Journal of Biometrics & Biostatistics
ISSN: 2155-6180
Chiba et al., J Biomet Biostat 2011, 2:5
DOI: 10.4172/2155-6180.1000e104
Editorial
Open Access

On the Identification of the Survivor Average Causal Effect

Yasutaka Chiba[1]*, Masataka Taguri[2] and Yukari Uemura[3]

[1]Division of Biostatistics, Clinical Research Center, Kinki University School of Medicine, Osaka, Japan
[2]Department of Biostatistics and Epidemiology, Graduate School of Medicine, Yokohama City University, Yokohama, Japan
[3]Department of Biostatistics/Epidemiology and Preventive Health Sciences, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

In randomized trials in which the outcome requires considerable follow-up, participants may die before the trial is complete. In such cases, for the individuals who die before follow-up is complete, the outcome is not simply missing, but is undefined. Some authors refer to this situation as one in which the outcome is "truncated by death" [1,2], to distinguish this scenario from cases in which the outcome is merely missing because of inadequate data collection. In these settings, a crude comparison of the outcome between those who survived in each treatment arm may give misleading results, because we no longer preserve randomization by conditioning on a post-treatment event (survival) and thus the crude comparison is not a comparison for the same population comparing different treatments, but a comparison of different populations.

A treatment comparison that makes sense in this setting would be to ask how the outcome differs between treated and untreated individuals in the subpopulation who would have survived under either arm. This effect is sometimes referred to as the survivor average causal effect (SACE) [3] or the principal strata effect [4]. We can circumvent the problem with the crude comparison by restricting the comparison to this subpopulation but, unfortunately, this subpopulation of interest is not identified. Trying to identify and estimate the SACE from the observed data is a subject to challenge. This editorial reviews the reported identifiability assumptions and gives their understandable interpretations, and then outlines a view of future research in this area.

We use the following notation. Suppose that A denotes the binary treatment variable (A = 1 for the treatment arm and A = 0 for the control arm), Y denotes an outcome of interest that is measured after some follow-up period, and S denotes an indicator of whether the individual survives (S = 1 if alive and S = 0 if dead). For individuals who died (S = 0), Y is undefined. For each individual, we can also consider potential outcomes [5] corresponding to what would have happened had an individual been in the arm other than the one they were, in fact, in. Let S_a denote the survival status if the individual had been in $A = a$, and let Y_a denote the outcome if the individual had been in $A = a$. The variable Y_a is defined only if $S_a = 1$. Otherwise, the individual would have died and Y_a would be undefined. Here, we assume no-interference [6], i.e., that the outcome and survival status of an individual do not depend on the treatment status of other individuals. We also assume randomization of the treatment, in which $Y_a \coprod A$ and $S_a \coprod A$ [7], where $B \coprod C$ denotes that B is independent from C. These independencies also hold conditional on X denoting a set of baseline covariates, which do not affect A, but are confounders between S and Y.

Using the notation, in the difference term, a crude comparison of the outcome, which compares the means of Y in each treatment arm among those who in fact survived, is formalized as $\mathrm{E}(Y \mid A = 1, S = 1) - \mathrm{E}(Y \mid A = 0, S = 1)$, and the SACE is formalized as:

$$\mathrm{SACE} \equiv \mathrm{E}\left(Y_1 \mid S_1 = S_0 = 1\right) - \mathrm{E}\left(Y_0 \mid S_1 = S_0 = 1\right).$$

The SACE compares the outcome under the treatment versus the control arm, but for the subpopulation that would have survived under either treatment arm. A subpopulation such as this is referred to as a principal stratum [8]. Note that individuals can be classified into four principal strata, as shown in Table 1, where it is assumed that some always-survivors exist.

First, we discuss the case in which X is not used to estimate the SACE. To identify the SACE, Gilbert et al. [9] introduced the following two assumptions:

Assumption 1: $S_1 \geq S_0$ for all individuals.

Assumption 2: $Y_1 \coprod S_0 \mid S_1$.

