Detecting Individual Level ‘Always Survivor’ Causal Effects Under ‘Truncation by Death’ and Censoring Through Time

Jaffer M. Zaidi, Tyler J. VanderWeele

May 28, 2019

Abstract

The analysis of causal effects when the outcome of interest is possibly truncated by death has a long history in statistics and causal inference. The survivor average causal effect is commonly identified with more assumptions than those guaranteed by the design of a randomized clinical trial or using sensitivity analysis. This paper demonstrates that individual level causal effects in the ‘always survivor’ principal stratum can be identified with no stronger identification assumptions than randomization. We illustrate the practical utility of our methods using data from a clinical trial on patients with prostate cancer. Our methodology is the first and, as of yet, only proposed procedure that enables detecting causal effects in the presence of truncation by death using only the assumptions that are guaranteed by design of the clinical trial.

1 Introduction

Researchers would often like to evaluate the effect of a treatment or exposure on an outcome that could be truncated by or missing due to death [4, 18, 10]. For individuals who die before the end of study, the outcome of interest remains undefined. Such issues occur both in clinical trials and in observational studies. Many researchers note that a comparison of risk differences of those populations that survive at least until the end of the study do not offer causal conclusions [11, 13, 10]. Studies based upon principal stratification, in which comparisons are made of average causal effects in the ‘always survivor’ group, attempt to provide causal conclusions in the presence of truncation by death.

We refer the interested reader to another paper [17], for a detailed description of the different approaches with regard to survivor average causal effects. Our approach differs from these previous approaches [3, 17] in that our causal estimand is not the survivor average causal effect estimand used in these previous papers [3, 17, 4, 13, 16]. Instead, we derive a null hypothesis that when falsified informs us that the ‘always survivor’ principal stratum must exist and that there are individuals in this principal stratum of ‘always survivors’ for whom the treatment has an effect of the post-survival outcome of interest. Our approach embeds testing for these causal effects firmly within the Neyman-Pearson paradigm. The identification assumptions used to detect such effects are exactly those assumptions made to identify the average treatment effect, that is, we make no identification assumption other than randomization of treatment.

2 Notation and assumptions

Suppose a binary treatment \(X \in \{0, 1\}\) is randomized at baseline. We assume data has been collected over time on survival and on a binary outcome of interest defined only among those who survive. We assume that censoring may take place either for survival or for the binary outcome. The individuals of our study, denoted by symbol \(\omega\), compose a finite population \(\Omega\). At pre-specified times \(T = \{1, \ldots, t_f\}\) over the course of the study duration \(t_f\), we define variable \(S(\omega, t) \in \{0, 1, 2\}\) of an individual \(\omega\) for each \(t \in T\). Here, \(S(\omega, t) = 2\) denotes that the individual’s survival response was censored or missing at time \(t\). If the individual’s response is not censored at time \(t\), then \(S(\omega, t) = 1\) denotes that the individual survived at least until time \(t\), and \(S(\omega, t) = 0\) denotes that the individual did not survive until time \(t\). A binary outcome of primary interest is also measured at each of the pre-specified times \(t \in T\), though at each \(t \in T\), the outcome might be truncated through death, that is undefined because the individual died. If the individual’s response is censored by time \(t\), then let \(Y(\omega, t) = 3\), and if the individual was
not censored but did not survive until \( t \in T \), then let \( Y(\omega, t) = 2 \). If the individual was not censored and did not die before time \( t \) then let \( Y(\omega, t) = 1 \) if the binary outcome is present and \( Y(\omega, t) = 0 \) if the binary outcome is absent. Consequently the normally binary outcome is transformed to a categorical variable. For this paper, it is important to note that censoring by time \( t \), denoted as \( Y(\omega, t) = 3 \), refers to either drop-out of the individual from the trial or missing data, and that death of an uncensored individual is treated as a different event and is denoted as \( Y(\omega, t) = 2 \). If the individual \( \omega \) survives at least until \( t \in T \), then \( Y(\omega, t) = 0 \) denotes that the primary outcome of interest is not present at follow up period \( t \), and \( Y(\omega, t) = 1 \) denotes that the primary outcome of interest is present at follow up period \( t \). We use this re-coding of the outcome variables \( Y \) and \( S \) for notational simplicity. However, we also offer an alternative proof of our results in the online supplement that does not require this re-coding.

Define potential outcomes \( Y(\omega, t) \) and \( S(\omega, t) \) to be the value of \( Y(\omega, t) \) and \( S(\omega, t) \) respectively at follow up time \( t \). Let \( \omega \) be the experimenter. This amendment or generalization of potential outcomes or counterfactuals allowing for the possibility of censoring at a given time \( t \) to the experimenter in the counterfactual world provides us with a theoretical framework to examine causal effects in the presence of censoring and truncation by death.

Consider now the setting of a randomized trial. With treatment randomized at baseline, we can make the ‘weak ignorability’ assumption \( (Y(\omega, t), S(\omega, t)) \perp \!
\!
\!
\!
\!
\!
\perp X \) for all \( t \in T \). We require the consistency assumption for both \( Y(\omega, t) \) and \( S(\omega, t) \), which means that when \( X(\omega) = x \), then \( Y(\omega, t) = Y(\omega, t) \) and \( S(\omega, t) = S(\omega, t) \). This assumption states that the value of \( Y(\omega, t) \) and \( S(\omega, t) \) that would be observed if \( X \) had been set to what in fact they were observed to be is equal respectively to the value of \( Y(\omega, t) \) and \( S(\omega, t) \) that was observed. The randomization and consistency assumption are assumed throughout this paper. Additional assumptions, whenever needed, are explicitly detailed in the relevant section, theorem and proposition.

For ease of notation, we drop the \( \omega \) in \( Y(\omega, t) \) and \( S(\omega, t) \) whenever the meaning is clear. Denote for all \( t \in T \), \( P_x(\{y_1, y_0, s_1, s_0\}) = P(Y(\{y_1, y_0\}) = y_1, Y(\{y_0\}) = y_0, S(\{s_1\}) = s_1, S(\{s_0\}) = s_0) \), where \( y_1 \) and \( y_0 \) take values in the set \( \{0, 1, 2, 3\} \), and \( s_1 \) and \( s_0 \) that take values in the set \( \{0, 1, 2\} \). Also, for all \( t \in T \), we shall use the notation \( P_{s. x}(t) = P(Y(t) = y, S(t) = s \mid X = x) \) which, as we will show, enables investigators to evaluate the magnitude of causal effects for always survivors using only empirical conditions from the observed data. The superscripts are meant to distinguish these probabilities for one another. The symbol \( c(t) \) is meant to remind the reader that this probability exists in the counterfactual world, and similarly the symbol \( r(t) \) is meant to remind the reader that this probability exists in our observed or real world. All proofs are provided in the appendix. Let \( I(\cdot) \) denote the usual indicator function. Define \( Y(\omega, t) \) and \( S(x) \) as follows:

\[
Y(\omega, t) = I(Y(\omega, t) = y), \quad S(x) = I(S(x) = s) \quad \text{for} \quad y \in \{0, 1, 2, 3\}, \quad s \in \{0, 1, 2\}, \quad \omega \in \Omega \quad \text{and} \quad t \in T .
\]

We say an individual is an ‘always survivor’ at time \( t \) if \( S(\omega, t) = 0 \) or \( S(\omega, t) = 1 \) for which \( S(\omega, t) = 1 \) regardless of treatment and there is a causal effect on \( Y \). We say an individual is an ‘always survivor’ at time \( t \) if \( S(\omega, t) = 0 \). Joint counterfactuals of the form \( Y(\omega, t) = 1 \) and \( S(\omega, t) = 0 \) are never considered, as the individual has not survived until time \( t \). Consequently, any counterfactual probability \( P(\{y_1, y_0, s_1, s_0\}) \) for which \( s_1 = 0 \) and \( y_1 = 0 \) or \( s_0 = 0 \) and \( y_0 = 0 \) is identically zero. Also \( P(\{y_1, y_0, s_1, s_0\}) = 0 \) whenever \( s_1 = 0 \) and \( y_0 \neq 3 \) or \( s_0 = 2 \) and \( y_0 \neq 3 \). Similarly, \( F_{g,s. x}(t) \) is equal to zero whenever any of the following conditions hold: (1) \( s = 2 \) and \( y \neq 3 \); (2) \( y = 3 \) and \( s \neq 2 \) (3) or \( s = 0 \) and \( y \neq 2 \); (4) \( y = 2 \) and \( s \neq 0 \). Situations where, at time \( t \), the realized survival status \( s \) is not censored or missing but the realized value \( y \) is censored are possible in practice. However, in our application below, these situations do not occur in our data analysis by design of the clinical trial. The theoretical framework to accommodate such situations entirely replicates the arguments presented here, and online supplement fully explicates the framework and results in the setting where there exists individuals \( \omega \) for whom \( Y(t, \omega) \) is censored but \( S(t, \omega) \) is not censored for specified time \( t \).
3 Identification without monotonicity

Here, we only require randomization and consistency for Theorem 1, Proposition 1, and Corollary 1. In the next section, we will examine the derivation of similar results in the presence of monotonicity assumptions.

**Theorem 1**

Suppose \( X \) is randomized at baseline. If for some \( t \in T \), \( P_{y,1.1}^t + P_{1-y,1.0}^t > 1 \) for \( y \in \{0,1\} \), then there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \) and \( Y_1(\omega, t) = y \).

Theorem 1 allows for the empirical detection of individual level total effects within the ‘always survivor’ principal stratum. A difference between our result and previous methods is that Theorem 1 effectively only requires randomization and yet still provides conditions for which the ‘always survivor’ principal stratum exists and has individuals for whom the treatment changes the outcome. This is the first time an empirical condition to detect causal effects is formulated in the presence of truncation by death and censoring without any additional assumptions other than those guaranteed by design of the clinical trial. Proposition 1 provides further context, comparing the proportion of individuals who display a positive (or negative) effect of treatment \( X \) on an outcome \( Y \) within the ‘always survivor’ principal stratum to the proportion of individuals who display a negative (or positive) effect of treatment \( X \) on outcome \( Y \) within the always survivor principal stratum. Corollary 1 provides further results for this difference for all individuals in the always survivor principal stratum (observed or censored).

