rightarrow Y$ and $A_2 \rightarrow Y$ in Figure 2b).

Note that a variable may be a good surrogate for the joint effect of a time-varying treatment, but not for the effect of one treatment alone. This can be assessed using measures of the "proportion mediated", which for the joint effect of A_1 and A_2 could be expressed as the ratio $E(Y^{a_1,a_2}-Y^{a_1,a_2,S^{a_1^*,a_2^*}})/E(Y^{a_1,a_2}-Y^{a_1^*,a_2^*})$, and similarly for the effect of A_1 could be expressed as $E(Y^{a_1}-Y^{a_1,S^{a_1^*}})/E(Y^{a_1}-Y^{a_1^*})$. Given assumptions of consistency and randomization of the treatments and surrogate, these measures of proportion mediated can be expressed in terms of the observed data as follows:

$$\frac{\mathrm{E}\left(\mathrm{Y}^{\mathrm{a}_1,\mathrm{a}_2}-\mathrm{Y}^{\mathrm{a}_1,\mathrm{a}_2,\mathrm{S}^{\mathrm{a}_1^*,\mathrm{a}_2^*}}\right)}{\mathrm{E}\left(\mathrm{Y}^{\mathrm{a}_1,\mathrm{a}_2}-\mathrm{Y}^{\mathrm{a}_1^*,\mathrm{a}_2^*}\right)} = \frac{\sum_{\mathrm{s}}\mathrm{E}(\mathrm{Y}|\mathrm{A}_1=\mathrm{a}_1,\mathrm{A}_2=\mathrm{a}_2,\mathrm{S}=\mathrm{s})\{\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1,\mathrm{A}_2=\mathrm{a}_2)-\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1^*,\mathrm{A}_2=\mathrm{a}_2^*)\}}{\sum_{\mathrm{s}}\{\mathrm{E}(\mathrm{Y}|\mathrm{A}_1=\mathrm{a}_1,\mathrm{A}_2=\mathrm{a}_2,\mathrm{S}=\mathrm{s})\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1,\mathrm{A}_2=\mathrm{a}_2)-\mathrm{E}(\mathrm{Y}|\mathrm{A}_1=\mathrm{a}_1^*,\mathrm{A}_2=\mathrm{a}_2^*,\mathrm{S}=\mathrm{s})\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1^*,\mathrm{A}_2=\mathrm{a}_2^*)\}},$$

and

$$\frac{\mathrm{E}\left(\mathrm{Y}^{\mathrm{a}_1}-\mathrm{Y}^{\mathrm{a}_1,\mathrm{S}^{\mathrm{a}_1^*}}\right)}{\mathrm{E}\left(\mathrm{Y}^{\mathrm{a}_1}-\mathrm{Y}^{\mathrm{a}_1^*}\right)} = \frac{\sum_{\mathrm{s}}\mathrm{E}(\mathrm{Y}|\mathrm{A}_1=\mathrm{a}_1,\mathrm{S}=\mathrm{s})\{\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1)-\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1^*)\}}{\sum_{\mathrm{s}}\{\mathrm{E}(\mathrm{Y}|\mathrm{A}_1=\mathrm{a}_1,\mathrm{S}=\mathrm{s})\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1)-\mathrm{E}(\mathrm{Y}|\mathrm{A}_1=\mathrm{a}_1^*,\mathrm{S}=\mathrm{s})\mathrm{pr}(\mathrm{S}=\mathrm{s}|\mathrm{A}_1=\mathrm{a}_1^*)\}}.$$

