

**University of Colorado, Denver**

---

**From the Selected Works of Debashis Ghosh**

---

Fall 2011

# A causal framework for surrogate endpoints with semi-competing risks data

Debashis Ghosh, *Penn State University*



Available at: [https://works.bepress.com/debashis\\_ghosh/51/](https://works.bepress.com/debashis_ghosh/51/)

# A causal framework for surrogate endpoints with semi-competing risks data

Debashis Ghosh

Departments of Statistics and Public Health Sciences

The Pennsylvania State University

514A Wartik Building

University Park, PA,16802 U.S.A.

ghoshd@psu.edu

## Summary

In this note, we address the problem of surrogacy using a causal modelling framework that differs substantially from the potential outcomes model that pervades the biostatistical literature. The framework comes from econometrics and conceptualizes direct effects of the surrogate endpoint on the true endpoint. While this framework can incorporate the so-called semi-competing risks data structure, we also derive a fundamental non-identifiability result. Relationships to existing causal modelling frameworks are also discussed.

*Key words:* Clinical Trial; Counterfactual; Dependence; Nonlinear response; Prentice Criterion; Rubin causal model.

# 1 Introduction

Surrogate markers are measurements that are proposed based on substantive biological knowledge about a disease and will generally require fewer subjects or can be collected in a shorter time window than most response variables in clinical trials (e.g., time to death). Prentice (1989) proposed a set of four criteria for a surrogate endpoint to be valid as a proxy for a true endpoint. The utility of the Prentice criteria have been debated and criticized heavily (e.g., Begg and Leung, 2000). Frameworks based on estimation have been given by Freedman et al. (1992) and Buyse and Molenberghs (1998).

More recently, authors have begun to study the problem of surrogacy assessment using ideas from causal inference. Gilbert and Hudgens (2008) proposed some estimation and inference procedures for quantifying surrogacy within the principal stratification framework initially proposed by Frangakis and Rubin (2002). Taylor et al. (2005) identified conditions that yield causal equivalences with the analysis approach of Friedman et al. (1992). A more general framework that synthesizes and compares various existing surrogacy measures in the single- and multiple-trial setting was given by Joffe and Greene (2008). More recently, various measures for quantifying surrogacy within the intermediate variable frameworks of Rubins and Greenland and Robins (1992) have been developed by Li et al. (2010). In addition, surrogacy with time-to-event for both endpoints (surrogate and true) has been explored by Ghosh (2009) and Ghosh et al. (2011). In that work, it was seen that the usual approaches for causal modelling from the biostatistical literature were not directly usable with semi-competing risks data; this argument is discussed in more detail in §2.3.

A central issue in the assessment of surrogacy that has not been explicitly considered in the literature involves the *timing* of the surrogate endpoint. Gilbert and Hudgens (2008) point out that the appropriate conditioning event for their causal measures must involve the surrogate marker being measured after randomization. For the principal stratification advocated by Frangakis and Rubin (2002), the principal strata are defined by counterfactual values of the surrogate endpoints so that the counterfactual values of the surrogates can effectively be treated as baseline (i.e., prerandomization) covariates.

The goal in our work here is to formalize the effects of timing of the surrogate endpoint event. This will be done within a proposed causal modelling framework. Doing so requires the framework related to that in Pearl (2009); in particular, the potential outcomes framework of Holland (1986) is not usable here. Taking such a view allows us to explicitly account for the timing of the surrogate endpoint as occurring before the true endpoint, which is the condition of practical interest (Ghosh, 2009; Ghosh et al., 2011). By utilizing this framework, we will come to the following conclusions:

1. The proposed causal modelling framework is compatible with a semi-competing risks data structure (Fine et al., 2001).
2. There is a fundamental nonidentifiability of subject heterogeneity with the causal effect of the surrogate endpoint on the true endpoint.

The framework we describe here have been previously proposed in econometrics (Abbring and van der Berg, 2003), but their application has not been brought to bear on the surrogate endpoint problem.

## 2 Methodology

### 2.1 Causal framework and nonidentification result

Define  $S$  to be the time to the surrogate endpoint,  $T$  to be the time to the true endpoint, and let  $Z$  be a binary treatment indicator taking values 0 for control and 1 for treatment.

Conceptually, we operationalize the surrogate endpoint as a type of “intervention” that is applied to the individual. After the intervention (i.e., surrogate endpoint) occurs, the effect is to modify the distribution of the time to the true endpoint.