**Proposition 1**

Suppose \( X \) is randomized at baseline. If for some \( t \in T \), \( P_{y,1.1}^t + P_{1-y,1.0}^t > 1 \) for \( y \in \{0,1\} \), then we have the following result concerning counterfactuals:

\[
P^c(t)(y, 1 - y, 1, 1) > P^c(t)(1 - y, y, 1, 1) + P^c(t)(2, y, 0, 1) + P^c(t)(1 - y, 2, 1, 0) \\
+ P^c(t)(2, 2, 0, 0) + P^c(t)(3, 3, 2, 2) + P^c(t)(3, y, 2, 1) \\
+ P^c(t)(3, 2, 2, 0) + P^c(t)(1 - y, 3, 1, 2) + P^c(t)(2, 3, 0, 2)
\]

**Corollary 1**

Suppose \( X \) is randomized at baseline. The expression \( P_{1,1.1}^t + P_{0,1.0}^t - 1 \) is a lower bound on the risk difference between the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom the treatment causes the outcome and the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom the treatment prevents the outcome of interest.

In counterfactual notation, the expression \( P_{1,1.1}^t + P_{0,1.0}^t - 1 \) is a lower bound on \( P^c(t)(1, 0, 1, 1) - [P^c(t)(0, 1, 1, 1) - P^c(t)(3, 3, 2, 2) - P^c(t)(3, 1, 2, 1) - P^c(t)(0, 3, 1, 2)] \). Note \( P^c(t)(1, 0, 1, 1) \) is a lower bound on the proportion of individuals randomized at baseline that are always survivors for whom the treatment causes the outcome \( Y \) as it does not count any censored individual that could be an always survivor for whom the treatment causes the outcome \( Y \). The term in the square brackets is an upper bound on the proportion of survivors randomized at baseline that are always survivors for whom the treatment prevents the outcome \( Y \), because it effectively treats any censored individual that could possibly be an always survivor for whom the treatment prevents the outcome \( Y \) as actually an always survivor for whom the treatment prevents the outcome \( Y \).

It follows from Corollary 1, the expression \( \max\{P_{1,1.1}^t + P_{0,1.0}^t - 1\} \) is thus a lower bound on the proportion of always survivors (observed or censored) at time \( t \) for whom the treatment causes the outcome: \( P^c(t)(1, 0, 1, 1) \). Similarly, the expression \( P_{0,1.1}^t + P_{1,1.0}^t - 1 \) is a lower bound on the risk difference between the proportion of always survivors (observed or censored) at time \( t \) for whom the treatment prevents the outcome and the proportion of always survivors (observed or censored) at time \( t \) for whom the treatment causes the outcome of interest. In counterfactual notation, the expression \( P_{0,1.1}^t + P_{1,1.0}^t - 1 \) is a lower bound on \( P^c(t)(0, 1, 1, 1) - P^c(t)(1, 0, 1, 1) - P^c(t)(3, 3, 2, 2) - P^c(t)(3, 0, 2, 1) - P^c(t)(1, 3, 1, 2) \). The expression \( \max\{P_{0,1.1}^t + P_{1,1.0}^t - 1\} \) is a lower bound on the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom the treatment prevents the outcome: \( P^c(t)(0, 1, 1, 1) \).

Remarkably, without any assumption on the censoring mechanism, we can formulate lower bounds on proportion of individuals randomized at baseline that are always survivors (censored or not) at time \( t \) and for whom the
treatment causes (prevents) the outcome minus the proportion of individuals randomized at baseline that are always survivors (censored or not) at time \( t \) and for whom the treatment prevents (causes respectively) the outcome. To the best of our knowledge, no other paper provides empirical conditions to detect causal effects in the presence of censoring without assumptions on the censoring mechanism. Corollary 1 also gives us a contrast only involving real-world probabilities corresponding to the minimum proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom treatment has an effect.

Theorem 1, Proposition 1 and Corollary 1 provide scientists with interpretations, under weaker assumptions, that are not possible in any of the previous approaches \([6 \ 17 \ 3]\) to deal with truncation by death. Previous approaches that examine the survivor average causal effect make more assumptions than randomization and consistency \([3 \ 6 \ 17]\). Bounds to the survivor average causal effect is provided in the online supplement.

### 4 Identification of always survivor causal effects under monotonicity

Monotonicity assumptions are often employed in methods to evaluate principal stratum direct effects. Recall, we denote \( Y_y^0(\omega, t) = I(Y_{x}(\omega, t) = y) \), and \( S_y^0(\omega, t) = I(S_{x}(\omega, t) = s) \) for \( y \in \{0, 1, 2, 3\} \), \( s \in \{0, 1, 2\} \), and \( t \in T \). Here we will consider additional results under two different monotonicity assumptions. In the first monotonicity assumption, we will assume \( S_y^0(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \) and for a pre-specified \( t \in T \). Heuristically, this assumption translates to it being the case that no individual exists who would be alive and not censored at least until time \( t \) when forced to take control, but would be dead and not censored by time \( t \) when forced to take treatment. This means that we do not observe any individual for whom \( S_1(\omega, t) = 0 \) and \( S_0(\omega, t) = 1 \) for the specified \( t \in T \). In the absence of censoring, this assumption is simply that for no individual does the treatment itself cause death. If treatment is considered to be non-smoking and the control is smoking, scientific literature might lead us to believe that such an assumption is reasonable to make. Subject matter expertise could provide guidance for which \( t \in T \), \( S_y^0(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for all \( \omega \in \Omega \), \( S_y^0(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) holds for only a subset of \( T_2 \subset T \), then Theorem 2A and Proposition 2A, provided below, applies only for those \( t \in T_2 \).

While such monotonicity constraints are never verifiable, they are falsifiable, and again can sometimes be justified with subject matter knowledge.

For the second monotonicity assumption, we shall assume that \( S_y^0(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \) and for a pre-specified \( t \in T \). Heuristically, this assumption means that there is no individual who would drop out of the study by time \( t \) when forced to take treatment but who would not drop out of the study and is alive when forced to take control condition (e.g. a prior or older treatment). This means that we do not observe any individual for whom \( S_1(\omega, t) = 2 \) and \( S_0(\omega, t) = 1 \) for the specified \( t \in T \). Again, such monotonicity constraints are falsifiable and could be judged using subject matter knowledge. For instance, if clinicians believe that the new treatment is less toxic than the old treatment, then this could provide some evidence that patients might be less disposed to drop out if they are forced to take the less toxic new treatment in comparison to the more toxic old treatment. We will consider analogous results to Theorem 1 and Proposition 1 when both of these two monotonicity assumptions hold. If such monotonicity assumptions are untenable for the population in question, investigators can still use the empirical conditions provided earlier that do not need any monotonicity assumptions. Causally interpretable sensitivity analyses are also presented in Section 6 when one or both of the monotonicity assumptions do not hold.

**Theorem 2A**

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S_y^0(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for some \( y \in \{0, 1\} \) the empirical condition \( P_y^t(1-y,1,0) - P_y^t(1-y,1,1) - P_y^t(3,2,1) > 0 \) holds, then there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \), \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y \), \( Y_0(\omega, t) = 1-y \).

**Proposition 2A**

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S_y^0(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for some \( y \in \{0, 1\} \) the empirical condition \( P_y^t(1-y,1,0) - P_y^t(1-y,1,1) - P_y^t(3,2,1) > 0 \) holds, then we have the following result concerning counterfactuals:

\[
P_c^t(y, 1-y, 1, 1) = \begin{align*}
P_c^t(1-y, y, 1, 1) + P_c^t(1-y, 2, 1, 0) \\
+ P_c^t(3, 3, 2, 2) + P_c^t(3, y, 2, 1) \\
+ P_c^t(3, 2, 2, 0) + P_c^t(1-y, 3, 1, 2)
\end{align*}
\]
Corollary 2A

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). The expression \( P^{r(t)}_{0,1,0} - P^{r(t)}_{0,0,1} - P^{r(t)}_{3,2,1} \) is a lower bound on the risk difference between the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom the treatment causes the outcome and the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom the treatment prevents the outcome of interest.

The interpretations for Theorem 2B and Proposition 2B are similar to those provided for Theorem 1 and Proposition 1. Specifically, if we believe that no individual exists that would die and not be censored by time \( t \) when forced to take control and if \( P^{r(t)}_{0,1,0} - P^{r(t)}_{0,0,1} - P^{r(t)}_{3,2,1} > 0 \), then there exists a set of individuals that would live at least until time \( t \) and for whom the treatment causes the outcome. From Corollary 2A, we learn that the proportion of always survivors (observed or censored) at time \( t \) for whom the treatment prevents (causes) the outcome of interest is at least greater the proportion of always survivors (observed or censored) regardless of treatment and for whom the treatment prevents (causes respectively) outcome of interest. Also, Corollary 2A provides a contrast involving only real-world probabilities that is the minimum proportion of always survivors (observed or censored) at time \( t \) for whom the treatment causes (or prevents) the outcome. The other monotonicity assumption also provides useful results and is presented next.

Theorem 2B

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for some \( y \in \{0, 1\} \) the empirical condition \( P^{r(t)}_{1-y,1,0} - P^{r(t)}_{1-y,1,1} - P^{r(t)}_{2,0,1} > 0 \) holds, then there exists a non-empty subpopulation \( \Omega_y \) such that for every \( \omega \in \Omega_y \), \( S_1(\omega, t) = 1 \), \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y, Y_0(\omega, t) = 1 - y \).

