Journal of Biometrics & Biostatistics
ISSN: 2155-6180
Chiba, J Biom Biostat 2012, 3:7
DOI: 10.4172/2155-6180.1000e116
Editorial Open Access

Estimation and Sensitivity Analysis of the Survivor Average Causal Effect under the Monotonicity Assumption

Yasutaka Chiba*

Division of Biostatistics, Clinical Research Center, Kinki University School of Medicine, Japan

In a previous editorial in this journal, we reviewed the identification assumptions of the Survivor Average Causal Effect (SACE) [1]. In this editorial, we discuss the estimation and sensitivity analysis of the SACE under the monotonicity assumption.

In randomized trials in which the outcome requires considerable follow-up, participants may die before the trial is complete. In such cases, the outcome is undefined. This situation is sometimes referred to as "truncated by death" or "censored by death" [2,3]. In these settings, a crude comparison between the survivors of each treatment arm may give misleading results because the randomization is no longer preserved by conditioning on a post-treatment event (survival), and thus the crude comparison is not a comparison of the same population comparing different treatments but is a comparison of different populations [4].

A treatment comparison that makes sense in this setting is the SACE, which is a comparison of the outcome between treated and untreated individuals in the subpopulation that would have survived in either arm [1,5]. Because this subpopulation is essentially an underlying characteristic of the individual, the SACE can circumvent the problem with the crude comparison. However, unfortunately, this subpopulation of interest is not identified. Evaluating the SACE from the observed data is a challenging subject.

We use identical notation to past literature [1]. 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 [6] corresponding to what would have happened if an individual had been in an arm other than the one they were in. Let Sa denote the survival status if the individual were in $A=a$, and let Y_a denote the outcome if the individual were 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 [7], i.e., we assume 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 such that $Y_a \amalg A$ and $S_a \amalg A$ [8], where $B \amalg C$ denotes that B is independent from C. These independencies also hold conditional on X denoting a set of baseline covariates that do not affect A but do affect both S and Y.

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

$$SACE \equiv E(Y_1 \mid S_1=S_0=1)-E(Y_0 \mid S_1=S_0=1).$$

To identify the SACE, we introduce the following two assumptions [1]:

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

Assumption 2: $Y_1 \amalg S_0 \mid \{S_1, X\}$.

Assumption 1 is sometimes referred to as the monotonicity assumption and implies that there is no individual with S_1=0 and S_0=1; i.e., $\Pr(S_1=0, S_0=1)=0$. Under this assumption,

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

and

$$\begin{aligned} &E(Y_1 \mid S_1 = S_0 = 1) \\ &= E(Y_1 \mid A = 0, S = 1) \\ &= \sum_x E(Y_1 \mid A = 0, S = 1, X = x) \Pr(X = x \mid A = 0, S = 1) \end{aligned} \quad (2)$$

As this formula includes a potential outcome, the SACE cannot be identified under Assumption 1 only. However, by using Assumption 2, the following identification formula of $E(Y_1 \mid S_1=S_0=1)$ is derived:

$$\begin{aligned} &E(Y_1 \mid S_1 = S_0 = 1) \\ &= \sum_x E(Y_1 \mid S_1 = S_0 = 1, X = x) \Pr(X = x \mid A = 0, S = 1) \\ &= \sum_x E(Y_1 \mid S_1 = 1, X = x) \Pr(X = x \mid A = 0, S = 1) \\ &= \sum_x E(Y \mid A = 1, S = 1, X = x) \Pr(X = x \mid A = 0, S = 1) \end{aligned} \quad (3)$$

Thus, we can identify the SACE under Assumptions 1 and 2. Note that, as seen from (2) and (3), we can use the following assumption to identify the SACE [9] instead of Assumption 2:

Assumption 2*: $Y_1 \amalg A \mid \{S, X\}$.

This alternative assumption holds if all factors affecting both S and Y are observed and they are conditioned. Therefore, under Assumption 1, the SACE can be identified if all baseline covariates are observed.

Under Assumptions 1 and 2 (or 2*), the SACE can very simply be estimated because the identification formulas (1) and (3) commonly include S=1 in the conditions of expectations and probability. When we limit the analysis set to individuals with S=1 (i.e., survivors),

***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 November 01, 2012; **Accepted** November 03, 2012; **Published** November 09, 2012

Citation: Chiba Y (2012) Estimation and Sensitivity Analysis of the Survivor Average Causal Effect under the Monotonicity Assumption. J Biom Biostat 3:e116. doi:10.4172/2155-6180.1000e116

we can apply the estimation methods developed in the context of observational studies to estimate casual effects of an exposure on the outcome by adjusting for measured confounders, where the target population is the unexposed group. These major methods (the model-based standardization approach [10], the inverse probability weighting [IPW] approach [11], and the doubly robust estimation [12]) have been summarized elsewhere [13].

