Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Links Between Causal Effects and Causal Association for Surrogacy Evaluation in a Gaussian Setting

A.S.C. Conlon[a]*, J.M.G. Taylor[a], Y. Li[a], K. Diaz-Ordaz[b] and M.R. Elliott[a]

Two paradigms for the evaluation of surrogate markers in randomized clinical trials have been proposed: the causal effects (CE) paradigm and the causal association (CA) paradigm [1]. Each of these paradigms rely on assumptions that must be made in order to proceed with estimation and to validate a candidate surrogate marker (S) for the true outcome of interest (T). We consider the setting in which S and T are Gaussian and are generated from structural models that include an unobserved confounder. Under the assumed structural models, we relate the quantities used to evaluate surrogacy within both the CE and CA frameworks. We review some of the common assumptions made in order to aid in estimating these quantities and show that assumptions made within one framework can imply strong assumptions within the alternative framework. We demonstrate that there is a similarity, but not exact correspondence between the quantities used to evaluate surrogacy within each framework and show that the conditions for identifiability of the surrogacy parameters are different from the conditions which lead to a correspondence of these quantities. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: Causal association; Direct effects; Principal stratification; Surrogate markers; Unmeasured confounders

1. Introduction

The validation of an intermediate marker (S) as a surrogate marker for the true outcome of interest (T) in clinical trials has attracted much attention [2-4]. An intermediate marker shown to be a valid surrogate would allow trials to be run more cheaply and quickly by basing analyses on the earlier or more cheaply measured surrogate. To use an example we will refer to throughout this paper, in a clinical trial assessing the efficacy of a new therapy on lengthening overall survival (OS) time in ovarian cancer, the duration of the trial could be reduced if the treatment effect on progression free survival (PFS) time could be used to infer a treatment effect on OS time. However, in practice demonstrating the validity of a surrogate marker has proven challenging, possibly due to the disease process affecting T through pathways not mediated through the surrogate, or due to unobserved confounders, U, of S and T [5]. For instance, suppose PFS time (S) is being assessed as a surrogate marker for OS time (T) in ovarian cancer patients in a trial with a binary treatment (Z), where $Z = 0$ is standard of care and $Z = 1$ is a new treatment. If patients experiencing longer PFS are also more likely to eat a healthy diet (U) which is also associated with longer OS, a treatment that prolonged PFS would also appear to prolong OS, indicating PFS as a potentially valid surrogate marker even though the survival benefit was not due to the induced treatment effect on PFS.

Several causal frameworks have been explored to identify intermediate markers as valid surrogates. Joffe and Greene [1] group these frameworks into two paradigms. The first, termed the "causal-effects" (CE) paradigm, attempts to separate the direct effect of the treatment (Z) on T from the effect of Z on T that is mediated through S. The second paradigm, termed the "causal association" (CA) paradigm, focuses on the association of the treatment effect on the surrogate and the

[a] *Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[b] *Department of Biostatistics, London School of Hygiene and Tropical Medicine, London, UK*
* *Correspondence to: achern@umich.edu*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

treatment effect on the true endpoint. VanderWeele [6] argued that conceptually the CA paradigm is more appealing when assessing surrogacy, although that may not be universally accepted. Each of the approaches to surrogacy assessment rely on certain assumptions that one must be willing to make in order to proceed with estimation and the evaluation of S as a surrogate. Ten Have and Joffe [7] and Ensor *et al.* [8] provide a comprehensive reviews of the estimation methods used and the assumptions made within both the CE and CA paradigms. Here, we explore the connections between some of the typical assumptions made within each paradigm, and examine the implications of these assumptions on the quantities used to determine the validity of S as a surrogate marker within each of the CE and CA paradigms.

The consideration of surrogacy from a causal perspective has some similarities to causal considerations of compliance and mediation, which can both be considered as intermediate variables between an intervention or exposure and an outcome. A number of the assumptions we consider originate in the compliance and mediation literature [7, 9-12].

The CE paradigm can be represented as a structural model in which one can explicitly change the values of Z or S or both, and the model specifies how the outcome T would then change. The indirect effect of Z on T is then the part of the effect of Z that is explained by changes in S holding Z constant, and the direct effect is the part of the effect of Z on T when the value of S is held constant. If S is a good surrogate for T, the direct effect of Z on T should be zero for all values of S. The Prentice [2] criteria for assessing surrogacy can be considered to be in the CE paradigm. These criteria require that S and T be correlated, that S be affected by Z, and that T and Z be conditionally independent given S. If the coefficient of Z is zero in the regression model for $T|Z, S$, then S would be considered a valid surrogate. Since this model conditions on the post randomization variable S, it will in general not have a causal interpretation. For the Prentice criteria to be valid from a causal perspective requires the assumption of no unmeasured confounders of S and T. This assumption is often unlikely to hold in the surrogate marker setting, where S and T are frequently involved in the same disease process. In general, the parameters in the CE paradigm are not estimable without assumptions, of which no unmeasured confounders is an example. In the PFS and OS example in ovarian cancer, this would preclude the possibility of a healthy diet affecting both PFS and OS. Assumptions weaker than no unmeasured confounders have been suggested in the literature, and these will be considered below.

The assessment of surrogacy within the CA paradigm includes methods based on principal stratification [13], which consider the distribution of the potential outcomes of T conditional on principal strata based on the values of the potential outcomes of S. In this framework, each subject has two potential outcomes (in the case of a binary treatment), one under $Z = 0$ and one under $Z = 1$, for each of the surrogate and the final outcome. Measures of surrogacy are derived from the distribution of the potential outcomes of T conditional on principal strata based on the potential values of the outcomes of S. S is considered to be a valid principal surrogate if there is no expected treatment effect on T within the principal stratum where there is no treatment effect on S. As the potential outcomes of S are pre-randomization variables, they can be regarded as baseline covariates, thereby avoiding the issue of potential unobserved confounding between the post randomization observed values of S and T. However, as only two of the four potential outcomes of S and T are observed for each person, assumptions must be made to aid in the estimation of unidentifiable parameters. Common assumptions often involve restrictions or assumptions on certain model parameters through the use of prior distributional assumptions [14], or conditional independence assumptions between certain counterfactual outcomes [15-17] or concepts of monotonicity, under which negative effects of the treatment on the surrogate marker or outcome are precluded [18]. For the ovarian cancer example, the potential outcomes for S are the two PFS times that would have arisen under each of the treatment arms, and the potential outcomes for T are the two OS times that would have arisen under the two treatments. The CE paradigm additionally requires consideration of what the OS time would be if the PFS time could be externally manipulated. While it is hard to specify how that could be achieved, it is never the less part of the conceptual framework of the CE paradigm.

The CE paradigm is consistent with a mechanistic view of causality as it describes how the output will change if the inputs are separately manipulated. By allowing the manipulation of S for fixed values of Z, this framework represents a larger, more general model. By considering the potential outcomes of S and T under each treatment arm, the CA framework does not require manipulations of S, as it is concerned with how the causal treatment effect on S is associated with the causal treatment effect on T and not with the effect of S on T. Pearl [19] invited a discussion on the uses and limitations of estimating effects using potential outcomes. A common argument against the use of potential outcomes and the principal stratification approach is the unidentifiability of the principal strata. It is argued that this lack of identifiability makes it difficult to make progress in estimation within this framework. However, estimation methods within the CE framework also rely on untestable assumptions and on conceiving of interventions on S, which may not always be possible [6, 20].

In this paper, in order to illustrate the connections between the CE and CA frameworks, we consider the setting in which S and T are Gaussian. While the relationship between the CE and CA frameworks has been considered previously in a general setting [1, 21], restricting to the Gaussian setting facilitates consideration of a larger number of different assumptions and also allows algebraic development, providing a more clear and concrete understanding of the relationship between these two frameworks. We assume that S and T are generated from structural models that include an unobserved

confounder. This model is detailed in Section 2. In Section 3, we explore the relationship between the parameters of the assumed Gaussian structural model and the model parameters in the principal surrogacy framework, and relate the structural model parameters to the parameters used to evaluate surrogacy within both the CE and CA frameworks. In Section 4, we briefly explore the role of baseline covariates in aiding in parameter estimation and in surrogacy evaluation. In Section 5, we review some of the common assumptions made within the CE framework to achieve identifiability and consider the impact of these assumptions on the parameters and quantities used to evaluate surrogacy in the CA framework. Section 6 explores some of the assumptions used to aid in estimation within the CA framework and their impact on the parameters within the CE framework. Section 7 presents a numerical study of the correspondence between the metrics of surrogacy under the various assumptions described in sections 5 and 6. Section 8 briefly outlines estimation methods for the parameters that are typically made in the two frameworks and explores how the explicit expressions derived for the relationship between the parameters and identifying assumptions of the two frameworks could aid in the estimation of surrogacy evaluation quantities. We conclude with a discussion in Section 9.

2. The structural model

Throughout the paper we will assume that the truth is a fairly general structural model, which is a model within the CE framework. We assume that both the surrogate marker S and the true endpoint T are continuous. We assume that the observed S_i is generated from a structural model which depends on the treatment, Z_i ($Z_i = 0$ or 1), and on an unobserved confounder, U_i, for each subject i, $i = 1, ..., n$. The observed value of T_i is also generated from a structural model that depends on Z_i, U_i and on S_i. Figure 1 provides a graphical representation of the assumed model.

We use the potential outcomes framework, and assume *no interference*, i.e. the potential outcomes of individual i are unaffected by the treatment and surrogate value of all other individuals. The assumed structural models for S_i and T_i are given by:

$$S_i(Z_i) = \alpha_0 + \alpha_1 Z_i + \alpha_2 U_i + \alpha_3 U_i Z_i + e^{S_i(Z_i)} \quad (1)$$

$$T_i(Z_i, s) = \beta_0 + \beta_1 Z_i + \beta_2 s + \beta_3 U_i + \beta_4 s Z_i + \beta_5 U_i Z_i + e^{T_i(Z_i)} \quad (2)$$

where $U_i \sim N(0,1)$, $e^{S_i(0)} \sim N(0, \delta^2_{S0})$, $e^{S_i(1)} \sim N(0, \delta^2_{S1})$, $e^{T_i(0)} \sim N(0, \delta^2_{T0})$, $e^{T_i(1)} \sim N(0, \delta^2_{T1})$ and U_i, $e^{S_i(0)}$, $e^{S_i(1)}$, $e^{T_i(0)}$, $e^{T_i(1)}$ are all uncorrelated. Note that the model is quite general in the sense that it does allow the outcome to depend on interactions between Z and U and between Z and S. To preclude having non Gaussian error terms, the model for $T_i(Z_i, S_i)$ does not include any interactions between U_i and S_i. Additionally, while the error associated with the potential outcome of $T_i(Z_i, S_i)$ changes with Z, there is no additional measurement error associated with S beyond that induced by the error of the selected $S(z)$, only a location shift for $T(Z, S(z))$ conditional on $e^{S(z)}$. Also note that the model does not include any other baseline covariates except for the intervention, which we assume to be randomly assigned. The situation with baseline covariates will be discussed later.

From the structural model we have the following for the four potential outcomes:

$S_i(0) = \alpha_0 + \alpha_2 U_i + e^{S_i(0)}$

$S_i(1) = \alpha_0 + \alpha_1 + (\alpha_2 + \alpha_3) U_i + e^{S_i(1)}$

$T_i(0) = T_i(0, S_i(0)) = \beta_0 + \beta_2 \alpha_0 + (\beta_2 \alpha_2 + \beta_3) U_i + \beta_2 e^{S_i(0)} + e^{T_i(0)}$

$T_i(1) = T_i(1, S_i(1)) =$

$\beta_0 + \beta_1 + (\beta_2 + \beta_4)(\alpha_0 + \alpha_1) + [(\beta_2 + \beta_4)(\alpha_2 + \alpha_3) + (\beta_3 + \beta_5)] U_i + (\beta_2 + \beta_4) e^{S_i(1)} + e^{T_i(1)}$

In addition to no interference, the structural model (2) assumes *consistency*, $T_i = T_i(z, s)$ if $Z_i = z, S_i = s$, which allows the observed outcome to be related to the potential outcome; and *positivity*, $P(Z_i = z \mid U_i = u) > 0$, $P(S_i(z) = s \mid Z_i = z, U_i = u) > 0$ for all z, $u \in \mathcal{U}$, and $s \in \mathcal{S}$. This implies that all treatments can be observed at all levels of potential confounders and that all levels of the surrogate marker are observable at all levels of potential confounders for all levels of treatment. The first part of the positivity assumption is trivially satisfied in the setting of a randomized trial.

The structural model described above has 14 parameters, which is the same number of parameters as the principal surrogacy model described in the following section. For the data that can be collected in a randomized trial, under the assumption that S and T are Gaussian, there are ten estimable quantities corresponding to the means and variances of $S(0)$, $S(1)$, $T(0)$ and $T(1)$ and the correlations of $(S(0), T(0))$ and $(S(1), T(1))$. While the structural model is a mechanistic model, the parameters still cannot all be estimated without untestable assumptions. Similarly, in the principal surrogacy model, there are ten estimable parameters; in the following section, we explicitly link the parameters of these two models. When there are no unmeasured confounders in the structural model, (i.e. $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$ or both), then all of the remaining parameters of this model are identifiable, as are the parameters of the principal surrogacy framework. In Sections 4 and 5, we explore some of the common assumptions made within the CE and CA frameworks and the impact of these assumptions on the quantities used to evaluate S as a surrogate marker.