Proposition 2B

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for some \( y \in \{0, 1\} \) the empirical condition \( P^{r(t)}_{1-y,1,0} - P^{r(t)}_{1-y,1,1} - P^{r(t)}_{2,0,1} > 0 \) holds, then we have the following result concerning counterfactuals:

\[
P^{r(t)}(y, 1 - y, 1, 1) \geq P^{r(t)}(1 - y, 1 - y, 1, 1) + P^{r(t)}(1 - y, 2, 1, 0) + P^{r(t)}(2, y, 0, 1) + P^{r(t)}(2, 2, 0, 0) + P^{r(t)}(1 - y, 3, 1, 2) + P^{r(t)}(2, 3, 0, 2)
\]

Corollary 2B

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). The expression \( \max\{P^{r(t)}_{0,1,0} - P^{r(t)}_{0,0,1} - P^{r(t)}_{2,0,1}, 0\} \) is a lower bound on the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom treatment causes the outcome. Similarly, the expression \( \max\{P^{r(t)}_{1,1,0} - P^{r(t)}_{1,1,1} - P^{r(t)}_{2,0,1}, 0\} \) is a lower bound on the proportion of individuals randomized at baseline that are always survivors (observed or censored) at time \( t \) for whom treatment causes the outcome. The interpretations for Theorem 2B and Proposition 2B are similar to those provided for Theorem 2A and Proposition 2A. Specifically, if we believe that no individual exists that would be censored by time \( t \) when forced to take control and if \( P^{r(t)}_{0,1,0} - P^{r(t)}_{0,0,1} - P^{r(t)}_{2,0,1} > 0 \), then there exists a set of individuals who would live at least until time \( t \) and for whom the treatment causes outcome of interest. From Proposition 2B, we learn that the proportion of individuals
randomized at baseline that are observed to live at least until time \( t \) and for whom the treatment causes outcome is greater than the proportion of individuals randomized at baseline that are observed to live at least until time \( t \) and for whom the treatment prevents outcome \( Y \). Corollary 2B provides a lower bound on the always survivors for whom the treatment causes (or prevents) the outcome of interest. Finally, if we believe that the two monotonicity assumptions hold, then we have the following result.

**Theorem 2C**

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S^{[2]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) and \( S^{[0]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for some \( y \in \{0, 1\} \) the empirical condition \( P_{1-y,1,0}^{(t)}(s) - P_{1-y,1,1}^{(t)}(s) > 0 \) holds, then there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \), \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y, Y_0(\omega, t) = 1 - y \).

**Proposition 2C**

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S^{[2]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) and \( S^{[0]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). If for some \( y \in \{0, 1\} \) the empirical condition \( P_{1-y,1,0}^{(t)}(s) - P_{1-y,1,1}^{(t)}(s) > 0 \) holds, then we have the following result concerning counterfactuals:

\[
P^{(t)}(y, 1 - y, 1, 1) > P^{(t)}(1 - y, y, 1, 1) + P^{(t)}(1 - y, 2, 1, 0) + P^{(t)}(1 - y, 3, 1, 2)
\]

**Corollary 2C**

Suppose \( X \) is randomized at baseline. In addition, suppose that for some \( t \in T \), \( S^{[2]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) and \( S^{[0]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) for all \( \omega \in \Omega \). The expression \( \max\{P_{0,1,0}^{(t)} - P_{0,1,1}^{(t)}, 0\} \) is a lower bound on the proportion of always survivors (observed or censored) at time \( t \) for whom treatment causes the outcome.

Similarly, the expression \( \max\{P_{1,1,0}^{(t)} - P_{1,1,1}^{(t)}, 0\} \) is a lower bound on the proportion of always survivors (observed or censored) at time \( t \) for whom treatment prevents the outcome. The interpretations of Theorem, Proposition, and Corollary 2C are similar to those provided for Theorem, Proposition and Corollary 2B respectively. The conditions are to detect individual level always survivor causal effects are weaker the more monotonicity assumptions are made. Again, if such monotonicity assumptions cannot be justified on scientific grounds, investigators can still use the results that do not require such assumptions.

## 5 Inference for Individual Level Always Survivor Causal Effects

To investigate individual level always survivor causal effects, null hypotheses can be formulated that when falsified produce the inequalities associated with Theorem 1, 2A, 2B, or 2C. For Theorem 1, if we reject \( P_{y,1,1}^{(t)} + P_{1-y,1,0}^{(t)} \leq 1 \) for \( y \in \{0, 1\} \), then we believe with a fixed type one error rate that there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \), \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y, Y_0(\omega, t) = 1 - y \), and with the same fixed type one error rate we believe Proposition 1 to be true. Testing \( P_{y,1,1}^{(t)} + P_{1-y,1,0}^{(t)} \leq 1 \) is equivalent to testing \( P_{1-y,1,0}^{(t)} \leq P_{1-y,1,1}^{(t)} + P_{2,0,1}^{(t)} + P_{3,2,1}^{(t)} \), which can also be converted into a one-sided difference of proportions. Note testing \( P_{1-y,1,0}^{(t)} \leq P_{y,1,1}^{(t)} + P_{2,0,1}^{(t)} + P_{3,2,1}^{(t)} \) is equivalent to testing \( P(Y(t) = 1 - y, S(t) = 1 \mid X = 0) \leq P\{Y(t) = 1 - y, S(t) = 1 \mid Y(t) = 2, S(t) = 0\} \cup \{Y(t) = 3, S(t) = 2\} \mid X = 1\).

To use Theorem 2A, if assume \( S^{[0]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) for all \( \omega \in \Omega \) and we reject, \( P_{1-y,1,0}^{(t)} - P_{1-y,1,1}^{(t)} - P_{3,2,1}^{(t)} \leq 0 \) for some \( y \in \{0, 1\} \), then we believe with a fixed type one error rate that there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \), \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y, Y_0(\omega, t) = 1 - y \), and with the same type one error rate we believe Proposition 2A to be true. Note testing \( P_{1-y,1,0}^{(t)} \leq P_{y,1,1}^{(t)} + P_{2,0,1}^{(t)} \) is equivalent to testing \( P(Y(t) = 1 - y, S = 1 \mid X = 0) \leq P\{Y(t) = 1 - y, S(t) = 1\} \cup \{Y(t) = 2, S(t) = 0\} \mid X = 1\). Similarly, to use Theorem 2B, if assume \( S^{[2]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) for all \( \omega \in \Omega \) and we reject, \( P_{1-y,1,0}^{(t)} - P_{1-y,1,1}^{(t)} - P_{2,0,1}^{(t)} \leq 0 \) for some \( y \in \{0, 1\} \), then we believe with a fixed type one error rate that there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \), \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y, Y_0(\omega, t) = 1 - y \), and with the same type one error rate we believe Proposition 2B to be true. Note testing \( P_{1-y,1,0}^{(t)} \leq P_{y,1,1}^{(t)} + P_{2,0,1}^{(t)} \) is equivalent to testing \( P(Y(t) = 1 - y, S = 1 \mid X = 0) \leq P\{Y(t) = 1 - y, S(t) = 1\} \cup \{Y(t) = 2, S(t) = 0\} \mid X = 1\). Finally, to use Theorem 2C, if assume and \( S^{[2]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) and \( S^{[0]}_1(\omega, t) + S^{[1]}_0(\omega, t) \leq 1 \) for all \( \omega \in \Omega \) and we
reject, \( P_{1-y,1.0}^{c(t)} - P_{1-y,1.1}^{c(t)} \leq 0 \) for some \( y \in \{0, 1\} \), then we believe with a fixed type one error rate that there exists a non-empty subpopulation \( \Omega_s \) such that for every \( \omega \in \Omega_s \), \( S_1(\omega, t) = 1 \) \( S_0(\omega, t) = 1 \) and \( Y_1(\omega, t) = y \), \( Y_0(\omega, t) = 1 - y \), and with the same type one error rate we believe Proposition 2C to be true.

In our data application below we have five fixed time periods of interest and so we use Theorems 1 five times and apply a Bonferroni correction. We also apply Theorem 2A, 2B, 2C arguing that the relevant monotonicity assumptions likely hold, and demonstrate similar conclusions. Note for the data application, we use the equivalent test expressed in a one-sided difference of proportions instead of the complement of the inequality associated with Theorems 1, 2A, 2B, 2C.

The online supplement provides Bayesian and randomization based inferential methods to assess results associated with Theorem 1, 2A, 2B, 2C. For our data application, we use standard t-test for a one-sided difference of proportions to assess always survivor causal effects, because of the moderately large sample size and ease of implementation for readers. In the setting of smaller sample sizes, the randomization or Bayesian approach might be more appropriate.

6 Sensitivity Analysis for monotonicity assumptions

As demonstrated in Section 3, we provide methods for identification of individual level always survivor causal effects with only the assumptions guaranteed by design of the clinical trial. Theorem 1, Proposition 1 and Corollary 1 effectively only require randomization at baseline to first detect always survivor causal effects and secondly provide a population level characterization of such causal effects. Monotonicity assumptions on survival or censoring enable scientists to detect always survivor causal effects under assumptions that are justified under using subject matter knowledge. Sensitivity analysis can provide scientists with methods to interpret the causal estimands of interest when such monotonicity assumptions do not hold true.