One could also formalize the data using counting process notation based on  $N_S(y) = I(S \leq y)$ . The counting process takes values zero before the occurrence of the surrogate endpoint and one afterwards. The idealized data that the investigator would see for each subject is  $\{T, N_S(y) : 0 \leq y \leq T\}$ . This could also be represented as  $[\{N_T(t) : 0 \leq t \leq \infty\}, \{N_S(y) : 0 \leq y \leq T\}]$ . Note that we are not interested in the situation where the surrogate endpoint occurs after the true endpoint. This is addressed by researchers working

in semi-competing risks settings through the so-called “wedge constraint” (Fine et al., 2001; Ghosh, 2009).

**Remark.** While the paper makes  $S$  and  $T$  to be explicitly times to events, most surrogate endpoints involve time implicitly in the fact that the surrogate measurement is recorded after the treatment has been assigned. This was alluded to in the work of Gilbert and Hudgens (2008). For example, let us consider tumor shrinkage 6 months after treatment as a surrogate for overall survival in a colorectal cancer setting. Typically, it is treated as a binary endpoint based on whether or not the tumor shrinks by a certain percentage (e.g., 25%). Using the notation of the paper here, one could define  $S$  to be the time until the tumor shrinks by 25% and  $T$  to be time to death. Because most clinical decision rules are binary, this logic should apply to a variety of surrogate endpoints. There are two potential complications. First, in the tumor shrinkage example, there will be a positive probability mass at 6 months. Second, the censoring mechanism might become more complicated; in particular,  $S$  might have to be considered as interval-censored data.

Our notion of a counterfactual quantity within this framework differs substantially from that posed in the Rubin causal model (Holland, 1986). In particular, the causal model that is used here corresponds to the following thought experiment:

1. Nature draws from an element  $\omega$  from the filtration  $(\Omega, \mathcal{F}, \mathcal{P})$ , where  $\mathcal{F}$  denotes a  $\sigma$ -algebra defined on  $\Omega$ , and  $\mathcal{P}$  is a probability measure;
2. We manipulate the surrogate endpoint to occur at time  $s$ .
3. Generate an exponential random variable  $E \equiv E(s, \omega)$ .
4. Define the failure time as

$$T^*(s, \omega) = \Lambda^{-1}(E),$$

where  $\Lambda(t)$  is the cumulative hazard function of  $T^*(s, \cdot)$ .

Causal inferential comparisons are then made based on comparison of  $T^*(s_1, \cdot)$  and  $T^*(s_2, \cdot)$ .

Examples of causal contrasts are  $E[T^*(s_1, \Omega)] - E[T^*(s_2, \Omega)]$  and

$$\frac{\lambda_{T^*(s_1)}(y)}{\lambda_{T^*(s_2)}(y)}.$$

The former contrast is a difference in means with two values of the surrogate, while the latter is a hazard ratio.

Fundamentally, this type of causal model is quite different from that posed in the work of Rubin (1974) and Holland (1986). The only randomness that exists in their model arises from the fact that only one of the potential outcomes is observed in practice. Our framework allows for randomness as seen in step 3 of the thought experiment. By contrast, the Rubin causal outcome specifies the exact values of the random variables under the various counterfactual scenarios so that it is more deterministic relative to this framework used here. We assume measurability throughout so that random variables are well-defined. While the Rubin causal model has been used extensively in the biostatistical literature, the conceptual model proposed here has been advocated in the linear model case by Freedman (2004).

A key assumption for performing causal inference with continuous time-value treatments is the so-called “no-anticipation” or “consistency” assumption (Abbring and van der Berg, 2003; Lok, 2008):

*Assumption 1:* For all  $u, v \in [0, \infty)$ ,

$$\Lambda_{T^*(u)}(t) = \Lambda_{T^*(v)}(t) \quad \forall t \leq u \wedge v,$$

where  $a \wedge b$  denotes the minimum of  $a$  and  $b$ , and  $\Lambda^*$  denotes the integrated hazards function. The consistency assumption rules out the possibility that future values of the true endpoint will affect the current values of the surrogate. It is analogous to the predictability condition that is needed in much of survival analysis to define proper martingale processes (Fleming and Harrington, 1991).