3. The CA and CE frameworks

3.1. The causal association model

The CA paradigm of surrogacy evaluation includes methods based on the "principal surrogacy" framework of Frangakis and Rubin [13]. This framework focuses on the distribution of the potential outcomes of T conditional on principal strata defined by the values of the potential outcomes of S. Let $S_i(z)$ and $T_i(z)$ denote the potential outcomes of S_i and T_i, respectively, for subject i under treatment assignment $Z_i = z$. We assume the joint distribution of $(S_i(0), S_i(1), T_i(0), T_i(1))$ is multivariate normal [14] with mean μ and covariance matrix Σ, and has the following distribution:

$$\begin{pmatrix} S_i(0) \\ S_i(1) \\ T_i(0) \\ T_i(1) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{S_0} \\ \mu_{S_1} \\ \mu_{T_0} \\ \mu_{T_1} \end{pmatrix}, \begin{pmatrix} \sigma^2_{S_0} & \rho_s\sigma_{S_0}\sigma_{S_1} & \rho_{00}\sigma_{S_0}\sigma_{T_0} & \rho_{01}\sigma_{S_0}\sigma_{T_1} \\ & \sigma^2_{S_1} & \rho_{10}\sigma_{S_1}\sigma_{T_0} & \rho_{11}\sigma_{S_1}\sigma_{T_1} \\ & & \sigma^2_{T_0} & \rho_t\sigma_{T_1}\sigma_{T_0} \\ & & & \sigma^2_{T_1} \end{pmatrix} \right) \tag{3}$$

The mean μ and the variances corresponding to the diagonal elements of Σ, as well as the correlation parameters ρ_{00} and ρ_{11} are fully identifiable from the observed data. However, because only one of the counterfactual pairs of outcomes is observed for each subject, the correlation parameters ρ_s, ρ_{01}, ρ_{10}, and ρ_t are not identifiable from data. The parameters of the structural model detailed in Section 2 can be directly related to the principal surrogacy model when the joint distribution of the potential outcomes of S and T is multivariate normal. The formulas for all the μs, σs and ρs in terms of the αs, βs and δs are given in Appendix A.

While there is a direct mapping of the 14 parameters of the structural model to the 14 parameters of the principal surrogacy model, there is not an explicit formula to map the parameters of the principal surrogacy model back to the structural model parameters. For some combinations of parameters within the parameter space of the principal surrogacy model no parameter combinations within the parameter space of the structural model exist.

3.1.1. Measures of surrogacy. To evaluate S as a surrogate marker within the principal surrogacy framework, Gilbert and Hudgens [15] proposed two properties that a good surrogate should possess, "average causal necessity" (ACN) and "average causal sufficiency" (ACS). ACN requires that there be no conditional treatment effect on T within the principal stratum where there is no treatment effect on S, while ACS requires a non-zero conditional treatment effect on T within principal strata where there is a non-zero treatment effect on S. For the ovarian cancer trial example with PFS as a potential surrogate for OS, ACN would be met if patients who would experience the same PFS under either treatment arm would on average experience the same OS under either treatment. ACS would be met if patients who would experience greater PFS in one treatment arm would on average experience greater OS under this treatment arm. The primary quantities of interest from the multivariate normal model used to evaluate surrogacy can be derived from the conditional distribution of $(T(1) - T(0) \mid S(1) - S(0) = s)$, which in the joint Gaussian setting is normal with mean given by $E[T(1) - T(0)|S(1) - S(0) = s] = \gamma_0 + \gamma_1 s$, where:

$$\gamma_0 = (\mu_{T_1} - \mu_{T_0}) - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}} \right) (\mu_{S_1} - \mu_{S_0}) \tag{4}$$

$$\gamma_1 = \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}} \right). \tag{5}$$

ACN is then satisfied if $\gamma_0 = 0$ and ACS is satisfied if $\gamma_1 \neq 0$. We note that neither γ_0 nor γ_1 depend on ρ_t, however, the variance of $[T(1) - T(0)|S(1) - S(0) = s]$ does depend on ρ_t.

Based on the mapping of the parameters, the surrogacy quantities of interest in the CA framework γ_0 and γ_1 can be rewritten as:

$$\gamma_0 = \beta_1 + \beta_4(\alpha_0 + \alpha_1) - \alpha_1 \left(\frac{\beta_4(\alpha_3^2 + \alpha_3\alpha_2 + \delta^2_{S1}) + \beta_5\alpha_3}{\delta^2_{S0} + \delta^2_{S1} + \alpha_3^2} \right) \tag{6}$$

and

$$\gamma_1 = \beta_2 + \frac{\beta_4(\alpha_3^2 + \alpha_3\alpha_2 + \delta_{S1}^2) + \beta_5\alpha_3}{\delta_{S0}^2 + \delta_{S1}^2 + \alpha_3^2}. \tag{7}$$

The principal surrogacy criteria requiring that $\gamma_0 = 0$ and $\gamma_1 \neq 0$ will be met if $\beta_1 + \beta_4(\alpha_0 + \alpha_1) = \alpha_1\left(\frac{\beta_4(\alpha_3^2+\alpha_2\alpha_3+\delta_{S0}^2)+\beta_5\alpha_3}{\delta_{S0}^2+\delta_{S1}^2+\alpha_3^2}\right)$ and these quantities are greater than $-\beta_2\alpha_1$ for $\gamma_1 > 0$ and less than $-\beta_2\alpha_1$ for $\gamma_1 < 0$. Thus the requirements of the structural model within the CE framework for achieving principal surrogacy under the CA framework are not simple.

3.2. The causal effects model

3.2.1. Direct and indirect effects. The causal effects framework for surrogacy evaluation attempts to quantify the direct effect of Z on T, and the indirect effect of Z on T that is mediated through S. The notions of direct and indirect effects [22, 23] are defined by the counterfactual outcomes $S_i(z)$ and $T_i(z, s)$, where $S_i(z)$ is the value of S for subject i under treatment assignment $Z_i = z$ and $T_i(z, s)$ is the counterfactual outcome of T for subject i when Z_i is set to z and S_i is set to s. Robins and Greenland [22] and Pearl [23] provide definitions of the natural direct effect ($NDE(z)$), natural indirect effect ($NIE(z)$) and total effect (TE). The $NDE(z)$ measures the effect of Z on T when S is set to its potential value under treatment assignment z. The $NIE(z)$ measures the effect on T when Z is set to z and S is changed to what it would have been if Z were set to 1 compared to what it would have been if Z were set to 0. Finally, the TE of Z on T is equal to the sum of the $NIE(1)$ and $NDE(0)$ or to the sum of $NIE(0)$ and $NDE(1)$. Imai, Keele and Yamamoto [11] focus on the average causal effects of $NDE(z)$, $NIE(z)$ and TE defined by $E[NDE(z)]$, $E[NIE(z)]$, $E[TE]$ respectively. From the assumed structural model, these average causal effects are given in Table 1. The average causal effects correspond to the relevant component parameters in the structural model. For example, $E[NDE(0)]$ equals the direct effects of Z on T plus the effect of Z on S brought through the interaction between S and Z on T. Since $U \sim N(0, 1)$, the expected effects in Table 1 do not depend on the parameters associated with the unmeasured confounders. An additional notion to measure the direct effect of Z on T is the *controlled direct effect* [22, 23]. It measures the effect of a treatment on an outcome after intervening to fix the value of the surrogate S to the same value s for the whole population; in the context of the ovarian cancer example, this would correspond to an intervention that sets the PFS time to be equal across the population before estimating the treatment effect on OS. In terms of the counterfactuals, we can define it as $CDE = E[T(1, s) - T(0, s)] = \beta_1 + \beta_4 s$.

Note that if there is no interaction between the surrogate and the treatment in the structural model, then the CDE and the NDE coincide. However, the total effects decomposes into the sum of the NDE and the NIE, but such decomposition is not available when using the CDE, so we shall not consider the CDE any further in this paper.

3.2.2. Measures of surrogacy. A measure of surrogacy in the CE framework is the ratio of the indirect effect to the total effect, denoted by $PE(Z)$, which can also be interpreted as the proportion of treatment effect on T explained by S. From the assumed structural model, $PE(z)$ is given by:

$$PE(0) = \frac{E[T(0, S(1)) - T(0, S(0))]}{E[T(1, S(1)) - T(0, S(0))]} = \frac{\alpha_1\beta_2}{\beta_1 + \beta_2\alpha_1 + \beta_4(\alpha_0 + \alpha_1)} \tag{8}$$

and

$$PE(1) = \frac{E[T(1, S(1)) - T(1, S(0))]}{E[T(1, S(1)) - T(0, S(0))]} = \frac{\alpha_1(\beta_2 + \beta_4)}{\beta_1 + \beta_2\alpha_1 + \beta_4(\alpha_0 + \alpha_1)}. \tag{9}$$

For S to be considered a perfect surrogate marker, $E[NDE(z)]$ should be zero and $E[NIE(z)]$ should be non-zero, indicating that all of the effect of Z on T is mediated through S. The $PE(z)$ provides a measure of the proportion of treatment effect on T that is explained by S, and should be large for good surrogate markers and equal to one for a perfect surrogate. In the ovarian cancer example, an $E[NDE(z)]$ of zero would imply that the treatment only effects OS time through its effect on PFS time and a non-zero $E[NIE(z)]$ would be the effect on OS time due to the treatment effect induced on PFS time, net of any treatment effect.

3.2.3. Relationship to Prentice criteria. Special cases of the direct and indirect effects approach to determine surrogacy are the Prentice [2] criteria and the closely related mediation methods proposed by Baron and Kenny [9]. The Prentice criteria considers the regression model

$$E[T|S,Z] = \theta_0 + \theta_1 Z + \theta_2 S + \theta_3 SZ,$$

and S is considered a perfect surrogate if $\theta_1 = \theta_3 = 0$. From the structural model it can be shown that,

$$\theta_1 = \beta_1 - \beta_3 \frac{\delta^2_{S0}(\alpha_1+\alpha_0)(\alpha_2+\alpha_3)+\alpha_2^2(\alpha_2\alpha_1+\alpha_1\alpha_3-\alpha_0\alpha_3)-\alpha_0\alpha_2(\alpha_3^2+\delta^2_{S1})}{(\alpha_2^2+\delta^2_{S0})((\alpha_2+\alpha_3)^2+\delta^2_{S1})} - \beta_5 \frac{(\alpha_1+\alpha_0)(\alpha_2+\alpha_3)}{(\alpha_2+\alpha_3)^2+\delta^2_{S1}}$$

and

$$\theta_3 = \beta_4 - \beta_3 \frac{\alpha_2^2\alpha_3+\alpha_2(\alpha_3^2+\delta^2_{S1})-\delta^2_{S0}(\alpha_2+\alpha_3)}{((\alpha_2+\alpha_3)^2+\delta^2_{S1})(\alpha_2^2+\delta^2_{S0})} + \beta_5 \frac{(\alpha_2+\alpha_3)}{(\alpha_2+\alpha_3)^2+\delta^2_{S1}}.$$

The coefficients of the Prentice model (θ_0, θ_1, θ_2, θ_3) depend on the coefficients of the confounding variables in equations 1 and 2 (i.e. they depend on α_2, α_3, β_3, β_5). Therefore, for the assessment of surrogacy using the Prentice criteria to be a valid causal assessment of surrogacy, there must be no confounders of S ($\alpha_2 = \alpha_3 = 0$) or no confounders of T ($\beta_3 = \beta_5 = 0$) so that we have $\theta_1 = \beta_1$ and $\theta_3 = \beta_4$. In the absence of confounders, S will be considered a perfect surrogate marker for T based on the Prentice criteria if $\alpha_1 \neq 0$, $\beta_2 \neq 0$, $\beta_1 = 0$ and $\beta_4 = 0$ and subsequently $\theta_1 = \theta_3 = 0$. In this case, the direct effect of Z on T will be zero and all of the treatment effect will be completely mediated through S. The methods of Baron and Kenny [9] also require no unobserved confounders and additionally require there to be no interaction of Z and S ($\beta_4 = 0$). Then, if α_1 and β_2 are non-zero, $\alpha_1\beta_2$ can be interpreted as the mediation effect, or the effect of Z that is explained by S.

In the absence of unobserved confounders and no interaction effect of S and Z, Freedman, Graubard, and Schatzkin [3] proposed a quantity to measure the proportion of treatment effect explained by S, derived from the ratio of treatment effects estimated from two regression models for T, one with no adjustment for S and the other adjusting for S. Freedman's proportion explained is one minus this ratio, given by $p = 1 - \frac{\beta_1}{\beta_1+\alpha_1\beta_2}$, where $p = 1$ corresponds to a perfect surrogate. Wang and Taylor [24] proposed an estimate of the proportion of treatment effect explained by S that can be estimated from the observed data in the presence of an interaction of S and Z. For the structural model assumed here, this quantity is equivalent to $PE(z)$ in equations 8 and 9.