Theorem 3A

Suppose \( X \) is randomized at baseline. If for some \( y \in \{0, 1\} \) the empirical condition \( P_{1-y,1.0}^{c(t)} - P_{1-y,1.1}^{c(t)} - P_{3.2.1}^{c(t)} > d_m(t) \) holds for \( d_m(t) = P_{3.2.1}^{c(t)}(2, 1 - y, 0, 1) - P_{3.2.1}^{c(t)}(1 - y, 2, 1, 0) - P_{3.2.1}^{c(t)}(3, 2, 2, 0) \), then we have the following result concerning counterfactuals:

\[
P_{3.2.1}^{c(t)}(y, 1 - y, 1, 1) > P_{3.2.1}^{c(t)}(1 - y, y, 1, 1) + P_{3.2.1}^{c(t)}(3, 3, 2, 2) + P_{3.2.1}^{c(t)}(3, y, 2, 1) + P_{3.2.1}^{c(t)}(1 - y, 3, 1, 2).
\]

Theorem 3B

Suppose \( X \) is randomized at baseline. If for some \( y \in \{0, 1\} \) the empirical condition \( P_{1-y,1.0}^{c(t)} - P_{1-y,1.1}^{c(t)} - P_{2.0.1}^{c(t)} > c_m(t) \) holds for \( c_m(t) = P_{3.1}^{c(t)}(3, 1 - y, 2, 1) - P_{3.1}^{c(t)}(1 - y, 2, 1, 0) - P_{3.1}^{c(t)}(1 - y, 3, 1, 2) - P_{3.1}^{c(t)}(2, y, 0, 1) - P_{3.1}^{c(t)}(2, 2, 0, 0) - P_{3.1}^{c(t)}(2, 3, 0, 2) \), then we have the following result concerning counterfactuals:

\[
P_{3.1}^{c(t)}(y, 1 - y, 1, 1) > P_{3.1}^{c(t)}(1 - y, y, 1, 1).
\]

Theorem 3C

Suppose \( X \) is randomized at baseline. If for some \( y \in \{0, 1\} \) the empirical condition \( P_{1-y,1.0}^{c(t)} - P_{1-y,1.1}^{c(t)} > k_m(t) \) holds for \( k_m(t) = P_{3.1}^{c(t)}(2, 1 - y, 0, 1) + P_{3.1}^{c(t)}(3, 1 - y, 2, 1) - P_{3.1}^{c(t)}(1 - y, 2, 1, 0) - P_{3.1}^{c(t)}(1 - y, 3, 1, 2) \) then we have the following result concerning counterfactuals:

\[
P_{3.1}^{c(t)}(y, 1 - y, 1, 1) > P_{3.1}^{c(t)}(1 - y, y, 1, 1).
\]
Readers should realize that \( d_m(t), c_m(t) \) and \( k_m(t) \) are not point identified in randomized studies. A simple, yet effective, approach for sensitivity analysis would be to first estimate the left hand side of the inequalities associated with Theorem 3A, Theorem 3B, and Theorem 3C, and then crudely interpret \( d_m(t), c_m(t), k_m(t) \) as the proportion of individuals of a specific counterfactual form to render the respective inferences associated with Theorem 3A, 3B, 3C to be invalid. As an example, suppose the researcher estimates \( P_{0.1, 0}^{r(t)} - P_{0.1, 1}^{r(t)} \) for a specific \( t \in T \) as 0.2, then the proportion of individuals randomized at baseline that follow counterfactual response types \( \{\omega_1, \omega_2\} \in \Omega \)

\[
Y_1(\omega_1, t) = 2, \quad Y_0(\omega_1, t) = 1, \quad S_1(\omega_1, t) = 0, \quad S_0(\omega_1, t) = 1 \text{ or } Y(\omega_2, t) = 3, \quad Y_0(\omega_2, t) = 0, \quad S_1(\omega_2, t) = 2, \quad S_0(\omega_2, t) = 1
\]

has to be at least 0.2 at time \( t \) for there to be no individual of counterfactual response type \( Y_1(\omega, t) = 1, Y_0(\omega, t) = 0, S_1(\omega, t) = 1, \) and \( S_0(\omega, t) = 1 \) in our population. See the data example for an illustration of the utility of such a sensitivity analysis.

7 Application to Southwest Oncology Group Trial

Our results are applied to the data from the Southwest Oncology Group (SWOG) Trial, which was a phase III trial to compare docetaxal plus estramustine against mitoxantrone plus prednisone in men with metastatic, hormone-independent prostate cancer [7]. We use software [5] to reconstruct data from the Kaplan-Meir survival curves presented in the Petrylak paper [7]. The reconstructed dataset might be slightly different from those presented in the original article as a result of measurement error stemming from image analysis of the Kaplan-Meir curves [5]. Similar reconstructed data was analyzed in previous papers [3, 17]. The purpose of this applied example is to illustrate the broad utility of our methods to detect causal effects in clinical trials in which an outcome of interest is truncated by death or censored due to drop out.

For the trial, the primary end-point was overall survival [7]. A total of 338 patients were randomized to the docetaxal plus estramustine (henceforth referred to as docetaxal), and a total of 336 were randomized to the mitoxantrone plus prednisone (henceforth referred to as mitoxantrone). The trial found that docetaxal improved median survival in comparison to mitoxantrone [7]. Progression of cancer is an important secondary outcome of interest and may be subject to truncation by death. Our hypothesis is that for some pre-specified times \( t \in T_{SWOG} \), the proportion of individuals who would remain alive regardless of treatment at least until the pre-specified time \( t \) and whose cancer would progress under mitoxantrone but not under docetaxal at time \( t \) is greater than the proportion of individuals who would remain alive regardless of treatment at least until the pre-specified time \( t \) and whose cancer would progress under docetaxal but not under mitoxantrone at time \( t \).

The time periods are 1 month \((t = 1)\), 2 months \((t = 2)\), 3 months \((t = 3)\), 4 months \((t = 4)\), 6 months \((t = 5)\), 12 months \((t = 6)\), and 18 months \((t = 7)\) post treatment initiation \(X\). Here, \(X = 1\) denotes that the individual was placed on the docetaxal arm, and \(X = 0\) denotes that the individual was placed on the mitoxantrone arm. Outcome \(Y(t)\) is defined as a categorical variable that takes values in the set \(\{0, 1, 2, 3\}\) for each \(t \in T_{SWOG} = \{1, \ldots , 7\}\), where \(Y(t) = 0\) means that the individual’s cancer did not progress at time point \(t\), \(Y(t) = 1\) means that the individual’s cancer progressed at time point \(t\), \(Y(t) = 2\) means that the individual did not survive until time \(t\), and finally \(Y(t) = 3\) means that the individual’s observation was censored or missing at time \(t\). Survival status at time \(t\), denoted \(S(t)\), is also defined as a categorical variable that takes value in the set \(\{0, 1, 2\}\) for each \(t \in T_{SWOG}\), where \(S(t) = 0\) means that the individual did not survive at time period \(t\), \(S(t) = 1\) means that the individual survived at least time period \(t\), and \(S(t) = 2\) means that the individual’s survival status is censored or missing at time period \(t\). We provide the relevant contingency tables in the appendix and the relevant code that is used to produce all our results in the online supplement.

7.1 Data Analysis without Monotonicity Assumptions

Without any monotonicity assumptions, we need to test \(P_{0,1,1}^{r(t)} + P_{1,1,0}^{r(t)} \leq 1\) for each \(t \in \{1, \ldots , 7\}\) to evaluate at each time point whether there exist individuals for whom \(S_1(\omega, t) = 1, S_0(\omega, t) = 1, Y_1(\omega, t) = 0, Y_0(\omega, t) = 1\), that is these, individuals would survive at least until time \(t\) regardless of treatment, but treatment with docetaxal (vs. mitoxantrone) would prevent their cancer from progressing. This hypothesis test is equivalent to testing \(P_{1,1,0}^{r(t)} \leq P_{1,1,1}^{r(t)} + P_{2,0,1}^{r(t)} + P_{3,2,1}^{r(t)}\). To conduct this hypothesis test, we provide the following three contingency tables:
were always survivors for whom treatment had a causal effect on cancer progression. The assumption because of the focus, not on the survivor average causal effect, but rather detecting individuals who were unable to draw such conclusions [3, 17]. Note also that these other papers, in contrast to ours, made more assumptions than those guaranteed by randomization alone. 

We are able to draw conclusions here even with weaker assumptions than those guaranteed by principal stratification to deal with truncation by death [3, 17]. Two papers [3, 17] have used similar data and were unable to draw such conclusions [3, 17]. Note also that these other papers, in contrast to ours, made more assumptions than those guaranteed by randomization alone. We are able to draw conclusions here even with weaker assumption because of the focus, not on the survivor average causal effect, but rather detecting individuals who were always survivors for whom treatment had a causal effect on cancer progression.

### Table 1: Survive and cancer progression at each $t \in T_{SWOG}$, $(S(t) = 1, Y(t) = 1)$

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| $X = 1$ | 6       | 28       | 60       | 81       | 108      | 143       | 102       |
| $X = 0$ | 37      | 94       | 142      | 145      | 155      | 129       | 93        |

### Table 2: Censored at $t \in T_{SWOG}$, $(Y(t) = 3, S(t) = 2)$:

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| $X = 1$ | 2       | 5        | 8        | 11       | 16       | 33        | 61        |
| $X = 0$ | 2       | 5        | 8        | 11       | 17       | 35        | 52        |

### Table 3: Did not survive until $t \in T_{SWOG}$, $(Y(t) = 2, S(t) = 0)$:

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| $X = 1$ | 2       | 8        | 13       | 20       | 38       | 87        | 155       |
| $X = 0$ | 4       | 10       | 16       | 26       | 48       | 116       | 176       |

Two-sided confidence intervals for $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P(Y(t) = 1, S(t) = 1) \cup \{Y(t) = 3, S(t) = 2 \mid X = 1\}$ and one-sided $p$-values for Wald test of the null hypothesis $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P(Y(t) = 1, S(t) = 1) \cup \{Y(t) = 3, S(t) = 2 \mid X = 1\} \leq 0$ are provided below:

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| Estimate | 0.08    | 0.16     | 0.18     | 0.10     | -0.018   | -0.39     | -0.66     |
| 95% CI | (0.04, 0.12) | (0.10, 0.22) | (0.11, 0.26) | (0.02, 0.18) | (-0.10, 0.06) | (-0.47, -0.32) | (-0.72, -0.61) |
| 99% CI | (0.03, 0.13) | (0.08, 0.24) | (0.09, 0.28) | (0.00, 0.20) | (-0.12, 0.08) | (-0.49, -0.30) | (-0.74, -0.59) |
| p-value | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0047 | 0.652 | 1 | 1 |