Now we come to the issue of identifiability of causal effects. Assume that  $P(T = S) = 0$ . Given an infinite sample, the data will provide information on  $\bar{F}(s, t) \equiv P(T > t, S > s, T > S)$  and  $\bar{F}(t) \equiv P(T > t, T < S)$ . Define two sets of random variables  $(S, T)$  and  $(U, V)$  to be *observationally equivalent* if the latter two functions are equal. We have the following result, which is an application of Proposition 1 of Abbring and van den Berg (2003):

**Theorem 1.** For any joint distribution of  $(S, T)$  given  $Z$ , there exists a joint distribu-

tion  $(S^*, T^*)$  given  $Z$  that is observationally equivalent that satisfies the “no-anticipation” assumption along with the fact that  $N_{T^*}(S^*) \perp S^*$ , where  $N_{T^*}(t) = I(T^* \leq t)$ .

A proof of Theorem 1 is given in the Appendix. The implication of the theorem is that person-specific heterogeneity within this causal modelling framework is completely confounded with the causal effect of  $S$  on  $T$ . Thus, without further untestable assumptions, one cannot separate strong causal effects from heterogeneity in the population. The result presented here has the flavor of nonidentifiability results that are available for competing risks problems in survival analysis (Tsiatis, 1975). Abbring and van der Berg (2003) get around this nonidentifiability issue by incorporating covariate effects in order to induce identifiability. However, for this to work, it is necessary that the covariates are continuous, which implies sufficient variation for identifying both person-specific effects as well as causal effects. In our situation, the only additional covariate available to us is  $Z$ , which is a binary covariate. This will not satisfy the assumptions need by Abbring and van der Berg (2003, Propositions 2 and 3, p. 1506) for identifiability.

## 2.2 A probabilistic model and semi-competing risks data

In this section, we describe a probabilistic model for  $(S, T)$  that we term a structural frailty model. We will use the following model for associating  $S$  and  $T$ : conditional on a nonnegative bivariate random vector  $\mathbf{V}_Z = (V_{SZ}, V_{TZ})$

$$\lambda_{SZ}(s|\mathbf{V}_Z) = V_{SZ}\lambda_{0S}(s|Z) \tag{1}$$

$$\lambda_{TZ}(t|\mathbf{V}_Z) = \begin{cases} V_{TZ}\lambda_{0T}(t|Z) & \text{if } t < S \\ V_{TZ}\lambda_{0T}(t|Z)\theta(t|S, Z) & \text{if } t \geq S \end{cases} \tag{2}$$

where  $\lambda_{0S}$  and  $\lambda_{0T}$  denote hazard functions for the time to surrogate and true endpoints, conditional on treatment group, and  $\theta(t|S, Z)$  denotes the effect of  $S$  on  $T$  (conditional on  $Z$ ) after the surrogate endpoint event occurs. The vector  $\mathbf{V}_Z$  is known as the frailty and is used to account for selection effects in the population (e.g., Hougaard, 1984). In (1)-(2), the causal effect is represented by  $\theta(t|S, Z)$ . It represents the multiplicative effect on the baseline hazard function for  $\lambda_{0T}(t)$  due to the occurrence of  $S$ .

The implication of Theorem 1 is that  $\theta(t|S, Z)$  and the distribution of  $\mathbf{V}_Z$  are not jointly identifiable from observed data. For the sake of exposition, let us now make the grossly

oversimplifying assumption that  $\mathbf{V}_Z$  has a degenerate distribution, i.e., there is between-person variation. We now recognize from (1)-(2) that  $\theta(t|S, Z)$  is nothing more than the cross-ratio function (Oakes, 1989) between  $S$  and  $T$ . We further assume that  $\theta(t|S, Z) = \theta_Z$ ; this signifies a cross-ratio function and corresponds to the Clayton-Oakes model (Clayton, 1978; Oakes, 1982).

We now briefly review the data structure for semi-competing risks data. In this setup, we collect observations  $(X_i, \delta_i^X, Y_i, \delta_i^Y, Z_i)$ ,  $i = 1, \dots, n$  that are a random sample from  $\{X, \delta^X, Y, \delta^Y, Z\}$ , where  $X = S \wedge T \wedge C$ ,  $Y = T \wedge C$ ,  $\delta^X = I(S \leq T \wedge C)$ ,  $\delta^Y = I(T \wedge C)$ , and  $I(A)$  is the indicator function for the event  $A$ . While  $(S, T)$  are assumed to be dependent, they are independent of  $C$  given  $Z$ . For this setup, the constant cross-ratio model can be estimated using methods proposed in Fine et al. (2001), stratified on treatment group. Thus, with this sequence of assumptions, we see that there exists causal interpretations to estimators from semi-competing risks models.