3.3. Correspondence between the CA and CE models

Within the CE framework, S will be considered a valid surrogate when the natural direct effects are zero, corresponding to $\beta_1 = 0$ and $\beta_4 = 0$ and the natural indirect effects are non-zero (α_1 and β_2 are non-zero). Within the CA framework, S is considered a valid surrogate if $\gamma_0 = 0$ and $\gamma_1 \neq 0$. VanderWeele [21] referred to the expected treatment effect on T within principal strata where there is no treatment effect on S, here corresponding to γ_0, as the "principal strata direct effect" and the expected treatment effect on T within principal strata where there is a treatment effect on S, here corresponding to γ_1, as the "principal strata indirect effect". Working at the individual level, VanderWeele [21] showed that when the natural direct effects are zero for all subjects, corresponding to the assumption $\beta_1 = \beta_4 = \beta_5 = 0$ and $\delta^2_{S0} = \delta^2_{S1} = \delta^2_{T0} = \delta^2_{T1} = 0$, there is no principal strata direct effect, corresponding to $\gamma_0 = 0$, therefore meeting the CA surrogacy criteria. In our case, where we are interested in the expected natural direct effects, which are zero when $\beta_1 = 0$ and $\beta_4 = 0$, the surrogacy quantities of the CA model will be $\gamma_0 = -\frac{\alpha_1\beta_5\alpha_3}{\delta^2_{S0}+\delta^2_{S1}+\alpha_3^2}$ and $\gamma_1 = \beta_2 + \frac{\beta_5\alpha_3}{\delta^2_{S0}+\delta^2_{S1}+\alpha_3^2}$ when the expected natural direct effects are zero. Therefore, when the criteria for surrogacy are met within the CE framework, the criteria within the CA framework will not always be satisfied, but will be met if either $\alpha_3 = 0$ or $\beta_5 = 0$, i.e. if either of the U_iZ_i interactions in Equation 1 or Equation 2 are zero.

When there is no interaction effect between S and Z on T ($\beta_4 = 0$), and no interaction between the unmeasured confounder U and Z for either the outcome ($\beta_5 = 0$) or the surrogate marker ($\alpha_3 = 0$), then there is a simple relationship between the proportion explained ($PE(z)$) measure in the CE framework and the ACN and ACS parameters γ_0 and γ_1 in the CA framework. In particular, $\beta_4 = 0$ implies $NDE(0) = NDE(1) = NDE = \beta_1$, $\alpha_3 = 0$ implies $E(s) = E(S(1)) - E(S(0)) = \alpha_1$, while $\beta_4 = 0$ and $\beta_5 = 0$ together imply that $\gamma_0 = \beta_1$ and $\gamma_1 = \beta_2$. Thus

$$PE(0) = PE(1) = PE = 1 - \frac{NDE}{TE} = \frac{\gamma_1 E(s)}{\gamma_0 + \gamma_1 E(s)} = 1 - \frac{\gamma_0}{\gamma_0 + \gamma_1 E(s)}.$$

Thus γ_0 can only be treated as analogous to the direct effect in the CE framework if there is no interaction effect of S and Z on T and there is no treatment interaction with the unobserved confounder on either the outcome or the surrogate marker.

3.3.1. Simulation experiments. The above algebra showed that the metrics of surrogacy in the CE framework (NDE(Z), NIE(Z) and PE(Z)) do not correspond to the metrics of surrogacy in the CA framework (γ_0 and γ_1) unless special conditions are met. To further understand the magnitude of the differences between the parameters and measures of surrogacy in the CE model and the CA model, we undertook a simulation experiment. We simulated a broad range of reasonable parameter combinations in the structural model. Additionally, the average total effect ($\beta_1 + \beta_2\alpha_1 +$

$\beta_4(\alpha_0 + \alpha_1))$ was constrained to be positive. Drawing the CE parameters in this way ensured that $\alpha_1 > 0$, $\beta_2 > 0$ and $(\beta_2 + \beta_4) > 0$, which is a reasonable assumption in the surrogate marker setting where any S being considered as a potential surrogate for T would be known to have an association with the treatment and with T. Additionally, under the distributional assumptions, the magnitude of the coefficients of the confounding variable on S and on T must be less than the magnitude of the coefficient of α_1 and β_2, respectively, and δ^2_{S0} and δ^2_{S1} are constrained to have the same values, as are δ^2_{T0} and δ^2_{T1}. Restricting the average total effect to be greater than zero ensures that $PE(z)$ is greater than zero. For each parameter set, we calculated the corresponding parameters in the CA model and also the measures of surrogacy in both frameworks. Details of the distributions used to generate parameters for the simulations are provided in Appendix B. The range of R^2 values for regression models of $T|Z$, $T|S$ and $T|U$ is also shown in Appendix B, and demonstrates that the way in which the parameters were simulated was not overly restrictive and leads to a wide spectrum of scenarios. We explored the sensitivity of the simulation results to the chosen distributions and found that the results appear generalizable to parameters arising from different distributions.

Figure 2 provides scatter plots of the correlation parameters of the CA model for the simulated CE model parameters. The plots show that ρ_s, ρ_t, ρ_{00} and ρ_{11} are almost always positive, that ρ_{00} and ρ_{11} are generally larger than the other four correlation parameters and that ρ_s and ρ_t are generally larger than ρ_{01} and ρ_{10}. Figure 3 provides a scatter plot of $E[NDE(0)]$ versus γ_0. We see that there is a close correspondence between the direct effects and γ_0. Figure 4 provides plots of γ_0 vs. γ_1 for different values of PE. When PE is small, γ_0 tends to be greater than zero. As PE increases, the distribution of γ_0 becomes more centered around zero. The plots show that although there is not a perfect concordance between the surrogacy measures in the two frameworks, similar conclusions regarding the validity of S as a surrogate marker will often be drawn from the two frameworks.

The above figures represent the degree of agreement between the two concepts of surrogacy, as if the joint distribution of all the counterfactual outcomes were known, that is all the parameters in Equations 1 and 2 were known. In practice the parameters would have to be estimated from observed data.

4. Baseline Covariates

In many settings, observed baseline covariates (X) are available that may explain some of the dependence between S and T, and explain some of the effect of Z on S and T. Often baseline covariates are sought that will control for any confounding of S and T. If X is a binary or categorical covariate, the models and assumptions within both the CE and CA frameworks could be made within strata defined by X. If X is a continuous covariate or a continuous linear combination of covariates, additional parameters could be added to the structural model to give:

$$S_i(Z_i) = \alpha_0 + \alpha_1 Z_i + \alpha_2 U_i + \alpha_3 U_i Z_i + \psi_1 X_i + \psi_2 X_i Z_i + e^{S_i(Z_i)}$$
$$T_i(Z_i, S_i) = \beta_0 + \beta_1 Z_i + \beta_2 S_i + \beta_3 U_i + \beta_4 S_i Z_i + \beta_5 U_i Z_i + \omega_1 X_i + \omega_2 X_i Z_i + e^{T_i(Z_i)}.$$

This model now has 18 parameters to estimate and leads to a new CA model given by: $\begin{pmatrix} S_i(0) \\ S_i(1) \\ T_i(0) \\ T_i(1) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{S_0} + \psi_1 X_i \\ \mu_{S_1} + (\psi_1 + \psi_2) X_i \\ \mu_{T_0} + \omega_1 X_i \\ \mu_{T_1} + (\omega_1 + \omega_2) X_i \end{pmatrix}, \begin{pmatrix} \sigma^2_{S_0} & \rho_s \sigma_{S_0} \sigma_{S_1} & \rho_{00} \sigma_{S_0} \sigma_{T_0} & \rho_{01} \sigma_{S_0} \sigma_{T_1} \\ & \sigma^2_{S_1} & \rho_{10} \sigma_{S_1} \sigma_{T_0} & \rho_{11} \sigma_{S_1} \sigma_{T_1} \\ & & \sigma^2_{T_0} & \rho_t \sigma_{T_1} \sigma_{T_0} \\ & & & \sigma^2_{T_1} \end{pmatrix} \right)$. The mean parameters of this model are estimable and, as the covariance matrix does not change with the addition of baseline covariates, there are still four correlation parameters that are not estimable. Full development of the structural model given in Equations 1 and 2 can be assumed to be conditional on X, making the common assumptions of conditional independence and sequential ignorability discussed in the next section more plausible. In Appendix D we describe the consequence of including additional covariates on the natural direct and indirect effects, on the Prentice criteria and on γ_0 and γ_1.

5. Assumptions made within the CE framework

The structural model assumed in Section 2 is not identifiable from the observed data. Therefore, assumptions must be made in order to aid in the estimation of the parameters and identification of the direct and indirect effects. We review some of the common identifying assumptions made within the CE framework and explore the implications of these assumptions on the parameters of the principal surrogacy model. The *no interference* assumption is expanded to mean that the treatment level

of one individual has no effect on the surrogate of another, and we require *generalized consistency*, namely $S(z) = S$, and $T(z, S(z)) = T$ when $Z = z$.

5.1. No unmeasured confounders

A critical assumption to identification within the causal effects framework is that there are no unobserved confounders driving the association between the outcome and the treatment or between the surrogate marker and the outcome. In the ovarian cancer example, the assumption of no unobserved confounders between the outcome and the treatment will be met because it is a randomized clinical trial. The assumption of no unobserved confounders between the surrogate marker and the outcome precludes the possibility that diet affects both PFS time and OS time, and therefore would only be a reasonable assumption to make in this context if diet was not thought to be associated with both of these outcomes, or if covariate information was available to be included in the model to sufficiently control for this association. Different versions of the no unmeasured confounders assumption are made in the literature.

Let X denote a set of measured covariates. Pearl [23] required *conditional exchangeability*, meaning that conditional on measured covariates X, treatment Z is "random", and that once we stratify according to Z and X, their level of S is also essentially random. More formally,

$$\begin{aligned} T_i(z) &\perp Z \mid X_i = x \\ T_i(z', s) &\perp S_i(z) \mid Z_i = z, X_i = x \end{aligned}$$

for all z, z' and $x \in \mathcal{X}$, implying no $Z - T$ confounding conditionally on observed covariates X, and no $S - T$ confounding conditionally on observed covariates X and Z. The first assumption is automatically satisfied in randomized trials.

The conditional exchangeability assumption is replaced by Imai, Keele, and Yamamoto [11] by *sequential ignorability*, defined as

$$T_i(z', s), S_i(z) \perp Z_i \mid X_i = x$$

$$T_i(z', s) \perp S_i(z) \mid Z_i = z, X_i = x$$

for all z, z' and $x \in \mathcal{X}$. Again, the first assumption is automatically satisfied in randomized trials; the second is stronger, especially in the setting we have here without covariates.

Under our assumption of randomized treatment, Pearl and Imai, Keele, and Yamamoto correspond. Under the structural model (2) without covariates, sequential ignorability implies Pearl's conditions for identification.

Petersen, Sinsi, and van der Laan [25] replace the assumption $T_i(z', s), S_i(z) \perp Z_i \mid X_i = x$ of Imai, Keele, and Yamamoto with the weaker assumption that the outcome rather than the joint distribution of the surrogate and the outcomes is independent of treatment: $T_i(z', s) \perp Z_i \mid X_i = x$, but require the additional assumption that the magnitude of the direct effect is independent of the potential values of the surrogate marker conditional on observed covariates:

$$E_{S(z)}\left[Y_i(1, s) - Y_i(0, s) \mid S_i(z) = s, X = x\right] = E_{S(z)}\left[Y_i(1, s) - Y_i(0, s) \mid X = x\right].$$

As with the assumptions of Pearl [23], the Petersen, Sinsi, and van der Laan requirements match those of sequential ignorability in a randomized trial setting.

While these identification assumptions hold without further parametric assumptions, we can translate them into our parametric structural model by noting

$$S_i(0) = \alpha_0 + \alpha_2 U_i + e_i^{S(0)}$$

$$S_i(1) = \alpha_0 + \alpha_1 + (\alpha_2 + \alpha_3) U_i + e_i^{S(0)}$$

$$T_i(0, s) = \beta_0 + \beta_2 s + \beta_3 U_i + e_i^{T(0)}$$

$$T_i(1, s) = \beta_0 + (\beta_2 + \beta_4) s + (\beta_3 + \beta_5) U_i + e_i^{T(0)}.$$

The requirement that $T_i(z', s) \perp S_i(z) \mid Z_i = z, X_i = x$ implies $\alpha_2 = 0$ or $\beta_3 = 0$ when $Z = 0$, and $\alpha_2 + \alpha_3 = 0$ or $\beta_3 + \beta_5 = 0$ when $Z = 1$, or, more concisely, $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$, so that either $S(Z_i)$ or $T(Z_i, S_i)$ is independent of U_i and thus U_i no longer confounds the surrogate marker and the outcome.

5.2. No interaction

Recent work by VanderWeele [26] has highlighted the important role of interactions in mediation analysis. Baron and Kenny [9] propose methods for mediation analysis based on solving a system of linear equations. In order to obtain causal interpretations of the parameters of their models, an assumption of no unmeasured confounders as well as no interaction is necessary. This leads to the structural model of Section 2 with $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$ and $\beta_4 = 0$. Then, $E[NDE(0)] = E[NDE(1)] = \beta_1$, $E[NIE(1)] = E[NIE(0)] = \alpha_1\beta_2$ and $E[TE] = \beta_1 + \beta_2\alpha_1$. Under these assumptions we have $\rho_s = \rho_{01} = \rho_{10} = \rho_t = 0$, and if $\delta^2_{S0} = \delta^2_{S1}$ and $\delta^2_{T0} = \delta^2_{T1}$ then $\rho_{00} = \rho_{11}$. For our data example in ovarian cancer, these assumptions imply that diet does not affect both PFS time and OS time and OS time changes with PFS time to the same degree under both treatment arms. This assumption may therefore be reasonable to make if there is clinical knowledge to support the notion that longer (shorter) PFS times will result in similarly longer (shorter) OS times, regardless of treatment given. Under these assumptions, γ_0 is equal to the natural direct effect ($E[NDE(0)] = E[NDE(1)] = \gamma_0 = \beta_1$) and $\gamma_1 = \beta_2$, leading to exact correspondence between the CE and CA measures of surrogacy. Therefore, if $\beta_1 = 0$ and $\beta_2 \neq 0$, S will be a valid surrogate for T from both the CE or CA model perspective.