The conclusions drawn in this data analysis could not be made with any of the previous methods that used statistical evidence suggesting that the proportion of always survivors (observed or censored) at one month post treatment initiation whose cancer would not progress on docetaxal but would progress on mitoxantrone is greater by 0.08 with a 95 percent confidence interval of (0.04, 0.12) than the proportion of always survivors (observed or censored) at one month post treatment initiation whose cancer would progress on docetaxal but would not progress on mitoxantrone. We do not find evidence for such individual level always survivor causal effects after 6 months. At three months, statistical evidence suggests that the proportion of always survivors three months post treatment initiation whose cancer would not progress on docetaxal but would on mitoxantrone is at least 0.18 with a 95 percent confidence interval of (0.11, 0.26). Note that this is a lower bound of such individuals, and with randomization alone this proportion is not point identified. In terms of the 336 + 338 = 674 individuals that were enrolled in this clinical trial, on average at least 0.18 · 674 = 121 of these individuals should survive until 3 months regardless of which treatment arm they were randomized and their cancer would worsen under mitoxantrone but not docetaxal. The results from Table 4 demonstrate by Corollary 1 for the Southwest Oncology Group Trial, that there is statistical evidence suggesting that the proportion of always survivors (observed or censored) at one month post treatment initiation whose cancer would not progress on docetaxal but would progress on mitoxantrone is greater by 0.08 with a 95 percent confidence interval of (0.04, 0.12) than the proportion of always survivors (observed or censored) at one month post treatment initiation whose cancer would progress on docetaxal but would not progress on mitoxantrone. We do not find evidence for such individual level always survivor causal effects after 6 months. At three months, statistical evidence suggests that the proportion of always survivors three months post treatment initiation whose cancer would not progress on docetaxal but would on mitoxantrone is at least 0.18 with a 95 percent confidence interval of (0.11, 0.26). Note that this is a lower bound of such individuals, and with randomization alone this proportion is not point identified. In terms of the 336 + 338 = 674 individuals that were enrolled in this clinical trial, on average at least 0.18 · 674 = 121 of these individuals should survive until 3 months regardless of which treatment arm they were randomized and their cancer would worsen under mitoxantrone but not docetaxal. The conclusions drawn in this data analysis could not be made with any of the previous methods that used principal stratification to deal with truncation by death [3, 17]. Two papers [3, 17] have used similar data and were unable to draw such conclusions [3, 17]. Note also that these other papers, in contrast to ours, made more assumptions than those guaranteed by randomization alone. We are able to draw conclusions here even with weaker assumption because of the focus, not on the survivor average causal effect, but rather detecting individuals who were always survivors for whom treatment had a causal effect on cancer progression.
7.2 Data Analysis with Monotonicity Assumption $S^0_t(\omega, t) + S^1_t(\omega, t) \leq 1$

For each $t \in \{1, \ldots, 7\}$, we test the null hypothesis $P_{t, \omega}^{(1)} \leq P_{t, \omega}^{(2)} + P_{t, \omega}^{(3)}$ to evaluate for a fixed $t$ whether the proportion of individuals for whom $Y_1(\omega, t) = 0$, $Y_0(t, \omega) = 1$, $S^1_t(\omega, t) = 1$, and $S^0_t(\omega, t) = 1$ is greater than the proportion of individuals for whom $Y_1(\omega, t) = 1$, $Y_0(t, \omega) = 0$, $S^1_t(\omega, t) = 1$, and $S^0_t(\omega, t) = 1$ under the assumption that $S^0_t(\omega, t) + S^1_t(\omega, t) \leq 1$ for all $\omega \in \Omega$ and $t \in T_{SWOG}$. The following table provides 95% and 99% confidence intervals for $P_{t, \omega}^{(1)} - P_{t, \omega}^{(2)} - P_{t, \omega}^{(3)}$ for each $t \in T_{SWOG}$. Notice $P(S(t) = 0 \mid X = 1) + P(S(t) = 1 \mid X = 0) \leq 1$ for all $t \in T_{SWOG}$ and therefore we fail to falsify the monotonicity assumption, though of course this does not guarantee that the assumption holds.

Two-sided confidence intervals of $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P\{Y(t) = 1, S(t) = 1\} \cup \{Y(t) = 3, S(t) = 2\} \mid X = 1 \leq 0$ are provided below:

| t | Estimate | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|---|----------|---------|----------|----------|----------|----------|-----------|----------|
| 1 | 0.09     | 0.18    | 0.22     | 0.16     | 0.09     | -0.14    | -0.21     |
| 95% CI | (0.05, 0.13) | (0.12, 0.24) | (0.15, 0.29) | (0.09, 0.23) | (0.02, 0.17) | (-0.21, -0.5) | (-0.28, -0.13) |
| 99% CI | (0.03, 0.14) | (0.10, 0.26) | (0.13, 0.31) | (0.06, 0.26) | (-0.01, 0.19) | (-0.24, -0.04) | (-0.30, -0.11) |
| p-value | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.008    | 0.998    | 1         |

Table 5: Results Monotonicity $S^0_t(\omega, t) + S^1_t(\omega, t) \leq 1$ for all $\omega \in \Omega$

For this section, we have assumed that $S^0_t(\omega, t) + S^1_t(\omega, t) \leq 1$ for all $t \in T_{SWOG}$. Notice with this assumption, we have the same conclusions as before at one three months regarding always survivors, and now also have statistical evidence that the proportion of always survivors (observed or censored) six months post treatment initiation whose cancer would not progress on docetaxal but would progress on mitoxantrone is greater by 0.09 with a 95 percent confidence interval of (0.05, 0.13) than the proportion of always survivors (observed or censored) six months post treatment initiation whose cancer would progress on docetaxal but would not progress on mitoxantrone. Statistical evidence suggests that the proportion of always survivors three months post treatment initiation whose cancer would not progress on docetaxal but would on mitoxantrone is at least 0.22 with a 95 percent confidence interval of (0.15, 0.29). Also, statistical evidence suggests that the proportion of always survivors six months post treatment initiation whose cancer would not progress on docetaxal but would on mitoxantrone is at least 0.09 with a 95 percent confidence interval of (0.02, 0.17). In order for the inference of detecting always survivors six months post treatment initiation whose cancer would not progress on docetaxal but would on mitoxantrone to be invalid, we would need at least 9 percent of our study population at baseline to follow the counterfactual response $S_1(\omega, t) = 0$, $S_0(\omega, t) = 1$, and $Y_1(\omega, t) = 2$, $Y_0(\omega, t) = 0$ for $t = 5$, that is at least 9 percent of our population would need to not survive if they were put on the new treatment but would survive on the old treatment and their cancer would not progress by six months post treatment initiation.

7.3 Data Analysis with Monotonicity Assumption $S^2_t(\omega, t) + S^1_t(\omega, t) \leq 1$

For each $t \in \{1, \ldots, 7\}$, we test the null hypothesis $P_{t, \omega}^{(1)} \leq P_{t, \omega}^{(2)} + P_{t, \omega}^{(3)}$ to evaluate for a fixed $t$ whether the proportion of individuals for whom $Y_1(\omega, t) = 0$, $Y_0(t, \omega) = 1$, $S^1_t(\omega, t) = 1$, and $S^0_t(\omega, t) = 1$ is greater than the proportion of individuals for whom $Y_1(\omega, t) = 1$, $Y_0(t, \omega) = 0$, $S^1_t(\omega, t) = 1$, and $S^0_t(\omega, t) = 1$ under the assumption that $S^2_t(\omega, t) + S^1_t(\omega, t) \leq 1$ for all $\omega \in \Omega$ and $t \in T_{SWOG}$. The following table provides 95% and 99% confidence intervals for $P_{t, \omega}^{(1)} - P_{t, \omega}^{(2)} - P_{t, \omega}^{(3)}$ for each $t \in T_{SWOG}$. Notice $P(S(t) = 2 \mid X = 1) + P(S(t) = 1 \mid X = 0) \leq 1$ for all $t \in T_{SWOG}$, though of course this does not guarantee that the assumption holds.

Two-sided confidence intervals of $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P\{Y(t) = 1, S(t) = 1\} \cup \{Y(t) = 3, S(t) = 2\} \mid X = 1 \leq 0$ and one-sided p-values for Wald test of the null hypothesis $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P\{Y(t) = 1, S(t) = 1\} \cup \{Y(t) = 3, S(t) = 2\} \mid X = 1 \leq 0$ are provided below:
Under the assumption of monotonicity for censoring, statistical evidence suggests that the proportion of always survivors three months post treatment initiation whose cancer would not progress on docetaxal but would on mitoxantrone is at least 0.14 with a 95 percent confidence interval of (0.14, 0.28). We do not find evidence for such individual level always survivor causal effects six months post treatment initiation.

### 7.4 Data Analysis with Monotonicity Assumption $S_1^{[2]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1$ and $S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1$

For each $t \in \{1, \ldots, 7\}$, we test the null hypothesis $P_{1,1,0}^{(t)} \leq P_{1,1,1}^{(t)}$ to evaluate for a different fixed $t$ whether the proportion of individuals for whom $Y_1(\omega, t) = 0$, $Y_0(\omega, t) = 1$, $S_1(\omega, t) = 1$, and $S_0(\omega, t) = 1$ is greater than the proportion of individuals for whom $Y_1(\omega, t) = 1$, $Y_0(\omega, t) = 0$, $S_1(\omega, t) = 1$, and $S_0(\omega, t) = 1$ under the assumption that $S_1^{[2]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1$ and $S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1$ for all $\omega \in \Omega$ and $t \in T_{SWOG}$. The following table provides 95% and 99% confidence intervals for $P_{1,1,0}^{(t)} - P_{1,1,1}^{(t)}$ for each $t \in T_{SWOG}$. As before, notice $P(S(t) = 0 \mid X = 1) + P(S(t) = 1 \mid X = 0) \leq 1$ and $P(S(t) = 2 \mid X = 1) + P(S(t) = 1 \mid X = 0) \leq 1$ for all $t \in T_{SWOG}$ and therefore we fail to falsify the monotonicity assumption, though of course this does not guarantee that the assumption holds.