For example, consider the simple setting where all variables are binary and both treatments A_1 and A_2 are randomized and the surrogate S is sequentially ignorable, i.e., independent of future potential outcomes given A_1 and A_2, so that the triple (A_1, A_2, S) is independent of potential outcomes $Y^{a_1,a_2,s}$ and (A_1, A_2) is independent of potential outcomes S^{a_1,a_2} (and similarly for A_1 alone). Then, for the data-generating process given in Table 1 (and considering the two global contrasts (a_1 = 1, a_2 = 1) versus (a_1 = 0, a_2 = 0) and a_1 = 1 versus a_1 = 0), S is a good surrogate for A_1 and A_2 jointly, but not for A_1 alone. Specifically, one can calculate that the proportion mediated for the joint effect of A_1 and A_2 is 100%, indicating that all of this joint effect is mediated by S, but that the proportion mediated for A_1 alone is only 3%, indicating that relatively little of the marginal effect of A_1 is mediated by S. This is the case even though in this example the putative surrogate S is a "consistent surrogate" in the sense that the effects of A_1 and A_2 jointly and of A_1 alone (marginally) on the surrogate are in the same direction as the corresponding effects on the outcome.[6] This example illustrates that one can severely underestimate the quality of a surrogate by using only partial information about a time-varying treatment.

Focusing on global contrasts may be reasonable if there is a common mechanism for the effect of sequential treatments or if the sequential treatments at different times have the same contents. For example, the global contrast would make sense in randomized trials with time-varying adherence or in observational studies of repeated treatments. It would make less sense if the sequential treatments have very different contents such as in sequential, multiple assignment, randomized trials (SMARTs).[20] For a local contrast, the overall effect of treatment A_1, $Y^{a_1,a_2}-Y^{a_1^*,a_2}=Y^{a_1,a_2,S^{a_1,a_2}}-Y^{a_1^*,a_2,S^{a_1^*,a_2}}$, consists of two pathways, one that goes through S and one that does not go through S (Figure 3a), and the indirect effect of treatment A_1 is the pathway that goes through S, that is, $Y^{a_1,a_2,S^{a_1,a_2}}-Y^{a_1,a_2,S^{a_1^*,a_2^*}}$, (Figure 3b).

While we focus on the treatment effect for A_1 in the local contrast, we also have to control for A_2, e.g., $A_2 = a_2$. There are multiple possibilities to be considered. For instance, if we are investigating an optimal treatment regime, we may want to choose A_2 to maximize the

utility function or to optimize the outcome. This is not an easy task when a surrogate S is still unproven. We may also be interested in the effects of treatment A_2. Similarly, there are the overall effect of treatment A_2, $Y^{a_1,a_2}-Y^{a_1,a_2^*}=Y^{a_1,a_2,S^{a_1,a_2}}-Y^{a_1,a_2^*,S^{a_1,a_2^*}}$, and the indirect effect of treatment A_2 with multiple possibilities of treatment A_1, $Y^{a_1,a_2,S^{a_1,a_2}}-Y^{a_1,a_2,S^{a_1,a_2^*}}$; see Figure 3c and 3d. In general, a good surrogate for treatment A_1 or A_2 would be, in the causal-effects paradigm, when the indirect effect is close to the overall effect and where a positive effect on the surrogate implies a positive effect on the outcome. A good surrogate for A_1 is not necessary a good surrogate for A_2.

In the causal-association paradigm for a time-varying treatment, we focus on the blip of treatment for a particular time. For example, let θ_{jt} be the effect in study/group j of blip of treatment at t on S and let ϕ_{jt} be the effect in group/study j of blip of treatment at t on Y. The focus would be on the association between θ_{jt} and ϕ_{jt} in this paradigm. Similar to the case of scalar quantities, S would be a good surrogate for A_t, if there is strong association between θ_{jt} and ϕ_{jt}; and positive effect of A_t on S implies positive effect on Y. Note that S may be a good surrogate for A_1, not for A_2; or vice versa. Similarly as in the causal effects paradigm, we can also consider θ_{jt} and ϕ_{jt} for global and local contrasts.