5.3. Conditional independence assumption

Daniels *et al.* [12] work under the assumption of conditional independence between potential outcomes which assumes that $T(1, S(1))$, $T(1, S(0))$ and $T(0, S(0))$ are conditionally independent given $S(0)$ and $S(1)$. In the ovarian cancer example, this assumption implies that given two PFS times, s_0 and s_1, under $Z = 0$ and $Z = 1$, respectively, amongst the set of people who have potential outcomes s_0 and s_1, the OS times under $Z = 0$ and the OS times under $Z = 1$ are independent, and also independent of the OS time under $Z = 1$ for PFS time s_0. They note that this assumption is not necessary to estimate the direct and indirect effects, however, in their Bayesian estimation strategy for estimating NDE and NIE, these assumptions are needed to estimate features of the posterior distribution of these quantities, such as the posterior variance. The conditional covariances of these three outcomes from the structural model of Section 2 are as follows:

$$\text{Cov}[T(0, S(0)), T(1, S(1)) \mid S(0), S(1)] = \frac{\beta_3(\beta_3+\beta_5)\delta^2_{S_0}\delta^2_{S_1}}{\delta^2_{S_0}(\alpha_2+\alpha_3)^2+\delta^2_{S_1}(\alpha_2^2+\delta^2_{S_0})},$$

$$\text{Cov}[T(0, S(0)), T(1, S(0)) \mid S(0), S(1)] = \frac{\beta_3(\beta_3+\beta_5)\delta^2_{S_0}\delta^2_{S_1}}{\delta^2_{S_0}(\alpha_2+\alpha_3)^2+\delta^2_{S_1}(\alpha_2^2+\delta^2_{S_0})}, \text{ and}$$

$$\text{Cov}[T(1, S(1)), T(1, S(0)) \mid S(0), S(1)] = \frac{\delta^2_{S_0}\delta^2_{S_1}(\beta_3+\beta_5)^2+\delta^2_{S_0}\delta^2_{T_1}(\alpha_2+\alpha_3)^2+\delta^2_{S_1}\delta^2_{T_1}(\alpha_2^2+\delta^2_{S_0})}{\delta^2_{S_0}(\alpha_2+\alpha_3)^2+\delta^2_{S_1}(\alpha_2^2+\delta^2_{S_0})}$$

In order for the three conditional independence assumptions to hold, in the structural model we must have $(\beta_3 + \beta_5) = 0$ and the possibly unrealistic assumption that $\delta^2_{T_1} = 0$, making this assumption difficult to satisfy in most scenarios. In terms of the parameters of the causal association model, this assumption does not change the correlation parameters ρ_s, ρ_{00} and ρ_{10}, and only slightly alters ρ_t, ρ_{11} and ρ_{01}. The surrogacy quantities of interest, γ_0 and γ_1, are unchanged by this assumption.

5.4. Exclusion restriction

Many of the assumptions discussed so far required no unobserved confounding. The *instrumental variable* approach does not make assumptions about S–T confounding, but instead assumes that all of the effect of Z on the outcome is mediated by the intermediate variable S_i, i.e. that the direct effect of Z is zero, i.e. $T_i(1, s) = T_i(0, s)$. This assumption is called *exclusion restriction*. More specifically, in the setting where the intermediate variable is binary (e.g. binary mediator or binary indicator of compliance to treatment), the exclusion restriction assumption requires that the distribution of the potential outcomes of T be independent of treatment assignment in the principal strata defined by the potential intermediate variable. So for the *never-takers*, $(S_i(0) = S_i(1) = 0)$, and the *always-takers* $(S_i(0) = S_i(1) = 1)$, this implies $T_i(1, s) = T_i(0, s) = \beta_0 + \beta_2 s + \beta_3 U_i$ [31], and thus $\beta_1 = \beta_4 = \beta_5 = 0$. In the continuous setting, Holland [32] and Sobel [33] have a similar requirement for identifiability, requiring that $\beta_1 + \beta_4 s + \beta_5 U = 0$. While in the compliance literature, it is often reasonable to assume that the treatment has no direct effect on the outcome, we note that the exclusion restriction is not compatible with the goals of surrogacy evaluation, as it assumes that the direct effect of treatment on the outcome is zero, which in turn assumes that S is a valid surrogate marker [7], and would therefore never be a reasonable assumption to make in this setting.

6. Assumptions made within the CA framework

As in the CE setting, some parameters of the principal surrogacy model are unidentifiable from the data, requiring assumptions to be made to aid in estimation. The assumptions that are typically made vary based on the setting being

explored and on the quantities of interest. In some settings, baseline covariate information is available that can aid in estimating the missing potential outcomes of S, or a "constant biomarker" assumption can be made about the potential outcomes of S in the control arm [15, 27]. Outside of these settings, assumptions must be placed on certain model parameters or on certain relationships between potential outcomes in order to proceed with estimation. While there is not a one-to-one mapping of the principal surrogacy model parameters to the structural model parameters as there is from the structural model parameters to the principal surrogacy model parameters, the assumptions made in the principal surrogacy setting have implicit effects on the parameters of the structural model.

6.1. Prior assumptions on correlation parameters

Within the setting of multivariate normally distributed outcomes of $S(0)$, $S(1)$, $T(0)$, and $T(1)$, Conlon, Taylor and Elliott [14] used a Bayesian estimation strategy and placed different plausible prior assumptions on the unidentified correlation parameters. These assumptions, along with the positive definite restriction of the covariance matrix aided in estimation. The assumptions made include restricting the correlation parameters to be positive and a restriction with respect to the ordering of the magnitudes of the correlations. These assumptions are reasonable in many surrogate marker settings, where the surrogate marker and the final outcome are often part of the same disease process. In the context of the ovarian cancer example, it would be reasonable to assume that the correlation parameters are positive, especially if the observed correlations between PFS time and OS time within each treatment arm are positive. It may also be reasonable to assume that the correlation between PFS time and OS time within the same treatment arm, the correlation between PFS times across treatment arms and the correlation between OS times across treatment arms are larger than the correlations between PFS time and OS time in opposite treatment arms. The implications of these assumptions on the parameters of the CE model are explored below.

6.1.1. Positivity of correlations. One assumption made by Conlon, Taylor and Elliott [14] restricts all of the correlation parameters to be positive. This assumption is motivated by the fact that S and T are usually scientifically or biologically related, and therefore if a person has an inherent frailty then this will result in both S and T being higher (or lower) irrespective of the treatment that they receive. In terms of the structural model, if we assume that $(\beta_2 + \beta_4) \geq 0$, which would be expected in any setting where S is being considered as a potential surrogate marker, restricting the correlation parameters to be positive requires one of the following two settings: (1) $\alpha_2 > 0$, $(\alpha_2 + \alpha_3) > 0$, $(\beta_2\alpha_2 + \beta_3) > 0$ and $(\beta_3 + \beta_5) + (\beta_2 + \beta_4)(\alpha_2 + \alpha_3) > 0$ or (2) $\alpha_2 < 0$, $(\alpha_2 + \alpha_3) < 0$, $(\beta_2\alpha_2 + \beta_3) < 0$ and $(\beta_3 + \beta_5) + (\beta_2 + \beta_4)(\alpha_2 + \alpha_3) < 0$. These settings imply that the effect of U must act in the same direction on both S and T. In our ovarian cancer data example, this implies that healthy diets are associated with both longer PFS time and longer OS time and would not be associated with a longer PFS time combined with a shorter OS time or vice versa. Figure 3 provides scatter plots from the simulation experiment of the correlation parameters of the CA model for the simulated CE model parameters. The scatter plots show that under the assumed structural model, all six of the correlations are greater than zero a majority of the time, with ρ_{00} and ρ_{11} nearly always positive and ρ_s and ρ_t usually positive, indicating that the positivity assumption, at least for ρ_s, ρ_t, ρ_{00} and ρ_{11}, would be reasonable in this setting.

6.1.2. Ordering of correlations. Another assumption explored by Conlon, Taylor and Elliott [14] restricts all of the correlation parameters to be positive and also restricts ρ_{10} and ρ_{01} to be less than the other four correlation parameters. This constraint is reasonable as ρ_{10} and ρ_{01} are measures of the correlation between S and T in opposite treatment arms, which is unlikely to be larger than the correlation between the S and T within the same treatment arm, or the correlation between the surrogate responses or final treatment responses across treatment arms. As not all combinations of parameter values of the principal surrogacy model are possible under the assumed structural model, it can be shown that one such set of parameters arises under the restriction of positivity and ordering of the correlations. If only the assumption about the ordering of the correlations is imposed and positivity is not assumed, then one of the following two settings is implied in terms of the structural model: (1) $\alpha_2 > 0$, $(\alpha_2 + \alpha_3) > 0$, $(\beta_2\alpha_2 + \beta_3) < 0$, and $(\beta_2 + \beta_4)(\alpha_2 + \alpha_3) + (\beta_3 + \beta_5) < 0$ or (2) $\alpha_2 < 0$, $(\alpha_2 + \alpha_3) < 0$, $(\beta_2\alpha_2 + \beta_3) > 0$, and $(\beta_2 + \beta_4)(\alpha_2 + \alpha_3) + (\beta_3 + \beta_5) > 0$. These settings imply that the effect of U on S must be in the same direction for $Z = 0$ and $Z = 1$ and the effect of U on T must be in the opposite direction as that of U on S, but the effect of U on T must be in the same direction for $Z = 0$ and $Z = 1$. In terms of the ovarian cancer example, this would imply that patients with healthy diets have longer (shorter) PFS time, regardless of their treatment assignment, but shorter (longer) OS time in either treatment arm. The scatter plots in Figure 3 show that under the assumed structural model, the assumption that ρ_{00} and ρ_{11} are greater than ρ_{10} and ρ_{01} appears to hold nearly all the time, and the assumption that ρ_s is greater than ρ_{10} and ρ_{01} holds the majority of the time. However, the assumption that ρ_t is greater than ρ_{10} and ρ_{01} holds only about half of the time

6.2. Conditional independence assumptions

Another approach to estimation in this framework involves reducing the number of unidentified parameters that must be estimated through assumptions about conditional independences. One common conditional independence assumption that has been considered is that of conditional independence of $T(0)$ and $T(1)$ given $S(0)$ and $S(1)$ [15, 16, 28, 29]. This assumption reduces the number of unidentified parameters by one, as ρ_t becomes a function of the other five correlation parameters. Specifically, this implies that $\rho_t = \frac{\rho_{11}\rho_{10}+\rho_{01}\rho_{00}-\rho_s(\rho_{01}\rho_{10}+\rho_{11}\rho_{00})}{(1-\rho_s^2)}$. In terms of the structural model, conditional independence of $T(0, S(0))$ and $T(1, S(1))$ given $S(0)$ and $S(1)$ implies that $\beta_3(\beta_3 + \beta_5) = 0$. Therefore, in order for this conditional independence assumption to hold we must have either $\beta_3 = 0$ or $(\beta_3 + \beta_5) = 0$, i.e. there is zero effect of the unmeasured confounder in one of the treatment arms on the outcome T. This would imply in the ovarian cancer trial example that patients with healthy diets have similar OS times to those with unhealthy diets in at least one of the treatment arms, and would therefore only be reasonable to make if diet is not thought to be associated with both OS time and PFS time.

A different conditional independence assumption was made by Parast, McDermott and Tian [17] who assumed $S(0)$ and $T(1)$ were conditionally independent given $S(1)$ and that $S(1)$ and $T(0)$ were conditionally independent given $S(0)$, implying in the ovarian cancer example that given knowledge of PFS time under one treatment arm, OS time in the same treatment arm and PFS in the opposite treatment arm are independent. The consequence of these assumptions is the following

$$\frac{\rho_{01}}{\rho_{11}} = \frac{\rho_{10}}{\rho_{00}} = \rho_s.$$

The consequence if this, derived from the equations in Appendix C, requires $\alpha_2 = \alpha_3 = \beta_2 = \beta_4 = 0$ and also holds for selected other parameter combinations.

A similar, but weaker, conditional independence assumption [30] is that $S(0)$ and $T(1)$ were conditionally independent given $S(1)$ and $T(0)$ and that $S(1)$ and $T(0)$ were conditionally independent given $S(0)$ and $T(1)$, implying in the ovarian cancer example that given knowledge of both PFS time under one treatment arm and OS time in the opposite treatment arm, OS time and PFS in the other treatment arms are independent. The consequence of these assumptions are the following:

assuming $S(0) \perp T(1)|S(1), T(0)$ gives

$$\frac{\rho_S\rho_{11} - \rho_{00}\rho_{10}\rho_{11} + \rho_T\rho_{00} - \rho_S\rho_T\rho_{10}}{\rho_{01}(1-\rho_{10}^2)} = \frac{\sigma_{S_0}^2\sigma_{T_1}^2}{\sigma_{S_1}^2\sigma_{T_0}^2}$$

and assuming $S(1) \perp T(0)|S(0), T(1)$ gives

$$\frac{\rho_S\rho_{00} - \rho_{11}\rho_{01}\rho_{00} + \rho_T\rho_{11} - \rho_S\rho_T\rho_{01}}{\rho_{10}(1-\rho_{01}^2)} = \frac{\sigma_{S_1}^2\sigma_{T_0}^2}{\sigma_{S_0}^2\sigma_{T_1}^2}$$

6.3. Monotonicity assumption

Within the setting of a binary surrogate and final outcome, Li, Taylor and Elliott [18] impose a monotonicity assumption to aid in the problem of non-identifiability. Specifically, they require that $S_i(1) \geq S_i(0)$ and $T_i(1) \geq T_i(0)$ for all i. In terms of the structural model, this requires that $\alpha_1 + \alpha_3 U_i + e^{S_i(1)} \geq e^{S_i(0)}$ and $\beta_1 + \beta_4(\alpha_0 + \alpha_1) + \beta_2\alpha_1 + [\beta_4(\alpha_2 + \alpha_3) + \beta_2\alpha_3 + \beta_5]U_i + (\beta_2 + \beta_4)e^{S_i(1)} + e^{T_i(1)} \geq \beta_2 e^{S_i(0)} + e^{T_i(0)}$, which cannot be satisfied with Gaussian random variables. If monotonicity is only required to hold in expectation so that $E[S_i(1)] \geq E[S_i(0)]$ and $E[T_i(1)] \geq E[T_i(0)]$, this reduces to $\alpha_1 \geq 0$ and $\beta_1 + \beta_4(\alpha_0 + \alpha_1) + \beta_2\alpha_1 \geq 0$. As α_1 and β_2 are assumed to be positive within the surrogate marker setting, this assumption will hold as long as the average total effect of Z on T is positive. In the ovarian cancer setting, this holds if on average the combined effect of treatment and PFS time in the $Z = 1$ arm on OS time is greater than this combined effect on OS time in the $Z = 0$ arm, and would be reasonable to assume in this scenario for a treatment thought to improve OS time, as PFS time is known to be positively associated with the OS time.

