Preliminary Report

The Monotone Instrumental Variable in Randomized Trials with Noncompliance

Yasutaka Chiba

Department of Environmental Medicine and Behavioral Science,
Kinki University School of Medicine
e-mail: chibay@med.kindai.ac.jp

Noncompliance is an important problem in randomized trials. The estimation and bounds of average causal effects (ACEs) have been discussed as a way to address this issue. Previous studies have considered ACEs under the instrumental variable (IV) assumption, which postulates that potential outcomes are constant across subject sub-populations assigned to separate treatment regimens. However, the IV assumption may not be valid in unmasked trials. In the present analyses, the IV assumption is relaxed to the monotone IV (MIV) assumption, which replaces equality in the IV assumption with inequality. We propose bounds on ACEs under the MIV assumption in addition to the other existing assumptions. The results demonstrate that the intention-to-treat effect is an upper or lower bound under one assumption and the per-protocol effect is an upper or lower bound under the other assumption, even using the MIV assumption in place of the IV assumption. These proposed bounds are illustrated using a classic randomized trial.

Key words: Bounds; Causal inference; Intention-to-treat effect; Per-protocol effect; Potential outcome.

1. Introduction

In human clinical trials, ethical considerations pertaining to study subjects override a study's scientific requirements. One resource for coping with the inevitable trial-associated subject non-compliance is intention-to-treat (ITT) analysis. In ITT analysis, the parameter estimates are not affected by noncompliance, as subjects are analyzed according to the assigned treatment rather than the treatment actually received (Lee et al., 1991). However, although the ITT estimate may represent the effect of a treatment policy, it generally does not estimate the causal effect of a treatment in an unbiased manner (Sheiner and Rubin, 1995), where the causal effect is a comparison between the expected outcome if the subjects had received a test treatment versus that if they had received the control treatment.

As summarized in a recent review by Sato (2006), estimation of causal effects has been discussed by several researchers. These discussions have covered two types of causal effects:

the complier average causal effect (CACE) (Angrist et al., 1996), which is the causal effect among the subpopulation of potential compliers, and the average causal effect (ACE) (Robins and Greenland, 1996), which is the causal effect for the entire study population. If the study subjects are intended to represent the community, ACE, rather than CACE, would more closely correspond to a public health parameter of interest (Robins and Greenland, 1996). Thus, the analyses presented herein focus on ACE rather than CACE.

Although ACE estimators have been developed under specific assumptions (Robins, 1989; Chiba, 2010c), these assumptions are viewed as being overly strict and unpractical. Thus, ACEs will generally not be estimated without bias. Bounds and sensitivity analyses for ACE have been discussed previously (e.g., Cai et al., 2007; Chiba, 2010a) under the instrumental variable (IV) assumption, which states that potential outcomes are constant across sub-populations of subjects assigned to different study treatment arms. Although the IV assumption may hold in successfully masked randomized trials, it may often not hold in unmasked trials due to potential placebo effects. Therefore, this report discusses the estimation of ACEs when the IV assumption is relaxed.

Manski and Pepper (2000, 2009) have presented the monotone IV (MIV) assumption, which replaces equality in the IV assumption with inequality. Here, we propose bounds on the ACE under the MIV assumption. Section 2 reviews the potential outcomes framework, and Section 3 presents bounds on the ACE under the MIV assumption in addition to the other existing assumptions. The proposed bounds are illustrated using a classic randomized trial in Section 4. Section 5 presents a discussion of the findings.

2. Potential outcomes

Let $R(\omega)$ denote the treatment assigned to subject ω with $R(\omega) = 1$ if ω was assigned to the test treatment, and $R(\omega) = 0$ if ω was assigned to the control; $X(\omega)$ denote the treatment received by subject ω with $X(\omega) = 1$ if ω received the test treatment, and $X(\omega) = 0$ if ω received the control; and $Y(\omega)$ denote the outcome for subject ω .

We assume that, at the time of treatment assignment, the outcome that subject ω would have if the subject receives treatment x exists. This value is called the potential outcome of subject ω under treatment x (Rubin, 1974), and refers to what would have happened if the treatment had been something other than what it actually was. We express the potential outcome of subject ω as $Y_{X=x}(\omega)$, which is the outcome that ω would have if ω receives treatment x other than what it actually was. We require the stable unit treatment value assumption (SUTVA) and the consistency assumption. The SUTVA is to assume that $Y_{X=x}(\omega)$ is unrelated to the treatment status of other subjects (Rubin, 1974). The consistency assumption is to assume that the value of the potential outcome that would have been observed if ω had received treatment x is equal to the value of the observed outcome when ω actually received treatment x (Rubin,

1974). Therefore, this assumption implies that $E(Y_{X=x} | X = x) = E(Y | X = x)$ and further $E(Y_{X=x} | X = x, R = r) = E(Y | X = x, R = r)$ hold. In the framework of potential outcomes, the ACE is defined as $E(Y_{X=1}) - E(Y_{X=0})$. See Sato (2006) for details.