### 2.3 Relationship with other frameworks

The target estimand in this paper is the causal effect of the time of the surrogate endpoint on the time to the true endpoint. This is quite different from much of the biostatistical literature on causal inference which attempts to assess the causal effect of treatment in the study population or in a subpopulation, such as those who comply with the treatment (Frangakis and Rubin, 2002). Recognizing that both  $S$  and  $T$  are measured post-randomization, the potential outcomes framework entails formulating a joint distribution for  $(S, T)$  for each individual under the scenarios  $Z = 0$  and  $Z = 1$ . Frangakis and Rubin (2002) define principal strata as being formed by the joint distribution of a post-randomization adjustment variable under all possible treatment assignments. The upshot of this approach is that the surrogate becomes a latent pre-randomization variable whose values must be independent of randomized assignment. This approach is not directly usable here for two reasons. First, the “treatment” here is in fact the surrogate endpoint so that it becomes impossible to construct principal strata that are sensible. Second, if  $T$  is an endpoint such as death, then the semi-competing risks data structure of Fine et al. (2001) cannot be formalized within a potential outcomes



framework because of the constraint that  $S \leq T$ . In particular, the potential outcome for  $S$  cannot be defined when  $T$  occurs  $S$ ; Zhang and Rubin (2003) refer to this issue as "truncation by death".

Another causal modelling framework has been advanced by Pearl (2009). In his framework, a set of equations define the outcomes. The causal framework presented here corresponds to contrasts in  $T$  when we manipulate  $S$  to be some value. In the language of Pearl (2009), we are proposing a framework in which we calculate causal contrasts based on  $T|do(S = s)$ . Thus, this causal modelling framework is more compatible with the Pearl framework than the potential outcomes framework.

Theorem 1 has a very simple representation in the Pearl framework, especially if we make use of causal diagrams. This is illustrated in Figure 1, a version of which was described in Ghosh et al. (2010). The causal effect is represented as the arrow from  $S$  to  $T$ ; the between-person heterogeneity is represented as  $U$ . For the moment, we are assuming that we are conditioning on treatment group. Because there are arrows going from  $U$  to both  $S$  and  $T$ ,  $U$  is acting as a confounder in the model. Intuitively, Theorem 1 says that with information only on  $S$  and  $T$ , we are unable to identify the causal effect of  $S$  on  $T$  because of the presence of  $U$ . In the parlance of the Pearl framework,  $U$  is acting as a parent of  $S$ , and because it is not measured, the causal effect of  $S$  on  $T$  is nonidentifiable (Pearl 2009, Theorem 3.2.5).

### 3 Discussion

In this article, we have proposed a simple causal framework for modelling the causal effect of the surrogate endpoint on the true endpoint. The initial goal of this research was to determine if there existed a "causal" modelling framework that justified the semi-competing risks data structure that has been recently studied in the literature. While the answer to this question is in the affirmative, the result of Theorem 1 implies that one cannot disentangle the effects of strongly causal surrogate endpoints from strong selection (dependence) effects. This corresponds to the model given in Figure 1. One way to remove the effects of selection effects would be to randomize the surrogate endpoint, which is impossible. Another way would be to collect a sufficiently high-dimensional vector of covariates  $\mathbf{X}$  so that one could make the

assumption that  $(S, T) \perp U | \mathbf{X}$ . Finally, one could formulate a probabilistic model for  $U$  and its effect on the joint distribution on  $(S, T)$  where it is modelled parametrically. Again, by Theorem 1, we are unable to perform model checking for any distributional assumptions on  $U$ .

## Acknowledgments

The author would like to thank Professors Jeremy Taylor and Michael Elliott for useful discussions. This research is funded by R01 CA129402 from the National Institutes of Health.

## References

- Abbring, J. H. and van den Berg, G. J. 2003. The nonparametric identification of the treatment effects in duration models. *Econometrica* 71, 1491 – 1517.
- Begg, C. B. and Leung, D. H. Y. 2000. On the use of surrogate endpoints in randomized trials with discussion. *Journal of the Royal Statistical Society, Series A* 163, 15 – 28.
- Buyse, M. and Molenberghs, G. 1998. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 54, 1014 – 1029.
- Clayton, D. G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65, 141 – 151.
- Fine, J. P., Jiang, H. and Chappell, R. 2001. On semi-competing risks data. *Biometrika* 88, 907 – 919.
- Fleming, T. R and Harrington, D. P. 1991. *Counting Processes and Survival Analysis*. New York: Wiley.
- Frangakis, C. E. and Rubin, D. B. 2002. Principal stratification in casual inference. *Biometrics* 58, 21-29.
- Freedman, D. A. 2004. On specifying graphical models for causation, and the identification problem. *Evaluation Review* 26, 267 – 293.