7. Numerical study of impact of assumptions on correspondence between the CE and CA metrics of surrogacy

The assumptions described in the previous sections are made either because they are reasonable in the scientific context or because they aid in estimation of quantities of interest. In this section we evaluate whether making these assumptions also leads to closer correspondence between the metrics of surrogacy in the two frameworks. Using the simulation experiment described in Section 3.3.1, we plot the distribution of γ_0 and of γ_1 when $|\gamma_0| \leq 0.25$ for different ranges of PE(0). We

note that the conditional independence assumption made by Daniels *et al.* [12] is not included, as the condition cannot be met under the parameter distributions used in our simulations. In the simulation experiment, for each assumption we only retain the draws of the parameters that either exactly or approximately satisfy the assumption.

7.1. Ordering of correlations assumption in CA framework

Under the assumption of Section 5.1.2 that ρ_s, ρ_t, ρ_{00} and ρ_{11} are all positive and that $\rho_{01} < min(\rho_s, \rho_t, \rho_{00}, \rho_{11})$ and $\rho_{10} < min(\rho_s, \rho_t, \rho_{00}, \rho_{11})$ the boxplots in Figure 5(b) show that the correspondence between the measures of surrogacy in the CE and CA frameworks is slightly improved as compared to the model without parameter restrictions (boxplot shown in Figure 5(a)). There is an increase in concordance between γ_0 and PE(0), with γ_0 decreasing as PE(0) increases.

7.2. Conditional independence assumptions in CA framework

The first conditional independence assumption of Section 5.2 is that $T(0)$ and $T(1)$ are independent given $(S(0), S(1))$. The second conditional independence assumption of Section 5.2 is that $T(0)$ and $S(1)$ are independent given $S(0)$ and that $T(1)$ and $S(0)$ are independent given $S(1)$ and the third conditional independence assumption of Section 5.2 is that $T(1)$ and $S(0)$ are independent given $S(1)$ and that $T(0)$ and $S(1)$ independent given $S(0)$.

Under the first two assumptions, the boxplots in Figures 5(c) and 5(d) show that the relationship between γ_0 and PE(0) is brought into slightly higher concordance by making these assumptions. In our simulations there were no cases where $|\gamma_0| \leq 0.25$ when PE(0)$= 0.25$, indicating that when S is a poor surrogate, the CA framework and the CE framework would always agree. However, the relationship between γ_1 and PE(0) is in somewhat less concordance compared to the model with no parameter restrictions, with very little increase in γ_1 as PE(0) increases. Under the third conditional independence assumption, the boxplots in Figures 5(e) also show an increased concordance between the CE and CA measures of surrogacy, with γ_1 increasing as PE(0) increases but with slightly less concordance of γ_0 and PE(0) as compared to the first and second conditional independence assumptions.

7.3. Sequential ignorability assumptions in CE framework

Under the assumption of Section 4.1 that $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$ the boxplots in Figure 6(b) and Figure 6(c) show some increase in the concordance between γ_0 and PE(0), with γ_0 larger when $PE(0)$ is small and moving toward 0 as $PE(0)$ increases, but little additional concordance between γ_1 and $PE(0)$ is achieved by making this assumption.

7.4. Sequential ignorability and no interaction assumptions in CE framework

The sequential ignorability assumption of Section 4.3 together with the no interaction assumption of Section 4.4 is (i) $\beta_4 = 0$ and (ii) $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$. For the combined sequential ignorability and no interaction assumption, the boxplot in Figure 6(d) shows a similar relationship between the CE and CA measures of surrogacy as with the sequential ignorability assumption alone, with some increase in the concordance between γ_0 and PE(0), but little additional concordance of γ_1 with $PE(0)$ as compared to the model with no parameter restrictions.

8. Estimation and Sensitivity Analyses

8.1. Estimation

In the CA framework, the approach to estimation of the parameters in the multivariate normal model (Equation 3) is relatively straightforward. Either equality types of assumptions are made to make the model identifiable and likelihood based methods are used, or inequality types of constraints can be expressed in the form of prior distributions and a Bayesian approach can be taken using MCMC methods. From the estimates, inference about γ_0 and γ_1 is easy either from the delta method or directly from the MCMC draws. The Bayesian approach for not fully identified models is not without its challenges [34], especially if non-informative or only very weakly informative priors are used. In our experience [18], MCMC algorithms can be slow to converge, and from a frequentist perspective the coverage rates of 95% credible intervals can deviate from the desired level. In the CE framework, estimation of the direct and indirect effects derived from the parameters in Equations 1 and 2 usually proceeds by making identifying assumptions, such as sequential ignorability, and then non-parametrically estimating the direct and indirect effects [7,11,35,36].

Since the parameters in the CA framework are a direct function of the parameters in the CE framework, a different approach to estimation of γ_0 and γ_1 is to undertake estimation of the CE parameters in Equations 1 and 2 using a Bayesian approach, making assumptions that are appropriate for the context, and then mapping these directly to the CA parameters to obtain γ_0 and γ_1. It may even be possible to make reasonable, but not strong assumptions in both

frameworks simultaneously using prior distributions, and then undertake Bayesian estimation to obtain inference for γ_0 and γ_1. For example, one might assume approximate sequential ignorability (as in Section 4.2) and approximate conditional independence (as in Section 5.2) and make inequality assumptions (as in Section 5.1).

8.2. Sensitivity Analyses

The estimation methods presented here for the CE approach assume no unobserved confounding. In the context of a randomized treatment, as is the case here, this assumption translates to "no unobserved confounding" of the surrogate-outcome relationship. Many sensitivity analyses to this unobserved confounding have been proposed, but the strategy to follow will depend on whether there is an interaction between the treatment and the surrogate and the type of the outcome (i.e. continuous or binary). In the absence of a treatment-surrogate interaction, the NDE and the CDE coincide, and thus simpler sensitivity analysis techniques, which are available for the CDE, can be employed. VanderWeele [37] develops an approach for binary confounders that computes bias in the CDE as the product of the expected difference in the outcome at the two levels of the confounder conditional on treatment and the expected difference in the confounder at the two levels of treatment. Imai *et al.* [11] propose a sensitivity analysis that fits more closely with the structural model proposed here, by introducing a correlation between the error terms in the structural equations for T and S. Beginning with equations analogous to Equations (1) and (2), we have,

$$S_i = \alpha_0 + \alpha_1 Z_i + \psi_1 X_i + e_S \tag{10}$$

$$T_i = \beta_0 + \beta_1 Z_i + \beta_2 s + \omega_1 X_i + \beta_4 s Z_i + e_T \tag{11}$$

Where X_1 is a measured baseline confounder and allow for the error terms e_S and e_T to be correlated. The correlation between these error terms, ρ, thus becomes the sensitivity parameter that the user must specify. Imai *et al.* [11] then give expressions for the NDE and the NIE in terms of the correlation term, and other parameters that can be estimated from the observed data. This method works for both continuous and binary outcomes, and has been implemented in the command `medsens`, as part of the R software package `mediation`. In the context of Equations (1) and (2), if we assume no confounder treatment interaction then the correlation between e_S and e_T is proportional to $\alpha_2\beta_3$. Thus a sensitivity analysis could consist of estimation with this product held fixed.

An alternative approach within the CE framework, not yet attempted to our knowledge, would be to study the sensitivity to departures from the identification assumptions by using prior distributions with small variances. For example, instead of sequential ignorability assumption 1 in Table 2 that $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$ set $\alpha_2 \sim N(0, \sigma^2_{\alpha_2})$, $\alpha_3 \sim N(0, \sigma^2_{\alpha_3})$ and $\beta_3 \sim N(0, \sigma^2_{\beta_3})$, $\beta_5 \sim N(0, \sigma^2_{\beta_5})$ where $\sigma^2_{\alpha_2}$, $\sigma^2_{\alpha_3}$, $\sigma^2_{\beta_3}$ and $\sigma^2_{\beta_5}$ are all small. For other parameters less informative priors would be used. Then proceed with Bayesian estimation.

9. Discussion

Within the setting of Gaussian surrogate and final outcome variables, we have explored the connection between the quantities used to evaluate surrogacy within the CE framework of surrogacy assessment and the CA framework of surrogacy assessment. Under the assumed structural models for S and T, there is a direct mapping of the 14 parameters of these models to the 14 parameters of a Gaussian principal stratification model. Not all of these 14 parameters can be identified from the observed data alone, and therefore assumptions must be made in order to proceed with estimation. We have reviewed some of the common assumptions made within the CE and the CA frameworks, and explored the consequences of these assumptions on model parameters and quantities used to determine surrogacy. With parameter values from the assumed structural model that are reasonable in the surrogate marker setting, there is a close correspondence between the natural direct effect and average causal necessity. Under the assumptions of Baron and Kenny [9] of no interaction ($\beta_4 = 0$) and no unobserved confounding ($\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$) the surrogacy evaluation quantities in the CE and CA framework are equivalent, with $E[NDE(0)] = E[NDE(1)] = \gamma_0 = \beta_1$. This equivalence also holds under slightly weaker conditions of $\beta_4 = 0$ and either $\alpha_3 = 0$ or $\beta_5 = 0$, however these conditions do not lead to identifiability. With the exception of the assumptions made by Baron and Kenny [9], the assumptions made within the CE or the CA framework that aid in estimation do not aid in bringing the surrogacy evaluation quantities in closer alignment.

Most estimation methods within the CE framework rely on assumptions about the absence of post-treatment unobserved confounders of S and T. This assumption is untestable and may be unlikely to hold in the surrogate marker setting, where S and T are usually involved in the same disease process. In contrast, the CA framework does not require assumptions about the absence of post-treatment confounders, as it focuses on the potential outcomes of S, which can be treated as baseline covariates. However, due to unobserved potential outcomes, assumptions must be made to aid in the estimation

of unidentified parameters. If baseline covariate information is available, this may aid in the estimation of the unobserved principal strata of S. Baseline covariate information can also be used within the CE framework to relax assumptions about post-treatment confounding. However, estimation methods in this case require the presence of a baseline covariate that has an interaction effect with Z on S [1, 38], and we have shown in Appendix C that when such interactions exist, S will not be a valid surrogate within the CA framework.

The CE and CA frameworks have tradeoffs in terms of assumptions, bias in parameter estimation and variability [39]. Estimation methods within the CA framework have been shown to have less bias, but more variability than standard methods within the CE framework [40]. As the parameters of the proposed structural model have a direct mapping to the parameters of the CA model, these models offer the potential for assumptions that are reasonable to make in one framework to aid in informing the parameter values within the alternative framework. In this way, both the CE and CA models could be employed with reasonable, but not especially strong assumptions made in the evaluation of S as a surrogate marker. While the research in this paper has focused on the situation of surrogate markers, the frameworks of CA and CE have also been considered in mediation analysis. It would be of interest to evaluate the correspondence between the metrics of mediation in this setting too. We have focused on Gaussian variables and linear models and have shown a certain degree of correspondence for evaluating the surrogates, and the degree of correspondence increasing if certain assumptions are made. For non-Gaussian variables we hypothesize we might expect broadly similar findings, but with possibly a lower degree of correspondence, due to the nonlinear link functions in the models.

10. Acknowledgements

This research was supported by NIH grants CA129102 and CA083654.