Two-sided confidence intervals of $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P(Y(t) = 1, S(t) = 1 \mid X = 1)$ and one-sided p-values for Wald test of the null hypothesis $P(Y(t) = 1, S(t) = 1 \mid X = 0) - P(Y(t) = 1, S(t) = 1 \mid X = 1) \leq 0$ are provided below:

| 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|---------|---------|---------|---------|---------|---------|---------|
| Estimate | 0.09    | 0.20    | 0.25    | 0.19    | 0.14    | -0.04    | -0.02    |
| 95% CI   | (0.05, 0.13) | (0.14, 0.26) | (0.18, 0.31) | (0.12, 0.26) | (0.07, 0.22) | (-0.12, 0.03) | (-0.10, 0.05) |
| 99% CI   | (0.04, 0.14) | (0.12, 0.27) | (0.15, 0.34) | (0.10, 0.29) | (0.04, 0.24) | (-0.14, 0.06) | (-0.12, 0.07) |
| p-value  | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | 0.0001 | 0.831 | 0.736 |

Table 7: Results Monotonicity $S_1^{[2]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1$ and $S_1^{[0]}(\omega, t) + S_0^{[1]}(\omega, t) \leq 1$ for all $\omega \in \Omega$
8 Conclusion

This work demonstrates that testing for individual causal effects within ‘always survivor’ principal stratum can be conducted with no stronger identification assumptions than testing for a non-zero total effect. The identifiability assumptions we use to identify these individual level always survivor causal effects are no stronger than randomization. This embeds testing for individual level always survivor causal effects firmly within the Neyman-Pearson paradigm. We also enable statistician to assess the existence of such individual always survivor causal effects using randomization based and Bayesian inference. Previous literature required stronger assumptions to identify or provide bounds for survivor average causal effect, which is a different causal estimand of interest. We discuss the similarity between our Theorems on detecting individual level always survivor causal effects and previous literature on detecting individual level causal effects for sufficient cause interaction [15, 8, 14] and the binary instrumental variable model [1, 2, 9, 12] in the online supplement. Our results are generally applicable to clinical trials in the situation where an outcome of interest is potentially not observed when death of the individual occurs prior to the end of study. Our identifiability assumptions are guaranteed by design in a randomized trial. To the best of our knowledge, this is the first time causal effects are detected in always survivors without any assumption regarding the censoring mechanism.

Our proposed methodology provides a procedure to evaluate whether a new treatment would delay cancer progression. The effect of a new treatment on overall survival is a different scientific question than whether new treatment delays cancer progression. Our data analysis demonstrates the usefulness of our methods in this setting. In oncology trials, determining treatment effectiveness sooner rather than later can extend and improve patient lives as the progression to mortality for many type III cancers can be rapid. Our methodology is the first and, as of yet, only proposed procedure that enables detecting causal effects in the presence of truncation by death and censoring using only the assumptions that are guaranteed by design of the clinical trial.

Acknowledgement

The authors thank Zack McCaw and Ryan Sun for generating a dataset used in Section 6 from Kaplan-Meir curves provided in the New England Journal of Medicine Article, ‘Docetaxal and Estramustine compared with Mitoxantrone and Prednisone for Advanced Refractory Prostate Cancer,’ using software. This research was funded by the National Institutes of Health.

Appendix

Let \( I(\cdot) \) denote the usual indicator function. Define \( Y_x[0](\omega, t) = I(Y_x(\omega, t) = y) \), and \( s_x[0](\omega, t) = I(S_x(t, \omega) = s) \) for \( y \in \{0, 1, 2, 3\} \), \( s \in \{0, 1, 2\} \), and \( t \in T \). Also note \( Y_x[0](\omega, t)S_x[0](\omega, t) = I(Y_x(\omega, t) = y, S_x(\omega, t) = s) \) as a property of indicator functions. For ease of notation and space considerations, we drop the \( \omega \) in \( Y(\omega, t), S(\omega, t), S_x(\omega, t) \), whenever the meaning is clear. The main paper and this supplement assumes deterministic counterfactuals. The online supplement provides the same results with stochastic counterfactuals and derives the same results. The online supplement also provides results for the situation where there exists individuals \( \omega \) whose variable \( S(\omega, t) \) observed at time \( t \), but the outcome \( Y(\omega, t) \) is censored.

The proofs below assume deterministic potential outcomes for ease of exposition. The online supplement provides the same proofs with stochastic counterfactuals.

Proof of Theorem 1

We prove the contrapositive. Assume that no individual \( \omega \) of response type \( Y_1(\omega, t) = y \), \( Y_0(\omega, t) = 1 - y \), \( S_1(\omega, t) = 1 \) and \( S_0(\omega, t) = 1 \) exists in our population for a fixed integer \( t \in T \) and \( y \in \{0, 1\} \). Then for all individuals \( \omega \) in our population \( \Omega \), \( I(Y_1(\omega, t) = y, S_1(\omega, t) = 1) + I(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1) \leq 1 \). Taking expectations,

\[
P(Y_1(\omega, t) = y, S_1(\omega, t) = 1) + P(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1) \leq 1 \iff \]
\[
P(Y_1(\omega, t) = y, S_1(\omega, t) = 1 \mid X = 0) + P(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1 \mid X = 1) \leq 1 \iff \]
\[
P(Y(t) = y, S(t) = 1 \mid X = 0) - P(Y(t) = 1 - y, S(t) = 1 \mid X = 1) \leq 1.
\]

The first to second line follows from \( (Y_x, S_x) \parallel X \) and the second to third line follows from consistency of counterfactuals.
Proof of Proposition 1

Taking expectation of \( Y_1^{[1]}(\omega, t)S_1^{[1]}(\omega, t) + Y_0^{[0]}(\omega, t)S_0^{[1]}(\omega, t) - 1 \) and \( Y_1^{[0]}(\omega, t)S_1^{[1]}(\omega, t) + Y_0^{[1]}(\omega, t)S_1^{[1]}(\omega, t) - 1 \) gives the result. A table with the relevant frequencies of the counterfactuals is provided below. Set the counterfactual probabilities that do not satisfy the monotonicity assumptions to zero.

Proof of Corollary 1

Set all counterfactual probabilities in Table 8 that do not satisfy the required monotonicity assumptions to zero. Consider, from arguments presented in Theorem 1 and Proposition 1,

\[
P_r^{(t)}(y, 1, 1, 1, 0) - 1 = E[I(Y_1(\omega, t) = y, S_1(\omega, t) = 1) + I(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1) - 1] \\
= P_r^{(t)}(y, 1 - y, 1, 1) - P_r^{(t)}(1 - y, y, 1, 1) - P_r^{(t)}(2, y, 0, 1) \\
- P_r^{(t)}(1 - y, 2, 1, 0) - P_r^{(t)}(2, 2, 0, 0) - P_r^{(t)}(3, 3, 2, 2) \\
- P_r^{(t)}(3, y, 2, 1) - P_r^{(t)}(3, 2, 2, 0) - P_r^{(t)}(1 - y, 3, 1, 2) \\
- P_r^{(t)}(2, 3, 0, 2).
\]

Therefore,

\[
P_r^{(t)}(y, 1 - y, 1, 1) - P_r^{(t)}(1 - y, y, 1, 1) - P_r^{(t)}(3, y, 2, 1) \\
- P_r^{(t)}(1 - y, 3, 1, 2) - P_r^{(t)}(3, 3, 2, 2) \\
= P_{r, y, 1, 1}^{(t)} + P_{1 - r, 1, 0}^{(t)} - 1 \\
+ P_r^{(t)}(2, y, 0, 1) + P_r^{(t)}(1 - y, 2, 1, 0) \\
+ P_r^{(t)}(2, 2, 0, 0) \\
+ P_r^{(t)}(3, 2, 2, 0) + P_r^{(t)}(2, 3, 0, 2).
\]

Consequently,

\[
P_{r, y, 1, 1}^{(t)} + P_{1 - r, 1, 0}^{(t)} - 1 < P_r^{(t)}(y, 1 - y, 1, 1) - P_r^{(t)}(1 - y, y, 1, 1) \\
- P_r^{(t)}(3, y, 2, 1) - P_r^{(t)}(1 - y, 3, 1, 2) \\
- P_r^{(t)}(3, 3, 2, 2).
\]

This demonstrates \( P_{r, y, 1, 1}^{(t)} + P_{1 - r, 1, 0}^{(t)} - 1 \) serves as a lower bound for the counterfactual contrast

\[
P_r^{(t)}(1, 0, 1, 1) - \left[ P_r^{(t)}(0, 1, 1, 1) + P_r^{(t)}(3, 1, 2, 1) + P_r^{(t)}(3, 3, 2, 2) + P_r^{(t)}(0, 3, 1, 2) \right].
\]

The first term in this counterfactual contrast, \( P_r^{(t)}(1, 0, 1, 1) \), is a lower bound of the proportion of individuals (observed or censored) that are always survivors and for whom the treatment causes the outcome, because it does not include any censored individuals that are also always survivors and for whom the treatment causes the outcome. The term in the square brackets is an upper bound on the always survivors (observed or censored) for whom the treatment prevents the outcome, because it effectively treats any censored individual that could possibly be an always survivor and for whom the treatment prevents the outcome is actually an always survivor and for whom the treatment prevents the outcome. To see that \( P_{r, y, 1, 1}^{(t)} + P_{0, 1, 0}^{(t)} - 1 \) serves as a lower bound to \( P_r^{(t)}(1, 0, 1, 1) \), note

\[
P_{r, y, 1, 1}^{(t)} + P_{0, 1, 0}^{(t)} - 1 \leq P_r^{(t)}(1, 0, 1, 1) - \left[ P_r^{(t)}(0, 1, 1, 1) + P_r^{(t)}(3, 1, 2, 1) + P_r^{(t)}(3, 3, 2, 2) + P_r^{(t)}(0, 3, 1, 2) \right] \implies \\
P_{r, y, 1, 1}^{(t)} + P_{0, 1, 0}^{(t)} - 1 \leq P_r^{(t)}(1, 0, 1, 1).
\]

The proofs of the other bounds are similar. We complete our proof.
Proof of Theorem 2A

We prove the contrapositive. For a fixed integer \( t \in T \), assume that no individual \( \omega \) of response type \( Y_1(\omega,t) = 1 \), \( Y_0(\omega,t) = 0 \), \( S_1(\omega,t) = 1 \) and \( S_0(\omega,t) = 1 \) exists in our population. Then for all individuals \( \omega \) in our population \( \Omega \), \( I(Y_0(\omega,t) = 0, S_0(\omega,t) = 1) - I(Y_1(\omega,t) = 0, S_1(\omega,t) = 1) - I(Y_1(\omega,t) = 3, S_1(\omega,t) = 2) \leq 0 \). This last assertion is true after examining the counterfactual table provided in tables 2.1-2.3 in the online supplement. Taking expectations, we have
\[
P(Y_0(\omega,t) = 0, S_0(\omega,t) = 1) - P(Y_1(\omega,t) = 0, S_1(\omega,t) = 1) - P(Y_1(\omega,t) = 3, S_1(\omega,t) = 2) \leq 0 \quad \iff \quad P(Y_0(t) = 0, S_0(t) = 1 \mid X = 0) - P(Y_1(t) = 0, S_1(t) = 1 \mid X = 1) - P(Y_1(t) = 3, S_1(t) = 2 \mid X = 1) \leq 0 \quad \iff \quad P(Y(t) = 0, S(t) = 1 \mid X = 0) - P(Y(t) = 0, S(t) = 1 \mid X = 1) - P(Y(t) = 3, S(t) = 2 \mid X = 1) \leq 0.
\]
The first to second line follows from \((Y_s, S_s)\) \(\Pi X\) and the second to third line follows from consistency of counterfactuals. The other inequality associated with Theorem 2A can be derived similarly. The proofs of Theorem 2B and 2C can be derived similarly.