Assumptions 1 and 2 are sometimes referred to as the assumptions of monotonicity and explainable nonrandom survival [3], respectively. Under these assumptions, the SACE is equivalent to the crude comparison:

$$\begin{aligned}\mathrm{SACE} &\equiv \mathrm{E}(Y_1 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_0 \mid S_1 = S_0 = 1) \\ &= \mathrm{E}(Y_1 \mid S_1 = 1) - \mathrm{E}(Y_0 \mid S_0 = 1) \\ &= \mathrm{E}(Y \mid A = 1, S = 1) - \mathrm{E}(Y \mid A = 0, S = 1), \end{aligned} \quad (1)$$

where the first term in the second equation is by Assumption 2 and the second term is by Assumption 1. The third equation holds by randomization of treatment.

Although it might be easy to verify that Assumptions 1 and 2 identify the SACE as shown in (1), it is difficult to understand what these assumptions mean. Therefore, we provide understandable interpretations for these assumptions. Assumption 1 implies that there is no individual with $(S_1, S_0) = (0, 1)$, because $S_1 = 0$ and $S_0 = 1$ cannot hold simultaneously under $S_1 \geq S_0$. Therefore, this assumption can be interpreted as the fact that no defier exists (see Table 1). Assumption 2 derives that $\mathrm{E}(Y_1 \mid S_1 = S_0 = 1) = \mathrm{E}(Y_1 \mid S_1 = 1)$ as shown in (1). Using this equation, $\mathrm{E}(Y_1 \mid S_1 = S_0 = 1) = \mathrm{E}(Y_1 \mid S_1 = 1, S_0 = 0)$ can be derived under $\Pr(S_1 = 1, S_0 = 0) > 0$, because

Principal stratum	Name	Survival status under treatment arm	Survival status under control arm
$(S_1, S_0) = (1, 1)$	Always-survivor	Alive	Alive
$(S_1, S_0) = (1, 0)$	Complier	Alive	Dead
$(S_1, S_0) = (0, 1)$	Defier	Dead	Alive
$(S_1, S_0) = (0, 0)$	Never-survivor	Dead	Dead

Table 1: Principal strata.

***Corresponding author:** Yasutaka Chiba, Division of Biostatistics, Clinical Research Center, Kinki University School of Medicine, 377-2, Ohno-higashi, Osakasayama, Osaka 589-8511, Japan, Tel: +81-72-366-0221; Fax: +81-72-368-1193; E-mail: chibay@med.kindai.ac.jp

Received December 22, 2011; **Accepted** December 22, 2011; **Published** December 22, 2011

Citation: Chiba Y, Taguri M, Uemura Y (2011) On the Identification of the Survivor Average Causal Effect. J Biomet Biostat 2:e104. doi:10.4172/2155-6180.1000e104

$$\mathrm{E}(Y_1 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_1 \mid S_1 = 1)$$

$$= \mathrm{E}(Y_1 \mid S_1 = S_0 = 1) - \sum_{j=0}^{1} \mathrm{E}(Y_1 \mid S_1 = 1, S_0 = j) \Pr(S_0 = j \mid S_1 = 1)$$

$$= \frac{\{\mathrm{E}(Y_1 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_1 \mid S_1 = 1, S_0 = 0)\} \Pr(S_1 = 1, S_0 = 0)}{p_1}, \quad (2)$$

where $p_a = \Pr(S = 1 \mid A = a)$. Therefore, Assumption 2 can be interpreted as the fact that the treatment arm distribution of outcome is the same in always-survivors and compliers (see Table 1), when at least a complier exists.

We can also identify the SACE without Assumption 1, by applying the following assumption, which is somewhat stronger than Assumption 2:

Assumption 3: $Y_a \coprod S_{1-a} \mid S_a$ for $a = 0, 1$.