A. Relationship of CE and CA model parameters

The parameters of the principal surrogacy model (Equation 3) relate to those of the assumed structural model (Equations 1 and 2) assuming $U \sim N(0,1)$ in the following way:

$$\mu_{S_0} = \alpha_0$$
$$\mu_{S_1} = \alpha_0 + \alpha_1$$
$$\mu_{T_0} = \beta_0 + \beta_2\alpha_0$$
$$\mu_{T_1} = \beta_0 + \beta_1 + (\beta_2 + \beta_4)(\alpha_0 + \alpha_1)$$
$$\sigma^2_{S_0} = \alpha_2^2 + \delta^2_{S0}$$
$$\sigma^2_{S_1} = (\alpha_2 + \alpha_3)^2 + \delta^2_{S1}$$
$$\sigma^2_{T_0} = (\beta_2\alpha_2 + \beta_3)^2 + \beta_2^2\delta^2_{S0} + \delta^2_{T0}$$
$$\sigma^2_{T_1} = [(\beta_2 + \beta_4)(\alpha_2 + \alpha_3) + (\beta_3 + \beta_5)]^2 + (\beta_2 + \beta_4)^2\delta^2_{S1} + \delta^2_{T1}$$
$$\rho_s = \frac{\alpha_2(\alpha_2+\alpha_3)}{\sqrt{\alpha_2^2+\delta^2_{S0}}\sqrt{(\alpha_2+\alpha_3)^2+\delta^2_{S1}}}$$
$$\rho_{00} = \frac{\alpha_2(\beta_2\alpha_2+\beta_3)+\beta_2\delta^2_{S0}}{\sqrt{\alpha_2^2+\delta^2_{S0}}\sqrt{(\beta_2\alpha_2+\beta_3)^2+\beta_2^2\delta^2_{S0}+\delta^2_{T0}}}$$
$$\rho_{01} = \frac{\alpha_2[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]}{\sqrt{\alpha_2^2+\delta^2_{S0}}\sqrt{[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]^2+(\beta_2+\beta_4)^2\delta^2_{S1}+\delta^2_{T1}}}$$
$$\rho_{10} = \frac{(\alpha_2+\alpha_3)(\beta_2\alpha_2+\beta_3)}{\sqrt{(\alpha_2+\alpha_3)^2+\delta^2_{S1}}\sqrt{(\beta_2\alpha_2+\beta_3)^2+\beta_2^2\delta^2_{S0}+\delta^2_{T0}}}$$
$$\rho_{11} = \frac{(\alpha_2+\alpha_3)[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]+(\beta_2+\beta_4)\delta^2_{S1}}{\sqrt{(\alpha_2+\alpha_3)^2+\delta^2_{S1}}\sqrt{[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]^2+(\beta_2+\beta_4)^2\delta^2_{S1}+\delta^2_{T1}}}$$
$$\rho_t = \frac{(\beta_2\alpha_2+\beta_3)[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]}{\sqrt{(\beta_2\alpha_2+\beta_3)^2+\beta_2^2\delta^2_{S0}+\delta^2_{T0}}\sqrt{[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]^2+(\beta_2+\beta_4)^2\delta^2_{S1}+\delta^2_{T1}}}.$$

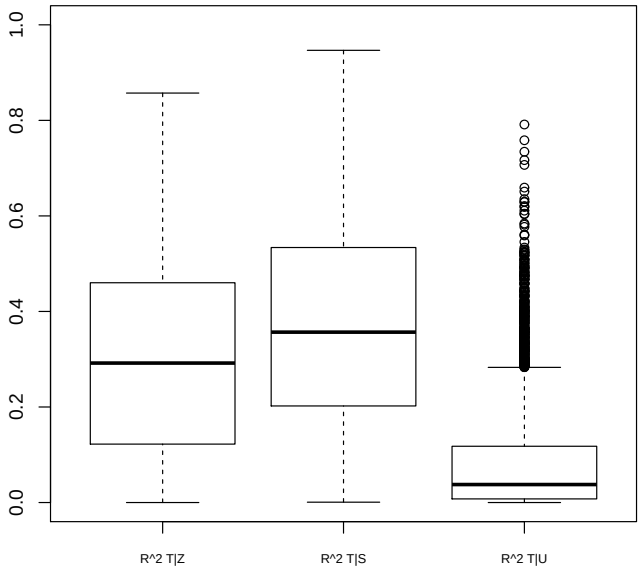
B. Parameter distributions and R^2 values for regression models of $T|Z$, $T|S$ and $T|U$ from simulation experiment

For the simulation experiment, the CE model parameters were generated such that
$\alpha_0, \beta_0 = 0$,
$\alpha_1 \sim U(0.25, 1.5)$,
$\alpha_2 \sim U(max(-0.5, min(-\alpha_1 + 0.05, \alpha_1 - 0.05)), max(-\alpha_1 + 0.05, \alpha_1 - 0.05))$,
$\alpha_3 \sim U(min(-\alpha_2/2, \alpha_2/2), max(-\alpha_2/2, \alpha_2/2))$,

$\beta_1 \sim U(-0.3, 1.5)$,
$\beta_2 \sim U(0.1, 1.5)$,
$\beta_3 \sim U(max(-0.5, min(-\beta_2 + 0.05, \beta_2 - 0.05)), max(-\beta_2 + 0.05, \beta_2 - 0.05))$,
$\beta_4 \sim U(min(-\beta_2/3, \beta_2/3), max(-\beta_2/3, \beta_2/3))$,
$\beta_5 \sim U(min(-\beta_3/2, \beta_3/2), max(-\beta_3/2, \beta_3/2))$,
$\delta^2_{S0}, \delta^2_{T0} \sim U(0.3, 1)$,
$\delta^2_{S1} = \delta^2_{S0}$,
$\delta^2_{T1} = \delta^2_{T0}$.

Figure B.1 below provides boxplots of the R^2 values across all of the parameter draws for regression models of $T|Z$, $T|S$ and $T|U$, where S and T are 10,000 random samples from each set of parameter draws. The plots show that the simulated parameters lead to a broad range of R^2 values, indicating that the way in which the parameters were simulated was not overly restrictive and leads to a wide spectrum of scenarios.

Figure B.1. Proportion of variance of T explained by Z, S and U for a range of plausible values from the assumed structural model.



C. Consequence of conditional independence assumptions in the CA framework on parameters in the CE model

In terms of the structural model, the assumptions that $S(0)$ and $T(1)$ are conditionally independent given $S(1)$ and that $S(1)$ and $T(0)$ are conditionally independent given $S(0)$ requires the following

$$\frac{(\beta_2\alpha_2+\beta_3)(\alpha_2+\alpha_3)(\alpha_2^2+\delta^2_{S0})-\alpha_2^2(\alpha_2+\alpha_3)^2-(\alpha_2(\beta_2\alpha_2+\beta_3)+\beta_2\delta^2_{S0})^2}{\alpha_2^2+\delta^2_{S0}} =$$
$$\frac{\alpha_2[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]((\alpha_2+\alpha_3)^2+\delta^2_{S1})-\alpha_2^2(\alpha_2+\alpha_3)^2-([(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)](\alpha_2+\alpha_3)+(\beta_2+\beta_4)\delta^2_{S1})^2}{(\alpha_2+\alpha_3)^2+\delta^2_{S1}} = 0.$$

D. Baseline covariates

Under the structural model with covariates, the direct and indirect effects of the CE framework become: $E[NDE(0)|X = x] = \beta_1 + \alpha_0\beta_4 + (\beta_4\psi_1 + \omega_2)x$, $E[NDE(1)|X = x] = \beta_1 + \beta_4(\alpha_0 + \alpha_1) + (\beta_4(\psi_1 + \psi_2) + \omega_2)\,x$, $E[NIE(0)|X = x] = \beta_2(\alpha_1 + \psi_2 x)$, $E[NIE(1)|X = x] = (\beta_2 + \beta_4)(\alpha_1 + \psi_2 x)$, and $E[TE|X = x] = (\beta_2 + \beta_4)\alpha_1 +$

$\alpha_0\beta_4 + \beta_1 + ((\beta_2 + \beta_4)\psi_2 + \beta_4\psi_1 + \omega_2)\,x$. If there is no interaction effect of X and Z on S ($\psi_2 = 0$), then the indirect effect will not be changed by the presence of baseline covariates. Under the assumption of sequential ignorability, estimates of the direct and indirect effects in the presence of a baseline covariate can be obtained non-parametrically by integrating over the distribution of X. The Prentice model in the presence of baseline confounders becomes: $E[T|S,Z] = \theta_0 + \theta_1 Z + \theta_2 S + \theta_3 SZ + \theta_4 X + \theta_5 XZ$, where θ_1 and θ_3 are as in Section 3.2 and $\theta_4 = \omega_1 - \left(\frac{\alpha_2(\beta_2\alpha_2+\beta_3)+\beta_2\delta^2_{S0}}{\alpha_2^2+\delta^2_{S0}}\right)\psi_1$ and $\theta_5 = \omega_2 + \left(\frac{\alpha_2(\beta_2\alpha_2+\beta_3)+\beta_2\delta^2_{S0}}{\alpha_2^2+\delta^2_{S0}}\right)\psi_1 - \left(\frac{(\alpha_2+\alpha_3)[(\beta_2+\beta_4)(\alpha_2+\alpha_3)+(\beta_3+\beta_5)]+(\beta_2+\beta_4)\delta^2_{S1}}{(\alpha_2+\alpha_3)^2+\delta^2_{S1}}\right)(\psi_1+\psi_2)$. Therefore, if there are no unmeasured confounders, the Prentice criteria will be a valid measure of surrogacy ($\theta_1 = \theta_3 = \theta_5 = 0$) if there is no interaction effect of X and Z on either S or T ($\omega_2 = \psi_2 = 0$) and additionally if either ψ_1 or β_4 is zero. In this case, the baseline covariate information aids in estimation and does not affect the ability to determine surrogacy. Under certain conditions, it is possible to relax the sequential ignorability assumption when baseline covariates are available. For example, when sequential ignorability cannot be assumed and baseline covariates are available for which $E[S(1) \mid X] - E[S(0) \mid X]$ varies with X (i.e. $\psi_2 \neq 0$), and there is no interaction effect of either Z and S on T or of Z and X on T (i.e. $\beta_4 = \omega_2 = 0$), Joffe and Greene [1] showed that a two-stage least squares procedure can be used to estimate the direct and indirect effects. Ten Have *et al.* [38] estimate the direct and indirect effects under the same conditions as Joffe and Greene [1] by assuming the following rank preserving model for T: $T(z,s) = g(\mathbf{x}) + \gamma_Z z + \gamma_S s + \epsilon$ and using a G-estimation procedure. Within the CA framework, if baseline covariates are present, the surrogacy quantities of interest become: $E[T(1) - T(0)|S(1) - S(0) = s, X = x] = \quad (\mu_{T_1} - \mu_{T_0}) - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)(\mu_{S_1} - \mu_{S_0}) +$ $\left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)s + \left(\omega_2 - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)\psi_2\right)x \quad =$ $\gamma_0 + \gamma_1 s + \left(\omega_2 - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)\psi_2\right)x$. When there is no interaction of X and Z on either S or T ($\psi_2 = \omega_2 = 0$), neither γ_0 nor γ_1 are affected by the presence of baseline covariates. In this case, controlling for X is helpful in explaining some of the variance of the potential outcomes and does not affect the ability to estimate γ_0 or γ_1. When there is an interaction of X and Z on either S or T, γ_1 is not affected but γ_0 becomes a function of x: $\gamma_0 = (\mu_{T_1} - \mu_{T_0}) - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)(\mu_{S_1} - \mu_{S_0}) +$ $\left(\omega_2 - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)\psi_2\right)x$. In this case, when there is no treatment effect on S, the expected treatment effect on T depends on the baseline covariate, implying that S may only be a valid principal surrogate for T within certain subgroups defined by X. In order for ACN to be met and S to be considered a valid principal surrogate, we would need to have $\gamma_0 = 0$ for all X, requiring $\int_x E[T(1) - T(0) \mid S(1) - S(0) = 0, X = x] f(X \mid S(1) - S(0) = 0) dx$ be equal to zero, which is unlikely to hold. Therefore, in order for S to be considered a valid principal surrogate, there can be no interaction of the baseline covariate X with Z, so that both ψ_2 and ω_2 are equal to zero. In the CA framework, baseline covariates have also been used to aid in estimating the principal strata of S. For example, a model for $f(S(1) \mid X, Z = 1)$ can be estimated using the surrogate response values in the $Z = 1$ arm and a model for $f(S(0) \mid X, Z = 0)$ can be estimated using the surrogate response values in the $Z = 0$ arm. These models can then be used to impute missing $S(1)$ values in patients in the $Z = 0$ arm and missing $S(0)$ values in patients in the $Z = 1$ arm, respectively [15,16]. Implicit in this assumption is that $[S(1) \mid X, S(0)] = [S(1) \mid X]$, requiring that $\rho_s = 0$, i.e. that either $\alpha_2 = 0$ or $\alpha_2 + \alpha_3 = 0$.

References

[1] Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics* 2009; **65**: 530–538.

[2] Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 1989; **8**: 431–440.

[3] Freedman L, Graubard B, Schatzkin A. Statistical validation of intermediate endpoints for chronic disease. *Statistics in Medicine* 1992; **11**: 167–178.

[4] Buyse M, Molenberghs G, Burzykowski D, *et al.* The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000; **1:** 49–67.

[5] Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: Are we being misled. *Annals of Internal Medicine* 1996; **125:** 605–613.

[6] VanderWeele TJ. Principal stratification–uses and limitations. *The International Journal of Biostatistics* 2011; **Vol. 7:Iss. 1**, Article 28.

[7] Ten Have TR, Joffe MM. A review of causal estimation of effects in mediation analyses. *Statistical Methods in Medical Research* 2010; **21**: 77–107.

[8] Ensor H, Lee RJ, Sudlow C, Weir CJ. Statistical approaches for evaluating surrogate outcomes in clinical trials: a systematic review. *Journal of Biopharmaceutical Statistics* 2015; DOI: 10.1080/10543406.2015.1094811

[9] Baron RM, Kenny DA. The moderator mediator variable distinction in social psychological-research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 1986; **51:** 1173–1182.

[10] Robins J. Correcting for non-compliance in randomised trials using structural nested mean models. *Communications in Statistics-Theory and Methods* 1994; **23**: 2379–2412.

[11] Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 2010; **25:** 51–71.