Here, we present a marginal structural model [11] to yield an IPW estimate of the PSE. After limiting the analysis set to individuals with S=1, this analysis can be conducted using a weighted regression model of A on Y with the weights w_i=1 for individuals with A_i=0 and w_i=Pr(A_i=1 | X_i)/Pr(A_i=0 | X_i) for individuals with A_i=1, where i=1, …, N denotes an individual. The value of Pr(Ai=a | X_i) is often predicted using a regression model; e.g., a logistic regression model. We present an SAS code to implement this analysis in the appendix. For example, we consider the simple hypothetical data shown in table 1, where only one binary baseline covariate is observed. The data yielded a SACE estimate of 0.30 (95% confidence interval [CI]: 0.25, 0.35), whereas the crude estimate was (480/800)–(300/600)=0.10 (95% CI: 0.05, 0.15).

For the sensitivity analysis when Assumption 2 is violated, we can also apply the methods for unmeasured confounding developed in the context of observational studies [13]. We set a sensitivity parameter as [4,14]

$$\delta_X = \mathrm{E}(Y_1 \mid A=1, S=1, X) - \mathrm{E}(Y_1 \mid A=0, S=1, X). \quad (4)$$

This sensitivity parameter is the average difference in the outcome that would have been observed under a treatment comparing two different subpopulations, in the stratum with X: the first is the subpopulation that would have survived with treatment (A=1, S=1); the second is the subpopulation that would have survived without treatment (A=0, S=1).

By substituting δX into (2), we obtain the following sensitivity analysis formula:

$$\mathrm{SACE} = \hat{\Delta} - \sum_x \delta_X \Pr(X = x \mid A = 0, S = 1),$$

where is a SACE estimator, which is the difference between (1) and (3), under Assumptions 1 and 2. Specifically, under an assumption that the values of δ_X do not vary between the strata of X, this formula simplifies to

$$\mathrm{SACE} = \hat{\Delta} - \delta \quad (5)$$

where δ=E(Y_1 | A=1, S=1)–E(Y_1 | A=0, S=1) (=δ_X). This sensitivity analysis formula implies that the sensitivity analysis can be easily conducted. The sensitivity parameter δ is set by the investigator according to what is considered plausible. The parameter can be varied over a range of plausible values to examine how conclusions vary according to different parameter values. The confidence interval of the true SACE for a fixed value of δ can be obtained simply by subtracting δ from the upper and lower confidence limits of $\hat{\Delta}$. Therefore, we can

		A=1			A=0		
		X=1	X=0	Total	X=1	X=0	Total
S=1	Y=1	300	180	480	250	50	300
	Y=0	20	300	320	200	100	300
S=0		70	130	200	290	110	400
Total		390	610	1000	740	260	1000

Table 1: A hypothetical data.

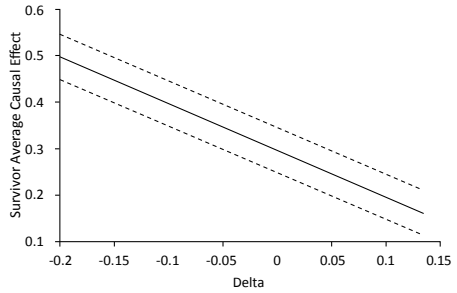

Figure 1: Sensitivity analysis of the survivor average causal effect; the solid line indicates the survivor average causal effect and broken lines indicate 95% confidence intervals.

readily display the results of the sensitivity analysis graphically, where the horizontal axis represents the sensitivity parameter, and the vertical axis represents the true SACE.

In some situations, it may be troublesome for investigators to determine the range of δ to examine. In such situations, the large sample bounds [2,15] can be used to determine this range. The bounds for E(Y_1 | A=0, S=1)=E(Y_1 | S_1=S_0=1) are derived from the number of individuals with S_1=S_0=1 in the subgroup with (A, S)=(1, 1). Because individuals in the subgroup with (A, S)=(1, 1) are limited to those with S_1=S0=1 and those with S_1=1 and S_0=0, under Assumption 1, this number is:

$$N_{11}\frac{\Pr(S_1 = S_0 = 1)}{\Pr(S_1 = S_0 = 1) + \Pr(S_1 = 1, S_0 = 0)}$$

$$= N_{11}\frac{\Pr(S_0 = 1)}{\Pr(S_1 = 1)}$$

$$= N_{11}\frac{\Pr(S = 1 \mid A = 0)}{\Pr(S = 1 \mid A = 1)},$$

where N_{11} is the number of individuals with (A, S)=(1, 1). For example, in the hypothetical data of table 1, this number is 800×(600/1000)/(800/1000)=600. Using this number, the large sample bounds for E(Y_1 | A=0, S=1) are calculated as:

$$\frac{0\times 320 + 1\times(600-320)}{600} \le \mathrm{E}(Y_1 \mid A=0, S=1) \le \frac{1\times 480 + 0\times(600-480)}{600}.$$