Another example of time-varying treatments can be found in SMARTs. In SMARTs, the goal is to find an optimal dynamic treatment regime. Let $\boldsymbol{H}_1$, ...,$\boldsymbol{H}_T$ denote the information on a subject's history up to times 1, ..., T; $\boldsymbol{H}_t$ can include the subject's treatment history before time (A_1,..., A_{t-1}), baseline covariates and intermediate variables that are measured after baseline but before or at time t. A dynamic treatment regime is a set of rules π_1 ($\boldsymbol{H}_1$),..., $\pi_T(\boldsymbol{H}_T)$ for assigning treatment at times 1, ..., T based on a patient's history. The Y-optimal dynamic treatment regime is the dynamic treatment regime which maximizes the expected value of the ultimate outcome of interest Y. See Laber et al. for discussion of estimating optimal dynamic treatment regimes.[21] Suppose there is a surrogate S that is measured after all of the treatments have been administered, i.e., after time T, but before the ultimate outcome of interest Y is measured. The S-optimal dynamic treatment regime is the dynamic treatment regime which maximizes the expected value of the outcome S. Our goal is to find a good treatment regime for the ultimate outcome of interest Y. The variable S is a good surrogate for this goal if the expected value of Y under the S-optimal dynamic treatment regime is close to the expected value of Y under the Y-optimal dynamic treatment regime, which can be evaluated by meta-analysis in the causal-association paradigm.

Time-varying surrogates

Here we briefly discuss some issues that can arise with time-varying surrogates. Considering generic surrogates S_1 and S_2 in Figure 4, each surrogate mediates a proportion of the effect of treatment; e.g., $A\rightarrow S_1\rightarrow Y$, $A\rightarrow S_2\rightarrow Y$, or $A\rightarrow S_1\rightarrow S_2\rightarrow Y$. In the HIV example, treatment A could be Zidovudine, surrogates S_1 and S_2 could be patient's CD4 count at the first and second time points of measurement, and clinical outcome Y could be time to death. Even if each surrogate mediates only a small proportion of the effect of treatment, the effects of treatment A on S_1 and S_2 individually may be good predictors of the effects of A on Y due to the common cause of S_1 and S_2, the two-headed black arc in Figure 4. In this case, there would be a divergence between causal-effects and causal-association measures for the

surrogacy for the individual level of *S*. When considering joint surrogacy of S_1 and S_2, both of them together as a vector could be better surrogates than each one individually. In the causal-association paradigm, let θ_{jv} be the effect in study/group *j* of treatment on S_v and let ϕ_j be the effect in group/study *j* of treatment on *Y*. The vector $\{S_1, ..., S_V\}$ could be a better predictor of ϕ_j than individual effects on single S_v.

In the causal-effects paradigm, the quality of S_1 and S_2 jointly versus individually can be assessed using "proportion mediated" measures. For instance, the ratio $E(Y^a - Y^{a,S_1^{a^*},S_2^{a^*,S_1^{a^*}}})/E(Y^a - Y^{a^*})$ represents the proportion mediated by both S_1 and S_2, and the ratio $E(Y^a - Y^{a,S_1^{a^*}})/E(Y^a - Y^{a^*})$ represents the proportion mediated by S_1 by itself (or marginally). Given assumptions of consistency, randomization of the treatment, and sequential ignorability of surrogates, these measures of proportion mediated can be expressed in terms of the observed data as follows:

$$\frac{\mathrm{E}\left(\mathrm{Y^a} - \mathrm{Y}^{\mathrm{a},\mathrm{S}_1^{\mathrm{a}^*},\mathrm{S}_2^{\mathrm{a}^*,\mathrm{S}_1^{\mathrm{a}^*}}}\right)}{\mathrm{E}\left(\mathrm{Y^a} - \mathrm{Y^{a^*}}\right)} = \frac{\mathrm{E(Y|A=a)} - \sum_{\mathrm{s_1}}\sum_{\mathrm{s_2}} \mathrm{E(Y|A=a, S_1=s_1, S_2=s_2)pr(S_2=s_2|A=a^*, S_1=s_1)pr(S_1=s_1|A=a^*)}}{\mathrm{E(Y|A=a)} - \mathrm{E(Y|A=a^*)}},$$

and

$$\frac{\mathrm{E}\left(\mathrm{Y^a} - \mathrm{Y}^{\mathrm{a},\mathrm{S}_1^{\mathrm{a}^*}}\right)}{\mathrm{E(Y^a - Y^{a^*})}} = \frac{\mathrm{E(Y|A=a)} - \sum_{\mathrm{s_1}} \mathrm{E(Y|A=a, S_1=s_1)pr(S_1=s_1|A=a^*)}}{\mathrm{E(Y|A=a)} - \mathrm{E(Y|A=a^*)}}.$$