Proof of Proposition 2A

Taking expectation of \( Y_0[1](\omega,t)S_0[1](\omega,t)-Y_1[1](\omega,t)S_1[1](\omega,t)-Y_1[3](\omega,t)S_1[2](\omega,t)\) and \(Y_0[0](\omega,t)S_0[1](\omega,t)-Y_1[0](\omega,t)S_1[1](\omega,t)-Y_1[3](\omega,t)S_1[2](\omega,t)\) gives the result. A table with the relevant frequencies of the counterfactuals is provided below. In Table 8, set any probability that does not satisfy the monotonicity assumption \( S_1[1](\omega,t)+S_0[0](\omega,t) \leq 1 \) to zero. The proofs of Proposition 2B and 2C can be derived similarly.

Proof of Corollary 2A

From Table 8,
\[
P_t^{(t)}(1-y,1.0) - P_t^{(t)}(1-y,1.1) - P_t^{(t)}(3,2.1)
= E[I(Y_0(\omega,t) = 1 - y, S_0(\omega,t) = 1) - I(Y_1(\omega,t) = 1 - y, S_1(\omega,t) = 1) - I(Y_1(\omega,t) = 3, S_1(\omega,t) = 2)]
= P_t^{(t)}(y,1-y,1,1) - P_t^{(t)}(1-y,y,1,1)
- P_t^{(t)}(1-y,2,1,0) - P_t^{(t)}(3,3,2,2)
- P_t^{(t)}(3,y,2,1) - P_t^{(t)}(3,2,2,0) - P_t^{(t)}(1-y,3,1,2).
\]

Therefore,
\[
P_t^{(t)}(y,1-y,1,1) - P_t^{(t)}(1-y,y,1,1)
- P_t^{(t)}(3,y,2,1) - P_t^{(t)}(1-y,3,1,2) + P_t^{(t)}(3,3,2,2)
= P_t^{(t)}(1-y,1.0) - P_t^{(t)}(1-y,1.1) - P_t^{(t)}(3,2.1)
+ P_t^{(t)}(1-y,2,1,0) + P_t^{(t)}(3,2,2,0).
\]

Consequently,
\[
P_t^{(t)}(1-y,1.0) - P_t^{(t)}(1-y,1.1) - P_t^{(t)}(3,2.1) < P_t^{(t)}(y,1-y,1,1) - P_t^{(t)}(1-y,y,1,1)
- P_t^{(t)}(3,y,2,1) - P_t^{(t)}(1-y,3,1,2)
- P_t^{(t)}(3,3,2,2).
\]

This demonstrates \( P_t^{(t)}(1,y,0,1,1) - P_t^{(t)}(0,1,1,1) - P_t^{(t)}(3,1,2,1) + P_t^{(t)}(3,3,2,2) + P_t^{(t)}(0,3,1,2) \) serves as a lower bound for the counterfactual contrast
\[
P_t^{(t)}(1,0,1,1) - \left[ P_t^{(t)}(0,1,1,1) + P_t^{(t)}(3,1,2,1) + P_t^{(t)}(3,3,2,2) + P_t^{(t)}(0,3,1,2) \right].
\]

The first term in this counterfactual contrast, \( P_t^{(t)}(1,0,1,1) \), is a lower bound of the proportion of individuals (observed or censored) that are always survivors and for whom the treatment causes the outcome, because it does not include any censored individuals that are also always survivors and for whom the treatment causes the outcome.
The term in the square brackets is an upper bound on the always survivors (observed or censored) for whom the treatment prevents the outcome, because it effectively treats any censored individual that could possibly be an always survivor and for whom the treatment prevents the outcome. To see that \( P_y^c(t) = P_0^c(t) - P_1^c(t) - P_y^r(t) \) serves as a lower bound to \( P_y^c(t) \), note

\[
P_y^c(t) - P_y^r(t) - P_y^r(t) \leq P_y^c(t)(1, 0, 1, 1) - \left[ P_y^c(t)(0, 1, 1, 1) + P_y^c(t)(3, 1, 2, 1) + P_y^c(t)(3, 3, 2, 2) + P_y^c(t)(0, 3, 1, 2) \right] 
\implies P_y^c(t) - P_y^r(t) - P_y^r(t) \leq P_y^c(t)(1, 0, 1, 1). 
\]

The proofs of the other bounds are similar. We complete our proof. The proofs of Corollary 2B and 2C can be derived similarly.

**Proof of Theorem 3A**

From Table 8, we have under consistency of counterfactuals and baseline randomization,

\[
P_{1-y,1.0} - P_{1-y,1.1} - P_{3,2,1}^{r(t)} \\
= E[I(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1) - I(Y_1(\omega, t) = 1 - y, S_1(\omega, t) = 1) - I(Y_1(\omega, t) = 3, S_1(\omega, t) = 2)] \\
= P_y^c(t)(y, 1 - y, 1, 1) - P_y^c(t)(1 - y, y, 1, 1) \\
- P_y^c(t)(1 - y, 2, 1, 0) - P_y^c(t)(3, 3, 2, 2) \\
- P_y^c(t)(3, y, 2, 1) - P_y^c(t)(3, 2, 2, 0) - P_y^c(t)(1 - y, 3, 1, 2) \\
+ P_y^c(t)(2, 1 - y, 0, 1).
\]

Therefore,

\[
P_y^c(t)(y, 1 - y, 1, 1) - P_y^c(t)(1 - y, y, 1, 1) \\
- P_y^c(t)(3, y, 2, 1) - P_y^c(t)(1 - y, 3, 1, 2) + P_y^c(t)(3, 3, 2, 2) \\
= P_{1-y,1.0} - P_{1-y,1.1} - P_{3,2,1}^{r(t)} \\
+ P_y^c(t)(1 - y, 2, 1, 0) + P_y^c(t)(3, 2, 2, 0) - P_y^c(t)(2, 1 - y, 0, 1).
\]

Consequently, if

\[
P_{1-y,1.0} - P_{1-y,1.1} - P_{3,2,1}^{r(t)} > d_m(t),
\]

where

\[
d_m(t) = P_y^c(t)(2, 1 - y, 0, 1) - P_y^c(t)(1 - y, 2, 1, 0) - P_y^c(t)(2, 1 - y, 0, 1),
\]

then,

\[
P_{1-y,1.0} - P_{1-y,1.1} - P_{3,2,1}^{r(t)} \\
> P_y^c(t)(y, 1 - y, 1, 1) - P_y^c(t)(1 - y, y, 1, 1) \\
- P_y^c(t)(3, y, 2, 1) - P_y^c(t)(1 - y, 3, 1, 2) - P_y^c(t)(3, 3, 2, 2).
\]

We therefore complete our proof.

**Proof of Theorem 3B**

From Table 8, we have under consistency of counterfactuals and baseline randomization

\[
P_{1-y,1.0} - P_{1-y,1.1} - P_{2,0.1}^{r(t)} \\
= E[I(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1) - I(Y_1(\omega, t) = 1 - y, S_1(\omega, t) = 1) - I(Y_1(\omega, t) = 2, S_1(\omega, t) = 0)] \\
= P_y^c(t)(y, 1 - y, 1, 1) - P_y^c(t)(1 - y, y, 1, 1) \\
- P_y^c(t)(1 - y, 2, 1, 0) - P_y^c(t)(2, 2, 0, 0) \\
- P_y^c(t)(1 - y, 3, 1, 2) - P_y^c(t)(2, 3, 0, 2) \\
+ P_y^c(t)(3, 1 - y, 2, 1) - P_y^c(t)(2, y, 0, 1).
\]
Therefore,
\[ P^{c(t)}(y, 1 - y, 1, 1) - P^{c(t)}(1 - y, y, 1, 1) \]
\[ = P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} - P^{r(t)}_{2,0.1} \]
\[ - P^{c(t)}(3, 1 - y, 2, 1) + P^{c(t)}(1 - y, 2, 1, 0) + P^{c(t)}(1 - y, 3, 1, 2) \]
\[ + P^{c(t)}(2, 2, 0, 0) + P^{c(t)}(2, 3, 0, 2) + P^{c(t)}(2, y, 0, 1) \]

Consequently, if
\[ P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} > c_m(t), \]
where
\[ c_m(t) = P^{c(t)}(3, 1 - y, 2, 1) - P^{c(t)}(1 - y, 2, 1, 0) - P^{c(t)}(1 - y, 3, 1, 2) \]
\[ - P^{c(t)}(2, y, 0, 1) - P^{c(t)}(2, 2, 0, 0) - P^{c(t)}(2, 3, 0, 2). \]

then,
\[ P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} > P^{c(t)}(y, 1 - y, 1, 1) - P^{c(t)}(1 - y, y, 1, 1). \]

We therefore complete our proof.