For $a = 1$, this assumption is a restatement of Assumption 2. The SACE is equivalent to the crude comparison under Assumption 3, because $\mathrm{E}(Y_0 \mid S_1 = S_0 = 1) = \mathrm{E}(Y_0 \mid S_0 = 1)$ under Assumption 3 with $a = 0$ ($Y_0 \coprod S_1 \mid S_0$). A calculation similar to (2) gives

$$\mathrm{E}(Y_0 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_0 \mid S_0 = 1)$$

$$= \frac{\{\mathrm{E}(Y_0 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_0 \mid S_1 = 0, S_0 = 1)\} \Pr(S_1 = 0, S_0 = 1)}{p_0}. \quad (3)$$

Therefore, Assumption 3 can be interpreted as the fact that the treatment arm distribution of outcome is the same in always-survivors and compliers (with $a = 1$), and the control arm distribution of outcome is the same in always-survivors and defiers (with $a = 0$), when at least a complier and a defier exist.

Next, we discuss the case in which X is used to estimate the SACE. Hayden et al. [3] introduced the following two assumptions to identify the SACE:

Assumption 4: $S_a \coprod S_{1-a} \mid X$ for $a = 0, 1$.

Assumption 5: $Y_a \coprod S_{1-a} \mid \{S_a, X\}$ for $a = 0, 1$.

Assumption 4 leads to

$$\Pr(S_1 = i, S_0 = j \mid X = x)$$

$$= \Pr(S_1 = i \mid X = x) \Pr(S_0 = j \mid X = x)$$

$$= \Pr(S = i \mid A = 1, X = x) \Pr(S = j \mid A = 0, X = x) \quad (4)$$

for $i = 0, 1$ and $j = 0, 1$, and it implies that the probabilities of the principal strata can be estimated in the stratum of x. Assumption 5 is the same as Assumption 3 conditional on X, and thus its interpretation is the same as that of Assumption 3 conditional on X. Although Assumption 5 may be a weaker assumption than Assumption 3, an additional assumption is required to identify the SACE, because $\mathrm{E}(Y_a \mid S_1 = S_0 = 1, X = x)$ can be identified under Assumption 5 only, but

$$\mathrm{E}(Y_a \mid S_1 = S_0 = 1)$$

$$= \sum_x \mathrm{E}(Y_a \mid S_1 = S_0 = 1, X = x) \Pr(X = x \mid S_1 = S_0 = 1)$$

$$= \frac{\sum_x \mathrm{E}(Y_a \mid S_1 = S_0 = 1, X = x) \Pr(S_1 = S_0 = 1 \mid X = x) \Pr(X = x)}{\sum_x \Pr(S_1 = S_0 = 1 \mid X = x) \Pr(X = x)}$$

cannot be identified until $\Pr(S_1 = S_0 = 1 \mid X = x)$ is identified. Assumption 4 identifies this as shown in (4), and thus Assumptions 4 and 5 identify the SACE. See Hayden et al. [3] for the estimation.

Note that Assumptions 1 and 5 with $a = 1$ also identify the SACE, because $\Pr(S_1 = S_0 = 1 \mid X = x) = \Pr(S_0 = 1 \mid X = x) = \Pr(S = 1 \mid A = 0, X = x)$ under Assumption 1. This is also because

$$\mathrm{E}(Y_1 \mid S_1 = S_0 = 1)$$

$$= \mathrm{E}(Y_1 \mid A = 0, S = 1)$$

$$= \sum_x \mathrm{E}(Y_1 \mid A = 0, S = 1, X = x) \Pr(X = x \mid A = 0, S = 1)$$

$$= \sum_x \mathrm{E}(Y_1 \mid S_1 = S_0 = 1, X = x) \Pr(X = x \mid A = 0, S = 1)$$

$$= \sum_x \mathrm{E}(Y \mid A = 1, S = 1, X = x) \Pr(X = x \mid A = 0, S = 1),$$

where the first and third equations are by Assumption 1 and the fourth equation is by Assumption 5 with $a = 1$. As shown in (1), $\mathrm{E}(Y_0 \mid S_1 = S_0 = 1) = \mathrm{E}(Y \mid A = 0, S = 1)$ under Assumption 1. See Egleston et al. [10] and Chiba [11] for the estimation of the SACE.