As E(Y_1 | A=1, S=1)=E(Y | A=1, S=1)=480/800, the range of δ becomes –0.20 ≤ δ ≤ 0.13. The result of the sensitivity analysis using (5) is shown in figure 1, where the lower and upper limits of the SACE are 0.16 (95% CI: 0.12, 0.21) and 0.50 (95% CI: 0.45, 0.55), respectively. When the larger value of the outcome denotes that the individual is healthier, if investigators are sure that individuals who survived even without treatment are likely to be healthier overall than those who would have survived with treatment, the second subpopulation in (4) is larger than the first subpopulation. In this situation, δ ≤ 0, and thus, the lower limit of the SACE is improved to 0.30 (95% CI: 0.25, 0.35).

In this editorial, we have presented simple methods for the estimation and sensitivity analysis of the SACE under the monotonicity assumption (Assumption 1). The methods are essentially identical to the methods developed in the context of observational studies to estimate the casual effects of an exposure on the outcome by adjusting for measured confounders. The methods presented here can be extended in a straightforward manner to other effect measures. Although the monotonicity assumption is a strong assumption,

whenever investigators consider that this assumption is plausible, the simple methods presented here can be applied.

Appendix

The following code is organized as follows. First, we used PROC LOGISTIC to fit the logistic model in the estimation of $\Pr(A=1 \mid X)$. Second, we used an SAS data step to calculate the weights for each individual from the estimated $\Pr(A=1 \mid X)$ of the previous logistic model. Third, we used PROC GENMOD to fit the weighted linear regression model that estimates the SACE. We note that the data file "IPW" contains only data of individuals with $S=1$ in the original data file.

```
/* Estimation of Predicted Values */
proc logistic data=IPW descending;
model A=X;
output out=PRED p=P;
run;
/* Calculation of the Weights */
data CIE;
set PRED;
if A=0 then WEIGHT=1; else WEIGHT=(1-P)/P;
run;
/* Weighted Analysis */
proc genmod data=CIE;
class i;
model Y=A/dist=normal link=identity;
weight WEIGHT;
repeated sub=i/type=ind;
run;
```

When the outcome is a binary variable, the SACE on the risk ratio scale can be estimated by replacing "dist=normal" and "link=identity" in the GENMOD procedure with "dist=poisson" and "link=log". Furthermore, when the "estimate 'beta' A 1/exp" is added in this procedure, the risk ratio and the 95% CI are displayed on the output.

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. Chiba Y, Taguri M, Uemura Y (2011) On the identification of the survivor average causal effect. J Biomet Biostat 2: e104.
2. 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.
3. Kurland BF, Heagerty PJ (2005) Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics 6: 241-258.
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. Hayden D, Pauler DK, Schoenfeld D (2005) An estimator for treatment comparisons among survivors in randomized trials. Biometrics 61: 305-310.
6. Little RJ, Rubin DB (2000) Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health 21: 121-145.
7. Cole SR, Hernán MA (2008) Constructing inverse probability weights for marginal structural models. Am J Epidemiol 168: 656-664.
8. Pearl J (2009) Causality: Models, Reasoning, and Inference. (2ndedn), Cambridge University Press.
9. Chiba Y (2011) Marginal structural models for estimating principal stratum direct effects under the monotonicity assumption. Biom J 53: 1025-1034.
10. Greenland S (1991) Estimating standardized parameters from generalized linear models. Stat Med 10: 1069-1074.
11. Robins JM, Hernán MA, Brumback BA (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11: 550-560.
12. Funk MJ, Westreich D, Weisen C, Stürmer T, Brookhart MA, et al. (2011) Doubly robust estimation of causal effects. Am J Epidemiol 173: 761-767.
13. Chiba Y (2012) A simple method for sensitivity analysis of unmeasured confounding. J Biomet Biostat 3: e113.
14. Chiba Y (2010) Bias analysis for the principal stratum direct effect in the presence of confounded intermediate variables. J Biomet Biostat 1: 101.
15. Chiba Y (2012) The large sample bounds on the principal strata effect with application to a prostate cancer prevention trial. Int J Biostat 8.