**Proof of Theorem 3C**

From Table 8, we have under consistency of counterfactuals and baseline randomization,
\[ P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} \]
\[ = E[I(Y_0(\omega, t) = 1 - y, S_0(\omega, t) = 1) - I(Y_1(\omega, t) = 1 - y, S_1(\omega, t) = 1)] \]
\[ = P^{c(t)}(y, 1 - y, 1, 1) - P^{c(t)}(1 - y, y, 1, 1) \]
\[ - P^{c(t)}(1 - y, 3, 1, 2) - P^{c(t)}(1 - y, 2, 1, 0) \]
\[ + P^{c(t)}(3, 1 - y, 2, 1) + P^{c(t)}(2, 1 - y, 0, 1). \]

Therefore,
\[ P^{c(t)}(y, 1 - y, 1, 1) - P^{c(t)}(1 - y, y, 1, 1) \]
\[ = P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} \]
\[ + P^{c(t)}(1 - y, 2, 1, 0) + P^{c(t)}(1 - y, 3, 1, 2) - P^{c(t)}(2, 1 - y, 0, 1) - P^{c(t)}(3, 1 - y, 2, 1). \]

Consequently, if
\[ P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} > k_m(t), \]
where
\[ k_m(t) = P^{c(t)}(2, 1 - y, 0, 1) + P^{c(t)}(3, 1 - y, 2, 1) - P^{c(t)}(1 - y, 2, 1, 0) - P^{c(t)}(1 - y, 3, 1, 2) \]

then,
\[ P^{c(t)}(y, 1 - y, 1, 1) - P^{c(t)}(1 - y, y, 1, 1) \]
\[ = P^{r(t)}_{1-y,1.0} - P^{r(t)}_{1-y,1.1} \]
\[ > 0 \]

We therefore complete our proof.
Associated Tables

The tables drop the explicit dependence on \( \omega \) and integer \( t \in T \), and use the shorthand \( Y_x(\omega, t) = Y_x \) and \( S_x(\omega, t) = S_x \) for \( x \in \{0, 1\} \) out of space considerations. Recall that we denote \( Y_x^{[y]}(\omega, t) = I(Y_x(\omega, t) = y) \), and \( S_x^{[s]}(\omega, t) = I(S_x(\omega, t) = s) \) for \( y \in \{0, 1, 2, 3\} \) and \( s \in \{0, 1, 2\} \). Likewise, we drop the dependence on \( \omega \) and \( t \), and use shorthand notation \( S_x^{[s]} \) and \( Y_x^{[y]} \) instead of \( S_x^{[s]}(\omega, t) \) and \( Y_x^{[y]}(\omega, t) \) for \( y \in \{0, 1\} \) out of space considerations. The reader should always assume that all counterfactuals presented in the tables below are implicitly dependent on \( \omega \) and \( t \) even if the shorthand notation does not explicitly indicate such a dependence. The tables below provide a complete enumeration of the different counterfactual response types. Without monotonicity assumptions there are 16 different counterfactual response types, and then with the monotonicity assumption that \( S_x^{[0]}(\omega, t) + S_x^{[1]}(\omega, t) \leq 1 \) for all individuals in our population, the response types \( \omega \in \{5, 6\} \) are no longer possible. Under monotonicity assumption that \( S_x^{[2]}(\omega, t) + S_x^{[1]}(\omega, t) \leq 1 \) response types \( \omega \in \{11, 12\} \) in Table 8 are no longer possible. All of our results can be derived using Table 8 only, but we provide further tables in the online supplement that could aid readers. The table below is for any fixed \( t \in T \).

| \( \omega \) | \( Y_1 \) | \( Y_0 \) | \( S_1 \) | \( S_0 \) | Probability |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | \( P(Y_1 = 1, Y_0 = 1, S_1 = 1, S_0 = 1) \) |
| 2 | 0 | 1 | 1 | 1 | \( P(Y_1 = 0, Y_0 = 1, S_1 = 1, S_0 = 1) \) |
| 3 | 1 | 0 | 1 | 1 | \( P(Y_1 = 1, Y_0 = 0, S_1 = 1, S_0 = 1) \) |
| 4 | 0 | 0 | 1 | 1 | \( P(Y_1 = 0, Y_0 = 0, S_1 = 1, S_0 = 1) \) |
| 5 | 2 | 1 | 0 | 1 | \( P(Y_1 = 2, Y_0 = 1, S_1 = 0, S_0 = 1) \) |
| 6 | 2 | 0 | 0 | 1 | \( P(Y_1 = 2, Y_0 = 0, S_1 = 0, S_0 = 1) \) |
| 7 | 1 | 2 | 1 | 0 | \( P(Y_1 = 1, Y_0 = 2, S_1 = 1, S_0 = 0) \) |
| 8 | 0 | 2 | 1 | 0 | \( P(Y_1 = 0, Y_0 = 2, S_1 = 1, S_0 = 0) \) |
| 9 | 2 | 2 | 0 | 0 | \( P(Y_1 = 2, Y_0 = 2, S_1 = 0, S_0 = 0) \) |
| 10 | 3 | 3 | 2 | 2 | \( P(Y_1 = 3, Y_0 = 3, S_1 = 2, S_0 = 2) \) |
| 11 | 3 | 1 | 2 | 1 | \( P(Y_1 = 3, Y_0 = 1, S_1 = 2, S_0 = 1) \) |
| 12 | 3 | 0 | 2 | 1 | \( P(Y_1 = 3, Y_0 = 0, S_1 = 2, S_0 = 1) \) |
| 13 | 3 | 2 | 2 | 0 | \( P(Y_1 = 3, Y_0 = 2, S_1 = 2, S_0 = 0) \) |
| 14 | 1 | 3 | 1 | 2 | \( P(Y_1 = 1, Y_0 = 3, S_1 = 1, S_0 = 2) \) |
| 15 | 0 | 3 | 1 | 2 | \( P(Y_1 = 0, Y_0 = 3, S_1 = 1, S_0 = 2) \) |
| 16 | 2 | 3 | 0 | 2 | \( P(Y_1 = 2, Y_0 = 3, S_1 = 0, S_0 = 2) \) |

Table 8: Counterfactual Distribution

9 Data

A total of 338 individuals were randomized to Docetaxal and 336 individuals were randomized to Mitoxantrone. The full dataset is provided here to enable reproducibility of our results.

9.1 Survive and Cancer Progression

| Time | \( X = 1 \) | \( X = 0 \) |
|---|---|---|
| 1 month | 6 | 37 |
| 2 months | 28 | 94 |
| 3 months | 60 | 142 |
| 4 months | 81 | 145 |
| 6 months | 108 | 155 |
| 12 months | 143 | 129 |
| 18 months | 102 | 93 |

Table 9: Survive and cancer progression at time \( t \) \( Y(t) = 1, S(t) = 1 \)
9.2 Survive and Cancer Did not Progress

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| \( X = 1 \) | 328     | 297      | 257      | 226      | 176      | 75        | 20        |
| \( X = 0 \) | 293     | 227      | 170      | 154      | 116      | 56        | 15        |

Table 10: Survive and cancer does not progress at time \( t \), \( (Y(t) = 0, S(t) = 1) \)

9.3 Did not Survive

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| \( X = 1 \) | 2       | 8        | 13       | 20       | 38       | 87        | 155       |
| \( X = 0 \) | 4       | 10       | 16       | 26       | 48       | 116       | 176       |

Table 11: Did not survive until \( t \), \( (Y(t) = 2, S(t) = 0) \)

9.4 Censored

|       | 1 month | 2 months | 3 months | 4 months | 6 months | 12 months | 18 months |
|-------|---------|----------|----------|----------|----------|-----------|-----------|
| \( X = 1 \) | 2       | 5        | 8        | 11       | 16       | 33        | 61        |
| \( X = 0 \) | 2       | 5        | 8        | 11       | 17       | 35        | 52        |

Table 12: Censored at time \( t \), \( (Y(t) = 3, S(t) = 2) \)

References

[1] Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997.

[2] Zhihong Cai, Manabu Kuroki, Judea Pearl, and Jin Tian. Bounds on direct effects in the presence of confounded intermediate variables. *Biometrics*, 64(3):695–701, 2008.

[3] Peng Ding, Zhi Geng, Wei Yan, and Xiao-Hua Zhou. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association*, 106(496):1578–1591, 2011.

[4] Constantine E Frangakis and Donald B Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

[5] Patricia Guyot, AE Ades, Mario JNM Ouwens, and Nicky J Welton. Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves. *BMC medical research methodology*, 12(1):9, 2012.

[6] Kosuke Imai. Sharp bounds on the causal effects in randomized experiments with “truncation-by-death”. *Statistics & probability letters*, 78(2):144–149, 2008.

[7] Daniel P Petrylak, Catherine M Tangen, Maha HA Hussain, Primo N Lara Jr, Jeffrey A Jones, Mary Ellen Taplin, Patrick A Burch, Donna Berry, Carol Moinpour, Manish Kohli, et al. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *New England Journal of Medicine*, 351(15):1513–1520, 2004.

[8] Roland R Ramsahai. Probabilistic causality and detecting collections of interdependence patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):705–723, 2013.
[9] Thomas S Richardson and James M Robins. Analysis of the binary instrumental variable model. *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, pages 415–444, 2010.

[10] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512, 1986.

[11] Donald B Rubin et al. Causal inference through potential outcomes and principal stratification: application to studies with “censoring” due to death. *Statistical Science*, 21(3):299–309, 2006.

[12] Sonja A Swanson, Miguel A Hernán, Matthew Miller, James M Robins, and Thomas Richardson. Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes. *Journal of the American Statistical Association*, (just-accepted), 2018.

[13] Eric J Tchetgen Tchetgen. Identification and estimation of survivor average causal effects. *Statistics in medicine*, 33(21):3601–3628, 2014.

[14] Tyler J VanderWeele, Thomas S Richardson, et al. General theory for interactions in sufficient cause models with dichotomous exposures. *The Annals of Statistics*, 40(4):2128–2161, 2012.

[15] Tyler J VanderWeele and James M Robins. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*, 95(1):49–61, 2008.

[16] Linbo Wang, Thomas S Richardson, and Xiao-Hua Zhou. Causal analysis of ordinal treatments and binary outcomes under truncation by death. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):719–735, 2017.

[17] Linbo Wang, Xiao-Hua Zhou, and Thomas S Richardson. Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*, 104(3):597–612, 2017.

[18] Junni L Zhang and Donald B Rubin. Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by “Death”. *Journal of Educational and Behavioral Statistics*, 28(4):353–368, 2003.