In previous studies incorporating ACE in randomized trials with noncompliance, the IV assumption has been required. This assumption states that the potential outcome $Y_{X=x}$ is not affected directly by the treatment assignment R ; rather, $Y_{X=x}$ is influenced only by the treatment that is actually received (Angrist et al., 1996). Thus, subjects' potential outcomes are independent of treatment assignment and are constant across the sub-populations of subjects assigned to different treatment arms. The IV assumption is formally presented as follows:

ASSUMPTION 1: Instrumental variable (IV). *The treatment assignment R is an IV, if $E(Y_{X=x} | R = 1) = E(Y_{X=x} | R = 0)$ for $x = 0, 1$.*

The IV assumption may hold in successfully masked randomized trials, because subjects are not aware of their assigned treatments in such trials and so the assigned treatments do not affect the potential outcomes. However, this assumption may often not hold in unmasked trials in which subjects are aware of the assigned treatment and this knowledge may affect the potential outcomes. We discuss cases in which the IV assumption is not required.

3. Assumptions and bounds

This section proposes bounds on ACE under specific assumptions. First, the IV assumption is relaxed to the MIV assumption as introduced in Section 3.1. The other assumptions are introduced in Sections 3.2 and 3.3, and the bounds are presented under combinations of these assumptions. Some of the results presented in Sections 3.1 and 3.2 represent substantially special cases of those presented by Manski and Pepper (2000, 2009), as the previous report discussed ordered categorical treatment variables with more than two categories, whereas the present analyses pertain to a binary treatment. However, the present analyses can still be regarded as an application of their results to randomized trials with noncompliance, as the previous report did not discuss such trials. Furthermore, the results presented in Section 3.3 are original. Although the ACE is discussed in terms of difference, the results can be readily applied to the causal risk ratio in the case of a binary outcome.

3.1 The monotone instrumental variable assumption

Manski and Pepper (2000, 2009) have previously presented the MIV assumption, which relaxes the IV assumption. In the setting of a binary treatment, this assumption is formally presented as follows:

ASSUMPTION 2.1: Monotone instrumental variable (MIV). *The treatment assignment R is a MIV, if $E(Y_{X=x} | R = 1) \geq E(Y_{X=x} | R = 0)$ for $x = 0, 1$.*

This assumption indeed replaces equality with inequality in the IV assumption, and means that the values of potential outcomes for subjects assigned to $R = 1$ are larger than those assigned to $R = 0$ on average. For example, consider an unmasked trial to compare a new treatment with a standard treatment, where the outcome is a measure such that a larger value is better for the subject's health. In such a trial, subjects may think that the new treatment is more effective than the standard treatment, and this thinking may give rise to better results for subjects assigned to the new treatment than those assigned to the standard treatment. Then, the values of potential outcomes for subjects assigned to a new treatment ($R = 1$) may be larger than those for subjects assigned to a standard treatment ($R = 0$) on average; this implies that $E(Y_{X=x} | R = 1) \geq E(Y_{X=x} | R = 0)$.

When the outcome Y has finite range $[K_0, K_1]$, the bounds on the ACE under the MIV are as follows:

$$\begin{aligned} & \{E(Y | X = 1, R = 0) \Pr(X = 1 | R = 0) + K_0 \Pr(X = 0 | R = 0)\} \\ & - \{E(Y | X = 0, R = 1) \Pr(X = 0 | R = 1) + K_1 \Pr(X = 1 | R = 1)\} \\ & \leq \text{ACE} \\ & \leq \{E(Y | X = 1, R = 1) \Pr(X = 1 | R = 1) + K_1 \Pr(X = 0 | R = 1)\} \\ & - \{E(Y | X = 0, R = 0) \Pr(X = 0 | R = 0) + K_0 \Pr(X = 1 | R = 0)\}. \end{aligned} \quad (1)$$

The derivation of this inequality is presented in Appendix A. Note that inequality (1) is a binary treatment version of the result from Manski and Pepper (2000, 2009), and $K_0 = 0$ and $K_1 = 1$ in the case of a binary outcome.