[12] Daniels MJ, Roy JA, Kim C, *et al.* Bayesian inference for the causal effect of mediation. *Biometrics* 2012; **68:** 1028–1036.

[13] Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**: 21–29.

[14] Conlon ASC, Taylor JMG, Elliott MR. Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal. *Biostatistics* 2013; **15:** 266–283.

[15] Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. *Biometrics* 2008; **64:** 1146–1154.

[16] Zigler CM, Belin TR. A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics* 2012; **68**: 922–932.

[17] Parast L, McDermott MM, Tian L. Robust estimation of the proportion of treatment effect explained by surrogate marker information. *Statistics in Medicine* 2015; DOI: 10.1002/sim.6820.

[18] Li Y, Taylor JMG, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics* 2010; **66**: 523–531.

[19] Pearl J. Principal stratification–a goal or a tool?. *The International Journal of Biostatistics* 2011; **Vol. 7:Iss. 1**, Article 20.

[20] Joffe M. Principal stratification and attribution prohibition: Good ideas taken too far. *The International Journal of Biostatistics* 2011; **Vol. 7:Iss. 1**, Article 35.

[21] VanderWeele TJ. Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters* 2008; **78**: 2957–2962.

[22] Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**: 143–155.

[23] Pearl J. Direct and indirect effects. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* 2001; pp. 411–420. San Francisco, CA: Morgan Kaufman.

[24] Wang Y, Taylor JMG. A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics* 2002; **58**: 803–812.

[25] Peterson ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology* 2006; **17(3)**: 276–284.

[26] VanderWeele TJ. A unification of mediation and interaction: a 4-way decomposition. *Epidemiology* 2014; **25**: 749–761.

[27] Qin L, Gilbert PB, Follmann D, *et al.* Assessing surrogate endpoints in vaccine trials with case-cohort sampling and the Cox model. *The Annals of Applied Statistics* 2008; **2(1)**: 386–407.

[28] Schwartz SL, Li F, Mealli F. A Bayesian semiparametric approach to intermediate variables in causal inference. *Journal of the American Statistical Association* 2011; **106**: 1331–1344.

[29] Bartolucci F, Grilli L. Modeling partial compliance through copulas in a principal stratification framework. *Journal of the American Statistical Association* 2011; **106:** 469–479.

[30] Parast L, Cai T, Tian L. Nonparametric Estimation of the Proportion of Treatment Effect Explained by a Surrogate Marker using Censored Data. *Technical Report* 2016.

[31] Angrist JD, Imbens GW, Rubin DB. Identifiability of path-specific effects. In *Proceedings of the International Joint Conference on Artificial Intelligence.* 2005; Morgan Kaufman, San Francisco, CA. MR2192340

[32] Holland PW. Causal inference, path analysis and recursive structural equation models (with discussion). In C. C. Clogg (Ed.), *Sociological methodology* 1988; (pp. 449-493). Washington, DC: American Sociological Association.

[33] Sobel ME. Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics* 2008; **Vol. 33, No. 2**: 230–251.

[34] Gustafson P. Bayesian inference for partially identified models. *The International Journal of Biostatistics* 2010;**Vol. 6: Iss. 2**, Article 17.

[35] Vansteelandt S. Estimation of direct and indirect effects. In *Causality: Statistical Perspectives and Applications* 2012;(pp. 126-150). Hoboken, NJ: Wiley.

[36] Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for PCP in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 1994; **89**: 737–749.

[37] VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* 2010; **21**: 540–551.

[38] VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. *Oxford University Press* 2015.

[39] Ten Have T, Joffe M, Lynch K, Maisto S, Brown G, Beck A. Causal mediation analyses with rank preserving models. *Biometrics* 2007; **63**: 926–934.

[40] Gallop R, Small D, Lin J, Elliott M, Joffe M, Ten Have T. Mediation analysis with principal stratification. *Statistics in Medicine* 2009; **28**: 1108–1130.

Table 1. Expressions for direct, indirect and average causal effects

$E[NDE(0)]$:	$E[T(1,S(0)) - T(0,S(0))] = \beta_1 + \alpha_0\beta_4$
$E[NDE(1)]$:	$E[T(1,S(1)) - T(0,S(1))] = \beta_1 + \beta_4(\alpha_0 + \alpha_1)$
$E[NIE(0)]$:	$E[T(0,S(1)) - T(0,S(0))] = \alpha_1\beta_2$
$E[NIE(1)]$:	$E[T(1,S(1)) - T(1,S(0))] = \alpha_1(\beta_2 + \beta_4)$
$E[TE]$:	$E[T(1,S(1)) - T(0,S(0))] = \beta_1 + \beta_2\alpha_1 + \beta_4(\alpha_0 + \alpha_1)$

Table 2. Consequence of the sequential ignorability and no interaction assumptions on parameters within the CA framework.

	Assumption	
	1. $\alpha_2 = \alpha_3 = 0$ No unmeasured confounders for S	2. $\beta_3 = \beta_5 = 0$ No unmeasured confounders for T
ρ_s	0	$\frac{\alpha_2(\alpha_2+\alpha_3)}{\sqrt{(\alpha_2^2+\delta_{S0}^2)((\alpha_2+\alpha_3)^2+\delta_{S1}^2)}}$
ρ_{00}	$\frac{\beta_2\delta_{S0}^2}{\sqrt{\delta_{S0}^2(\beta_3^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)}}$	$\frac{\beta_2(\alpha_2^2+\delta_{S0}^2)}{\sqrt{(\alpha_2^2+\delta_{S0}^2)((\beta_2\alpha_2)^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)}}$
ρ_{01}	0	$\frac{\alpha_2(\beta_2+\beta_4)(\alpha_2+\alpha_3)}{\sqrt{(\alpha_2^2+\delta_{S0}^2)([(\beta_2+\beta_4)(\alpha_2+\alpha_3)]^2+(\beta_2+\beta_4)^2\delta_{S1}^2+\delta_{T1}^2)}}$
ρ_{10}	0	$\frac{(\alpha_2+\alpha_3)\beta_2\alpha_2}{\sqrt{((\alpha_2+\alpha_3)^2+\delta_{S1}^2)((\beta_2\alpha_2)^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)}}$
ρ_{11}	$\frac{(\beta_2+\beta_4)\delta_{S1}^2}{\sqrt{\delta_{S1}^2((\beta_3+\beta_5)^2+(\beta_2+\beta_4)^2\delta_{S1}^2+\delta_{T1}^2)}}$	$\frac{(\beta_2+\beta_4)((\alpha_2+\alpha_3)^2+\delta_{S1}^2)}{\sqrt{((\alpha_2+\alpha_3)^2+\delta_{S1}^2)([(\beta_2+\beta_4)(\alpha_2+\alpha_3)]^2+(\beta_2+\beta_4)^2\delta_{S1}^2+\delta_{T1}^2)}}$
ρ_t	$\frac{\beta_3(\beta_3+\beta_5)}{\sqrt{(\beta_3^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)((\beta_3+\beta_5)^2+(\beta_2+\beta_4)^2\delta_{S1}^2+\delta_{T1}^2)}}$	$\frac{\beta_2\alpha_2(\beta_2+\beta_4)(\alpha_2+\alpha_3)}{\sqrt{((\beta_2\alpha_2)^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)([(\beta_2+\beta_4)(\alpha_2+\alpha_3)]^2+(\beta_2+\beta_4)^2\delta_{S1}^2+\delta_{T1}^2)}}$
γ_0	$(\beta_1 + \beta_4(\alpha_0 + \alpha_1)) - \alpha_1\beta_4\left(\frac{\delta_{S1}^2}{\delta_{S0}^2+\delta_{S1}^2}\right)$	$(\beta_1 + \beta_4(\alpha_0 + \alpha_1)) - \alpha_1\beta_4\left(\frac{\alpha_3^2+\alpha_3\alpha_2+\delta_{S1}^2}{\alpha_3^2+\delta_{S0}^2+\delta_{S1}^2}\right)$
γ_1	$\beta_2 + \beta_4\left(\frac{\delta_{S1}^2}{\delta_{S0}^2+\delta_{S1}^2}\right)$	$\beta_2 + \beta_4\left(\frac{\alpha_3^2+\alpha_3\alpha_2+\delta_{S1}^2}{\alpha_3^2+\delta_{S0}^2+\delta_{S1}^2}\right)$
	Assumption	
	3. $\beta_4 = \alpha_2 = \alpha_3 = 0$ No interaction, no unmeasured confounders for S	4. $\beta_4 = \beta_3 = \beta_5 = 0$ No interaction, no unmeasured confounders for T
ρ_s	0	$\frac{\alpha_2(\alpha_2+\alpha_3)}{\sqrt{(\alpha_2^2+\delta_{S0}^2)((\alpha_2+\alpha_3)^2+\delta_{S1}^2)}}$
ρ_{00}	$\frac{\beta_2\delta_{S0}^2}{\sqrt{\delta_{S0}^2(\beta_3^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)}}$	$\frac{\beta_2(\alpha_2^2+\delta_{S0}^2)}{\sqrt{(\alpha_2^2+\delta_{S0}^2)((\beta_2\alpha_2)^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)}}$
ρ_{01}	0	$\frac{\alpha_2\beta_2(\alpha_2+\alpha_3)}{\sqrt{(\alpha_2^2+\delta_{S0}^2)([\beta_2(\alpha_2+\alpha_3)]^2+\beta_2^2\delta_{S1}^2+\delta_{T1}^2)}}$
ρ_{10}	0	$\frac{(\alpha_2+\alpha_3)\beta_2\alpha_2}{\sqrt{((\alpha_2+\alpha_3)^2+\delta_{S1}^2)((\beta_2\alpha_2)^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)}}$
ρ_{11}	$\frac{\beta_2\delta_{S1}^2}{\sqrt{\delta_{S1}^2((\beta_3+\beta_5)^2+\beta_2^2\delta_{S1}^2+\delta_{T1}^2)}}$	$\frac{\beta_2((\alpha_2+\alpha_3)^2+\delta_{S1}^2)}{\sqrt{((\alpha_2+\alpha_3)^2+\delta_{S1}^2)([\beta_2(\alpha_2+\alpha_3)]^2+\beta_2^2\delta_{S1}^2+\delta_{T1}^2)}}$
ρ_t	$\frac{\beta_3(\beta_3+\beta_5)}{\sqrt{(\beta_3^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)((\beta_3+\beta_5)^2+\beta_2^2\delta_{S1}^2+\delta_{T1}^2)}}$	$\frac{\alpha_2\beta_2^2(\alpha_2+\alpha_3)}{\sqrt{((\beta_2\alpha_2)^2+\beta_2^2\delta_{S0}^2+\delta_{T0}^2)([\beta_2(\alpha_2+\alpha_3)]^2+\beta_2^2\delta_{S1}^2+\delta_{T1}^2)}}$
γ_0	β_1	β_1
γ_1	β_2	β_2

Figure 1. Causal graph for the intervention (Z), the surrogate (S) and the final outcome (T) with an unmeasured confounder (U).

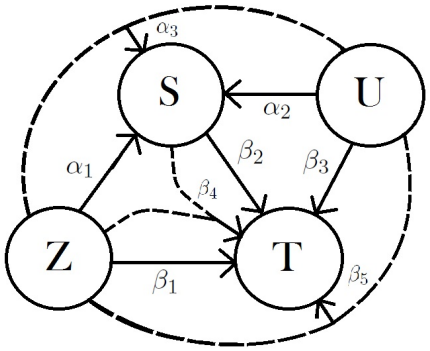
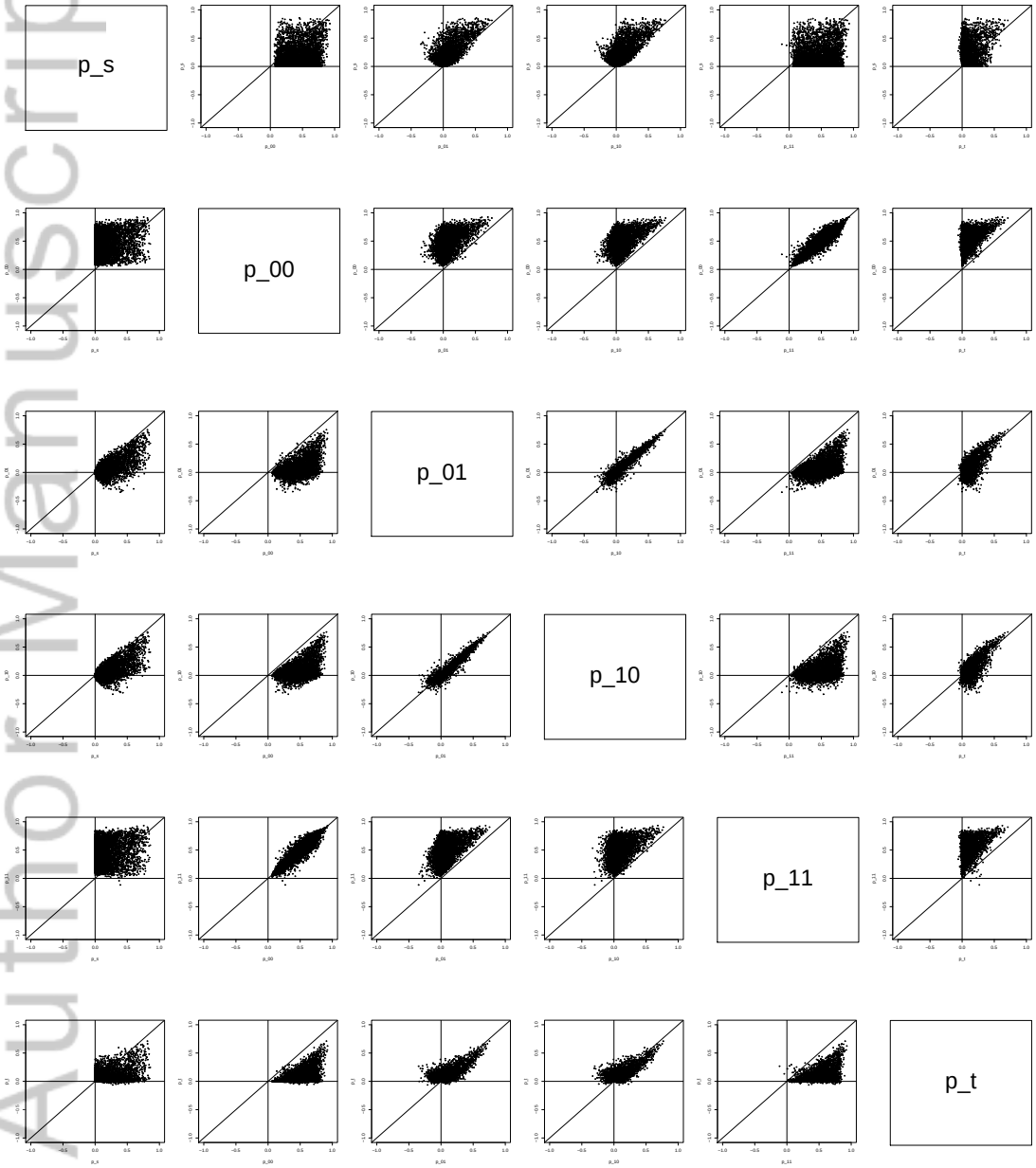


Figure 2. CA model parameters for a range of plausible values from the assumed structural model.