This editorial reviewed the reported identifiability assumptions for the SACE and gave understandable interpretations for them. Such interpretations are useful not only for understanding the assumptions but also for the sensitivity analysis. For example, using the interpretations for Assumptions 1–3, from (2) and (3), a simple sensitivity analysis formula can be derived:

$$\mathrm{SACE} = \mathrm{E}(Y \mid A = 1, S = 1) - \mathrm{E}(Y \mid A = 0, S = 1) + \frac{\alpha}{p_0}\beta_0 - \frac{p_1 - p_0 + \alpha}{p_1}\beta_1,$$

where $\alpha = \Pr(S_1 = 0, S_0 = 1)$, $\beta_0 = \mathrm{E}(Y_0 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_0 \mid S_1 = 0, S_0 = 1)$ and $\beta_1 = \mathrm{E}(Y_1 \mid S_1 = S_0 = 1) - \mathrm{E}(Y_1 \mid S_1 = 1, S_0 = 0)$ are the sensitivity parameters [4]. Note that $\alpha = 0$ and $\beta_1 = 0$ correspond to Assumptions 1 and 2, respectively, and $(\beta_0, \beta_1) = (0, 0)$ corresponds to Assumption 3. However, if we try to implement a sensitivity analysis by setting the sensitivity parameters corresponding to Assumptions 1–3 themselves, rather than α, β_0, and β_1 above, it will be more difficult to give interpretations of the sensitivity parameters and to determine ranges of the parameters to examine. In addition, the estimation of the SACE becomes complex [12].

Analysis of the SACE is important when assessing quality-of-life outcomes in settings in which some of the participants may die. Similarly, it is important when assessing the cost of different treatment options when individuals may die before the full costs of different treatments are incurred. A number of applications concerning the SACE have also been pursued in the literature [3,9,10,12-14]. Unfortunately, the SACE may not be identified without making any assumptions, and the identifiability assumptions reviewed here may not hold in many actual studies. Therefore, an important subject for future research is to introduce more reasonable identifiability assumptions, which may hold in many actual studies. Then, even if the assumptions are relaxed, the sensitivity analysis will yield a result with a narrow range. Another subject for future research is to extend it to the setting of observational studies [15].

Acknowledgements

This work was supported partially by Grant-in-Aid for Scientific Research (No. 23700344) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. Zhang JL, Rubin DB (2003) Estimation of causal effects via principal stratification when some outcomes are truncated by "death." J Educ Behav Stat 28: 353–368.
2. Kurland BF, Heagerty PJ (2005) Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by death. Biostat 6: 241–258.
3. Hayden D, Pauler DK, Schoenfeld D (2005) An estimator for treatment comparisons amongst survivors in randomized trials. Biometrics 61: 305–310.
4. Chiba Y, VanderWeele TJ (2011) A simple method for principal strata effects when the outcome has been truncated due to death. Am J Epidemiol 173: 745–751.
5. Little RJ, Rubin DB (2000) Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Publ Health 21: 121–145.
6. Cole SR, Hernán MA (2008) Constructing inverse probability weights for marginal structural models. Am J Epidemiol 168: 656–664.
7. Pearl J (2009) Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge: Cambridge University Press.
8. Frangakis CE, Rubin DB (2002) Principal stratification in causal inference. Biometrics 58: 21–29.
9. Gilbert PB, Bosch RJ, Hudgens MG (2003) Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. Biometrics 59: 531–541.
10. Egleston BL, Scharfstein DO, Freeman EE, West SK (2007) Causal inference for non-mortality outcomes in the presence of death. Biostat 8: 526–545.
11. Chiba Y (2011) Marginal structural models for estimating principal stratum direct effects under the monotonicity assumption. Biometrical J 53: 1025–1034.
12. Jemiai Y, Rotnitzky A, Shepherd BE, Gilbert PB (2007) Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs. JR Stat Soc B 69: 879–901.
13. Mattei A, Mealli F (2007) Application of the principal stratification approach to the Faenza Randomized Experiment on breast self-examination. Biometrics 63: 437–446.
14. Shepherd BE, Redman MW, Ankerst DP (2008) Does finasteride affect the severity of prostate cancer? A causal severity analysis. J Am Stat Assoc 103: 1392–1404.
15. Egleston BL, Scharfstein DO, MacKenzie E (2009) On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death. Biometrics 65: 497–504.