Using the inverse sign of the inequality in the MIV, the following inverse MIV (IMIV) assumption is applied:

ASSUMPTION 2.2: Inverse monotone instrumental variable (IMIV). *The treatment assignment R is an IMIV, if $E(Y_{X=x} | R = 1) \leq E(Y_{X=x} | R = 0)$ for $x = 0, 1$.*

In contrast to the MIV, the IMIV means that the values of potential outcomes for subjects assigned to $R = 1$ are smaller than those assigned to $R = 0$ on average. This may hold when the lower value of an outcome is better for subjects' health in the same comparison of treatments as the above example for the MIV.

The bounds on the ACE under the IMIV are

$$\begin{aligned} & \{E(Y | X = 1, R = 1) \Pr(X = 1 | R = 1) + K_0 \Pr(X = 0 | R = 1)\} \\ & - \{E(Y | X = 0, R = 0) \Pr(X = 0 | R = 0) + K_1 \Pr(X = 1 | R = 0)\} \\ & \leq \text{ACE} \\ & \leq \{E(Y | X = 1, R = 0) \Pr(X = 1 | R = 0) + K_1 \Pr(X = 0 | R = 0)\} \\ & - \{E(Y | X = 0, R = 1) \Pr(X = 0 | R = 1) + K_0 \Pr(X = 1 | R = 1)\}. \end{aligned}$$

The derivation of this inequality is similar to that presented for inequality (1).

3.2 The monotone treatment response assumption

Manski (1997) has previously presented the monotone treatment response (MTR) assumption. In the setting of a binary treatment, this assumption is formally presented as follows:

ASSUMPTION 3.1: Monotone treatment response (MTR). *For each ω , $Y_{X=1}(\omega) \geq Y_{X=0}(\omega)$.*

The MTR means that a subject ω takes a larger value of outcome if ω received the test treatment than if ω received the control treatment, and holds when it is apparent that an exposure (treatment) has a positive effect. An outstanding example of this assumption is the influence of smoking (X) on lung cancer (Y). It is well known that smoking is an important prognostic factor of lung cancer, and the potential occurrence of lung cancer if one smoked ($X = 1$) is higher than that if one had not smoked ($X = 0$) for all individuals; this implies that $Y_{X=1}(\omega) \geq Y_{X=0}(\omega)$.

By combining the MTR with the MIV, the lower bound on the ACE is improved as follows:

$$\text{ACE} \geq \max\{-\{E(Y | R = 1) - E(Y | R = 0)\}, 0\}, \quad (2)$$

where $E(Y | R = 1) - E(Y | R = 0)$ represents the ITT effect. The derivation of this inequality is presented in Appendix B. Note that inequality (2) is also a binary treatment version of the result from Manski and Pepper (2000, 2009).

As with the IMIV, the following inverse MTR (IMTR) assumption is applied:

ASSUMPTION 3.2: Inverse monotone treatment response (IMTR). *For each ω , $Y_{X=1}(\omega) \leq Y_{X=0}(\omega)$.*

The IMTR means that a subject ω takes a smaller value of outcome if ω received the test treatment than if ω received the control treatment, and holds when it is apparent that an exposure (treatment) has a negative effect. Again, consider the example of the influence of smoking (X) on lung cancer (Y). When subjects are smokers, the potential occurrence of lung cancer if one quits smoking ($X = 1$) is lower than that if one continues to smoke ($X = 0$) for all individuals; this implies that $Y_{X=1}(\omega) \leq Y_{X=0}(\omega)$.

Each combination of the MIV or IMIV and the MTR or IMTR can be used to improve the lower or upper bound on the ACE. The results are as follows:

Under the MIV + MTR, $\text{ACE} \geq \max\{-\text{ITT}, 0\}$ (inequality (2)),

Under the IMIV + MTR, $\text{ACE} \geq \max\{\text{ITT}, 0\}$,

Under the MIV + IMTR, $\text{ACE} \leq \min\{\text{ITT}, 0\}$,

Under the IMIV + IMTR, $\text{ACE} \leq \min\{-\text{ITT}, 0\}$,

where $\text{ITT} = E(Y | R = 1) - E(Y | R = 0)$. The derivation of these inequalities is similar to that presented for inequality (2).

3.3 The monotone treatment selection assumption

Manski and Pepper (2000, 2009) presented the monotone treatment selection (MTS) assumption as a special version of the MIV. In the setting of a binary treatment, this assumption is formally presented as follows:

ASSUMPTION 4.1: Monotone treatment selection (MTS). $E(Y_{X=x} \mid X = 1, R = r) \geq E(Y_{X=x} \mid X = 0, R = r)$ for $x = 0, 1$ and $r = 0, 1$.