As an example, consider a simple setting as in the Time-varying treatments Subsection where (*A*, S_1, S_2, *Y*) are all binary, and the treatment and surrogates are all sequentially ignorable. Then, for the distribution of observed data given in Table 2, the surrogates S_1 and S_2 taken together mediate a large proportion of the effect on *Y* but S_1 alone does not. Specifically, one can calculate that the proportion mediated by S_1 and S_2 together is approximately 99%, while the proportion mediated by S_1 alone is exactly zero. This illustrates that one can undervalue a surrogate by not incorporating its repeated measurements.

Discussion

We have discussed a conceptual framework for time-varying variables in both the causal-effects and causal-association paradigms, focusing in particular on time-varying treatments and surrogates. We extended notions of surrogacy for scalar variables to corresponding settings with time-varying variables. Both paradigms play important roles in the surrogate marker problem. In general, the causal-effects paradigm is often useful for proposing causal surrogates that are intermediate on the causal pathway from treatments to outcome, and the causal-association paradigm is often useful for evaluating proxy surrogates that are not necessarily on the causal pathway but may be correlated with causal surrogates. Different approaches within the two paradigms, for example using "proportion explained" measures,

indirect effects, meta-analysis, or principal stratification, each have their own advantages and disadvantages depending on the kind of surrogate being used and on the particular study objectives.

VanderWeele[6] discussed three types of surrogates: (1) mediator surrogates, those on the causal pathway from treatment to response; (2) proxy-mediator surrogates, those related to the intermediate variables on the causal pathway from treatment to response; and (3) prognostic surrogates, those not on the causal pathway and unrelated to the intermediate variables on the pathway. Different types of surrogates require different approaches. As presently developed in this paper, we have assumed time-varying surrogates in the causal-effects paradigm are also mediator surrogates. It is also possible that the time-varying surrogates in the causal-association paradigm are proxy-mediator or prognostic surrogates. How to think about the presence of time-varying proxy-mediator or prognostic surrogates in the causal-effects paradigm has not been fully investigated yet.

We have considered issues arising in surrogate assessment when there are time varying treatments or time varying surrogates. Another type of time varying setting that would be of interest to consider in future research is time varying outcomes, which could be either a repeated measure outcome or a survival outcome.

We have focused almost exclusively on conceptual issues in this paper. A very important topic we have not covered in much detail is how to do statistical estimation and inference in time-varying settings. In the supplemental material, we provide a brief illustration of statistical estimation in time-varying settings. This will be critical to explore in future work, for which we hope this paper has provided some motivation.

Acknowledgments

Grant support acknowledgement: This work was supported by the National Institutes of Health [grant number R01-DK090385].

Funding

This work was supported by the National Institutes of Health [grant number R01-DK090385].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med. 1997; 16:1515–1527. [PubMed: 9249922]
2. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Stat Med. 1997; 16:1965–1982. [PubMed: 9304767]
3. Taylor JMG, Wang Y, Thiebaut R. Counterfactual links to the proportion of treatment effect explained by a surrogate marker. Biometrics. 2005; 61:1102–1111. [PubMed: 16401284]
4. Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. Biometrics. 2008; 64:1146–1154. [PubMed: 18363776]
5. Li Y, Taylor JMG, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. Biometrics. 2010; 66:523–531. [PubMed: 19673864]