Prepared using simauth.cls

Figure 3. Correspondence between NDE and γ_0 for a range of plausible values from the assumed structural model.

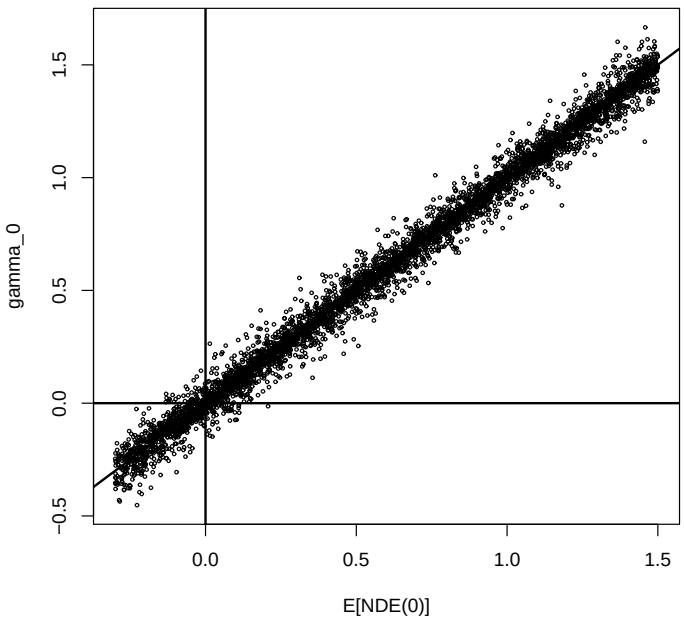
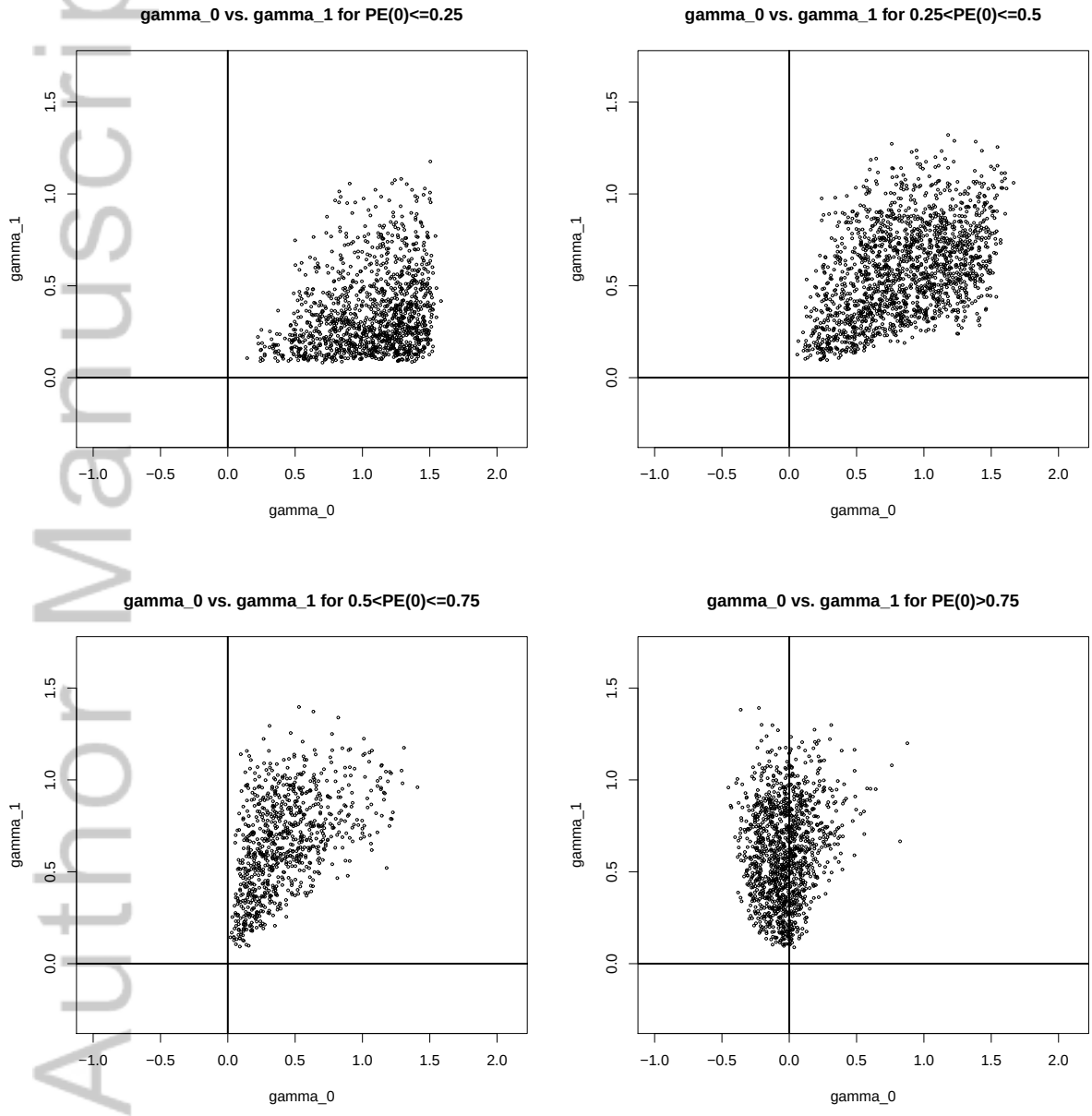


Figure 4. Correspondence between CA surrogacy measures (γ_0 and γ_1) and CE surrogacy measures (PE(0)) for a range of plausible values from the assumed structural model.

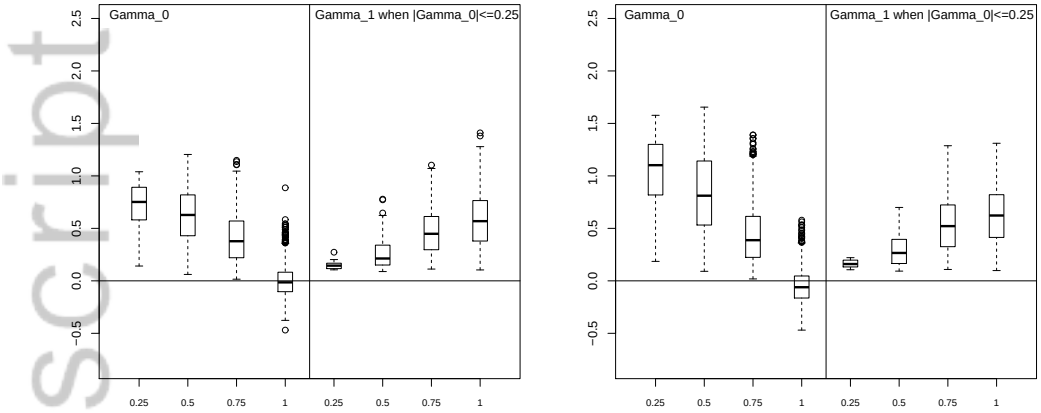


Prepared using simauth.cls

Figure 5. Correspondence between CA surrogacy measures (γ_0 and γ_1 shown on vertical axis) and CE surrogacy measure (PE(0) shown on horizontal axis) under assumptions made in the CA framework

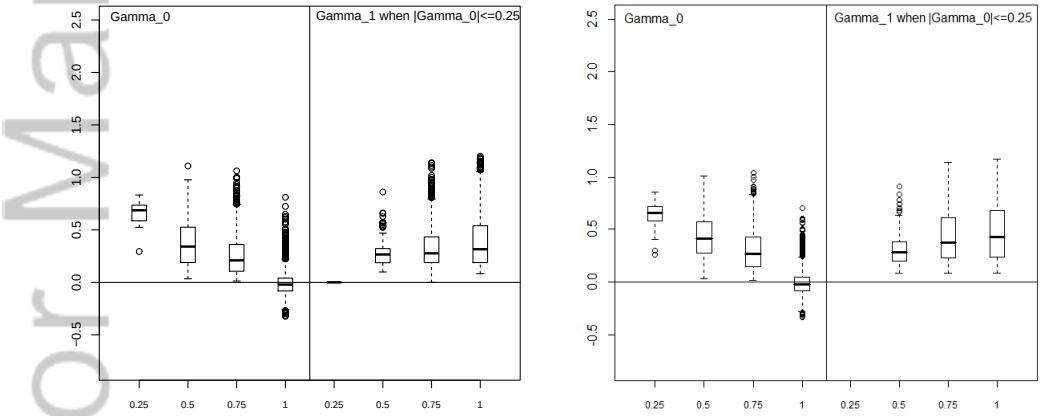
(a) No assumptions

(b) $(\rho_s, \rho_t, \rho_{00}, \rho_{11}) > 0$, and $(\rho_{01}, \rho_{01}) < min(\rho_s, \rho_t, \rho_{00}, \rho_{11})$



(c) $T(0) \perp T(1) \mid S(0), S(1)$

(d) $T(1) \perp S(0) \mid S(1)$, and $T(0) \perp S(1) \mid S(0)$



(e) $T(0) \perp S(1) \mid S(0)$, and $T(1) \perp S(0) \mid S(1)$

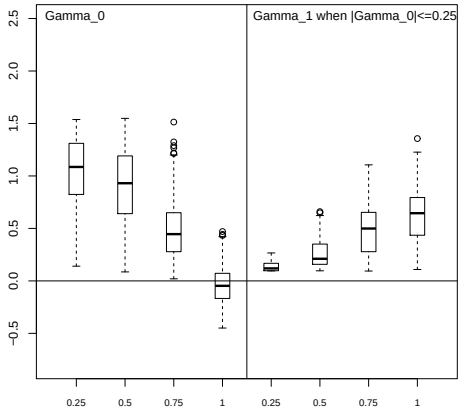
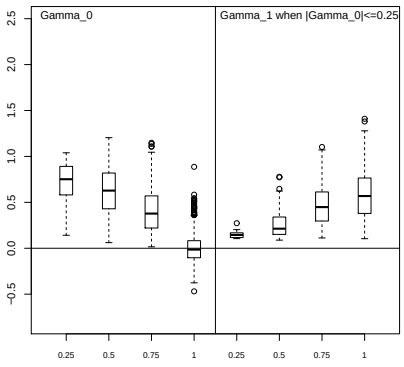
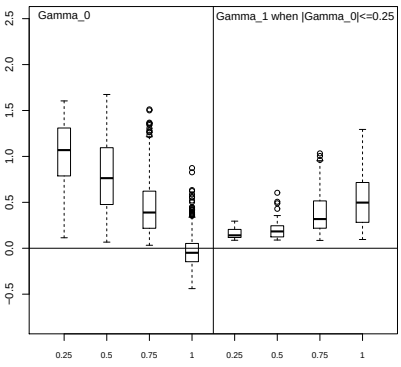


Figure 6. Correspondence between CA surrogacy measures (γ_0 and γ_1) and CE surrogacy measure (PE(0)) under assumptions made in the CE framework

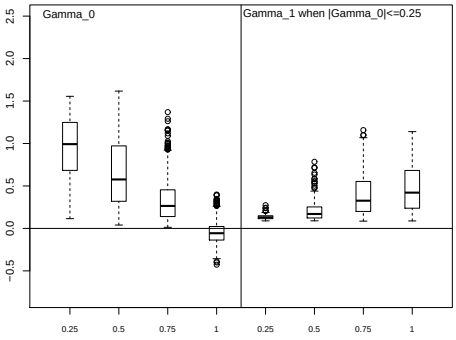
(a) No assumptions



(b) $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$



(c) $\alpha_2 = \alpha_3 = \beta_3 = \beta_5 = 0$



(d) $\alpha_2 = \alpha_3 = 0$ or $\beta_3 = \beta_5 = 0$ and $\beta_4 = 0$