- Freedman, L. S., Graubard, B. I. and Schatzkin, A. 1992. Statistical validation of intermediate endpoints for chronic disease. *Statistics in Medicine* 11, 167 – 178.
- Ghosh, D. 2009. On assessing surrogacy in a single-trial setting using a semi-competing risks paradigm. *Biometrics* 65, 521 – 529.
- Ghosh, D., Elliott, M. R. and Taylor, J. M. 2010. Links between surrogate endpoints and endogeneity. *Statistics in Medicine* 29, 2869 – 2879.
- Ghosh, D., Taylor, J. M. and Sargent, D. L. Meta-analysis for surrogacy: accelerated failure time modelling and semi-competing risks (with discussion), *Biometrics*, published online June 13, 2011. doi: 10.1111/j.1541-0420.2011.01633.x
- Gilbert, P. B. and Hudgens, M. G. 2008. Evaluating candidate principal surrogate endpoints. *Biometrics* 64, 1146 – 1154.
- Holland, P. W. 1986. Statistics and causal inference with discussion *Journal of the American Statistical Association* 81, 945 – 960.
- Hougaard, P. 1984. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* 71, 75 – 83.
- Joffe, M. M. and Greene, T. 2009. Related causal frameworks for surrogate outcomes. *Biometrics* 65, 530 – 538.
- Li, Y., Taylor, J.M.G., and Elliott, M.R. 2010. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics* 66, 523 – 531.
- Lok, J. J. 2008. Statistical modeling of causal effects in continuous time. *Annals of Statistics* 36, 1464 – 1507.
- Miller, D. R. 1977. A note on independence of multivariate lifetimes in competing risks models. *Annals of Statistics* 5, 576 – 579.
- Oakes, D. 1982. A model for association in bivariate survival data. *J. R. Statist. Soc. B* 44, 414 – 422.

- Oakes, D. 1989. Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 84, 487 – 493.
- Pearl, J. 2009. *Causality: Models, Reasoning and Inference*, 2nd edition. Cambridge University Press.
- Prentice, R. L. 1989. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* 8, 431 – 440.
- Robins, J. M. and Greenland, S. 1992. Identifiability and exchangeability of direct and indirect effects. *International Journal of Epidemiology* 3, 143-155.
- Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688 – 701.
- Taylor, J. M. G., Wang, Y. and Thiebaut, R. 2005. Counterfactual links to the proportion of treatment effect explained by a surrogate marker. *Biometrics* 61, 1102-1111.
- Tsiatis, A. 1975. A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. USA* 72, 20 – 22.
- Zhang, J. L. and Rubin, D. B. 2003. Estimation of causal effects via principal stratification when some outcomes are truncated by “death.” *Journal of Educational and Behavioral Statistics* 28, 353 – 368.

## Appendix

### Proof of Theorem 1:

Let  $(S, T)$  be given with subsurvival functions  $\bar{F}_S(s, t) = P(T > t, S > s, T > S)$  and  $\bar{F}_T(t) = P(T > t)$ . Then there exists an observationally equivalent  $(\tilde{S}, \tilde{T})$  with  $\tilde{S} \perp \tilde{T}$ , and  $\bar{F}_S(s, t) = \bar{F}_{\tilde{S}}(s, t)$  and  $\bar{F}_T(t) = \bar{F}_{\tilde{T}}(t)$  by Theorem 1 of Miller (1976). Define

$$\theta_{T^*(s)}(t) = \begin{cases} \theta_{\tilde{T}}(t) & \text{if } 0 \leq t \leq s \\ \theta_{\tilde{T}}(s) - \log[P(T > t | T > s, S = s)] & \text{if } y > s. \end{cases}$$

Define  $E$  to be an exponential random variable generated independently of  $S^*$ . Then let  $S^* = \tilde{S}$ ,  $\tilde{T}^*(s) = \theta_{T^*(s)}^{-1}(E)$  for all  $s \geq 0$  and  $T^* = T^*(S^*)$ . Then it is easy to show the following:

1.  $T^*$  satisfies Assumption 1.
2.  $T^* \perp S^*$ .
3.  $T^*$  and  $S^*$  have subsurvival functions  $\bar{F}_T(t)$  and  $\bar{F}_S(s, t)$ .

This concludes the proof.

**Figure 1.** *Graphical model conceptualizing Theorem 1.  $S$  denotes the surrogate endpoint,  $T$  the true endpoint, and  $U$  denotes between-person heterogeneity.*