6. Vanderweele TJ. Surrogate measures and consistent surrogates. Biometrics. 2013; 69:561–581. [PubMed: 24073861]
7. Peto T. Surrogate markers in HIV disease. J Antimicrob Chemother. 1996; 37 (Suppl B):161–170. [PubMed: 8818838]
8. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. Biometrics. 2009; 65:530–538. [PubMed: 18759836]
9. Neyman J. On the application of probability theory to agricultural experiments. Statist Sci. 1990; 5:465–480.
10. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974; 66:688–701.
11. Pearl, J. Causality: Models, Reasoning, and Inference. Cambridge, U.K: Cambridge University Press; 2000.
12. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992; 11:167–178. [PubMed: 1579756]
13. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. Stat Med. 1989; 8:431–440. [PubMed: 2727467]
14. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3:143–155. [PubMed: 1576220]
15. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Ann Stat. 2012; 40:1816–1845.
16. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58:21–29. [PubMed: 11890317]
17. Qin L, Gilbert PB, Follmann D, et al. Assessing surrogate endpoints in vaccine trials with case-cohort sampling and the Cox model. Ann Appl Stat. 2008; 2:386–407. [PubMed: 19079758]
18. Gabriel EE, Gilbert PB. Evaluating principal surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. Biostatistics. 2014; 15:251–265. [PubMed: 24337534]
19. Robins, JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran, ME.; Berry, D., editors. Statistical models in epidemiology, the environment, and clinical trials. New York: Springer; 2000. p. 95-133.
20. Murphy SA. An experiment design for the development of adaptive treatment strategies. Stat Med. 2005; 24:1455–1481. [PubMed: 15586395]
21. Laber EB, Lizotte DJ, Qian M, et al. Dynamic treatment regimes: Technical and challenges and applications. Electron J Stat. 2014; 8:1225–1272. [PubMed: 25356091]

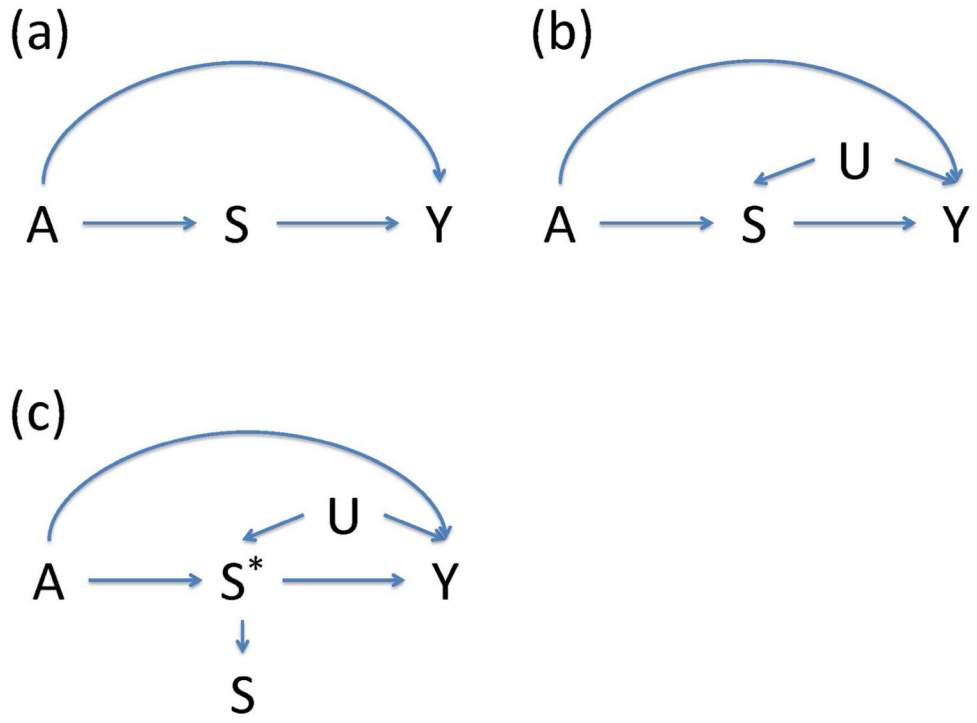

Figure 1.
Causal diagrams of a treatment *A*, a surrogate *S*, a proxy surrogate S^*, an unmeasured confounder *U*, and an outcome of interest *Y*

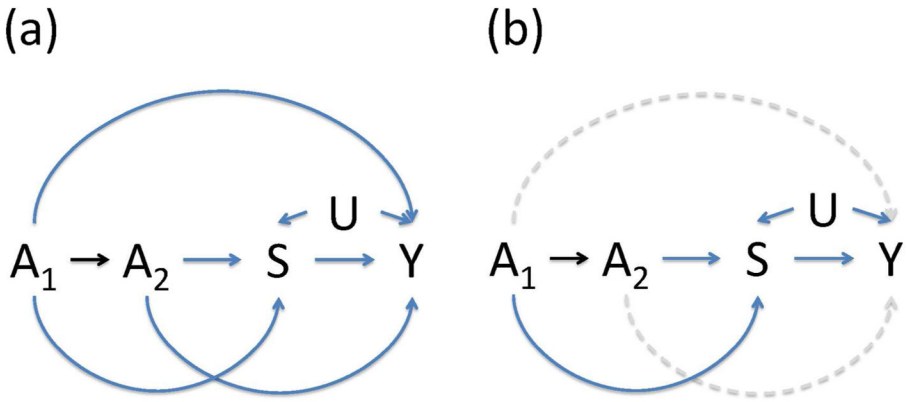

Figure 2.
Time-varying treatment causal diagrams of two treatments A_1 and A_2, a surrogate S, an unmeasured confounder U, and an outcome of interest Y

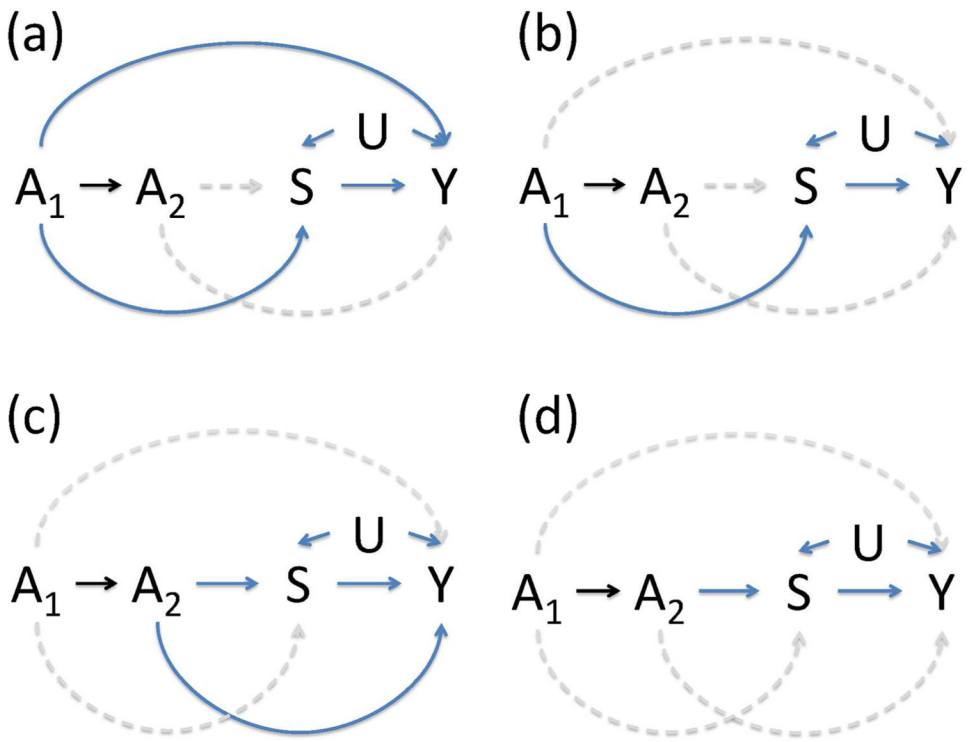

Figure 3.
Direct and indirect time-varying treatment causal diagram of two treatments A_1 and A_2, a surrogate *S*, an unmeasured confounder *U*, and an outcome of interest *Y*

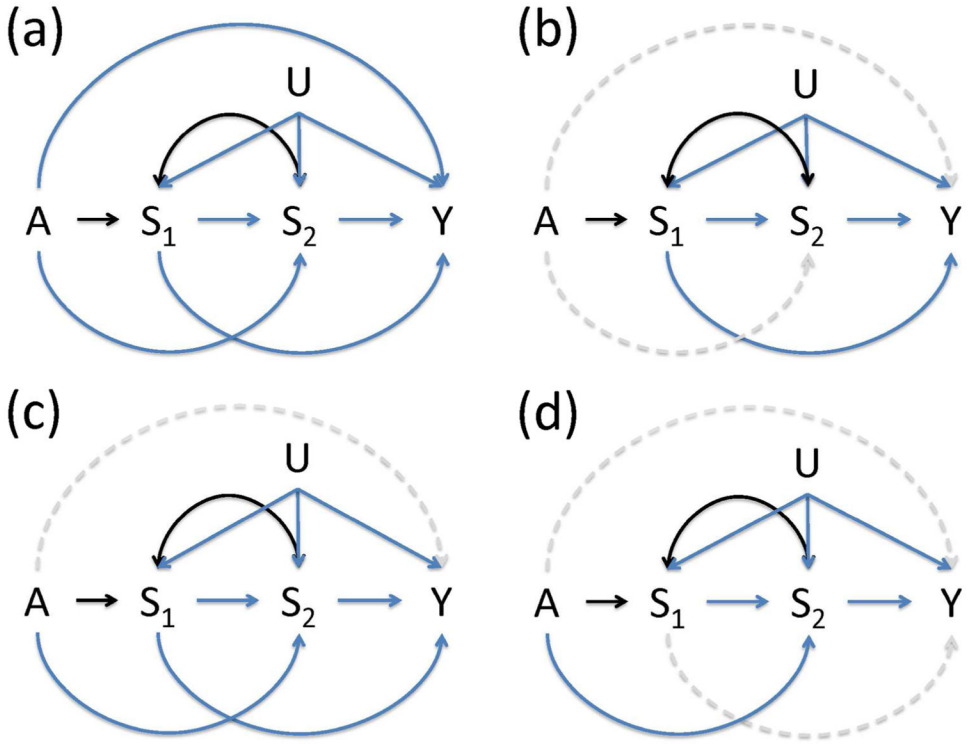

Figure 4.
Time-varying surrogate causal diagram of a treatment *A*, two surrogates S_1 and S_2, an unmeasured confounder *U*, and an outcome of interest *Y*

Table 1

Distribution of observed data for numerical example 1

S	A_1	A_2	pr(Y=1 | S, A_1, A_2)	A_1	A_2	pr(S=1 | A_1, A_2)
0	0	0	0.8	0	0	0.0
		1	0.9		1	0.3
	1	0	0.8	1	0	0.4
		1	0.8		1	0.5
1	0	0	-			
		1	0.3	A_1		pr(A_2=1 | A_1)
	1	0	0.8	0		0.8
		1	0.9	1		0.1

Table 2

Distribution of observed data for numerical example 2

A	S_1	S_2	pr(Y=1 | A, S_1, S_2)	S_1	A	pr(S_2=1 | S_1, A)
0	0	0	0.7	0	0	0.1
		1	0.1		1	0.9
	1	0	0.5	1	0	0.9
		1	0.1		1	0.1
1	0	0	0.6			
		1	0.9		A	pr(S_1=1 | A)
	1	0	0.7		0	0.5
		1	0.1		1	0.5