The MTS means that subjects who received the test treatment tend to have larger values of outcome than those who received the control within each study treatment-arm subpopulation. For example, when patients with a worse condition prefer to receive the new treatment, it should be anticipated that the incidence proportion of a bad event (Y) (e.g., death) will be higher for patients who receive the new treatment ($X = 1$), compared with those who receive the standard treatment ($X = 0$); this implies that $E(Y_{X=x} \mid X = 1, R = r) \geq E(Y_{X=x} \mid X = 0, R = r)$.

By combining the MTS with the MIV, the upper bound on the ACE is improved as follows:

$$ACE \leq E(Y \mid X = 1, R = 1) - E(Y \mid X = 0, R = 0), \quad (3)$$

where the right hand side is the per-protocol (PP) effect. The derivation of this inequality is presented in Appendix C. Note that Manski and Pepper (2000, 2009) did not derive this bound because the MTS was considered to be a special case of the MIV, *i.e.*, R in the MIV was only replaced to X , and the combination of the MIV and MTS was not considered. Thus, inequality (3) is newly derived in the present report.

Similar to the above assumptions, the following inverse MTS (IMTS) assumption is applied:

ASSUMPTION 4.2: Inverse monotone treatment selection (IMTS). $E(Y_{X=x} \mid X = 1, R = r) \leq E(Y_{X=x} \mid X = 0, R = r)$ for $x = 0, 1$ and $r = 0, 1$.

The IMTS means that subjects who received the test treatment tend to take smaller values of outcome than those who received the control within each study treatment-arm subpopulation. For example, when health-conscious individuals prefer to receive the new treatment, it should be anticipated that the incidence proportion of a bad event (Y) will be lower for individuals who receive the new treatment ($X = 1$), compared with those who receive the standard treatment ($X = 0$); this implies that $E(Y_{X=x} \mid X = 1, R = r) \leq E(Y_{X=x} \mid X = 0, R = r)$.

We can also improve the lower or upper bound on the ACE under each combination of the MIV or IMIV and the MTS or IMTS, as follows:

Under the MIV + MTS, $ACE \leq PP$ (inequality (3)),

Under the IMIV + MTS, $ACE \leq NPP$,

Under the MIV + IMTS, $ACE \geq NPP$,

Under the IMIV + IMTS, $ACE \geq PP$,

where $PP = E(Y | X = 1, R = 1) - E(Y | X = 0, R = 0)$ and $NPP = E(Y | X = 1, R = 0) - E(Y | X = 0, R = 1)$. The derivation of these inequalities is similar to that presented for inequality (3).

The results presented in Sections 3.2 and 3.3 above imply that both the lower and upper bounds can be improved under some combinations of the MTR or IMTR and the MTS or IMTS in addition to the MIV or IMIV. These improved bounds are summarized in Table 1.

Table 1. Bounds on the ACE under some combinations of the MTR or IMTR and the MTS or IMTS in addition to the MIV or IMIV, where $ITT = E(Y | R = 1) - E(Y | R = 0)$, $PP = E(Y | X = 1, R = 1) - E(Y | X = 0, R = 0)$, and $NPP = E(Y | X = 1, R = 0) - E(Y | X = 0, R = 1)$.

Assumptions	Bounds on ACE
MIV + MTR + MTS	$\max\{-ITT, 0\} \leq ACE \leq PP$
IMIV + MTR + MTS	$\max\{ITT, 0\} \leq ACE \leq NPP$
MIV + IMTR + IMTS	$NPP \leq ACE \leq \min\{ITT, 0\}$
IMIV + IMTR + IMTS	$PP \leq ACE \leq \min\{-ITT, 0\}$

4. Application

For illustration, the bounds presented in Section 3 are applied to data from the Multiple Risk Factor Intervention Trial (MRFIT) (MRFIT Research Group, 1982). The MRFIT was a large field trial to test the effect of a multifactor intervention program on mortality from coronary heart disease (CHD) in middle-aged men with sufficiently high risk levels attributed to cigarette smoking, high serum cholesterol, and high blood pressure. The intervention consisted of dietary advice on ways to reduce blood cholesterol, smoking cessation counseling, and hypertension medication. All subjects were randomly assigned to the intervention program or the control group.

For this illustration, attention is restricted to the effects of cessation of cigarette smoking. This restriction follows other studies (Sato, 2000; Matsui, 2005) and was applied due to the paucity of differences achieved for the other risk factors. Table 2 summarizes the incidence of subject mortality due to CHD during the 7-year follow-up based on the assigned treatment and the actual subject smoking status 1 year after study entry. Analytical interest focused on the ACE, defined as the causal effect of quitting smoking on CHD deaths among all participants, rather than the CACE, which is the causal effect among the subpopulation of potential compliers. In the calculations, R represents the assigned group ($R = 1$ for the test group and $R = 0$ for the control group), X is the actual smoking status 1 year after entry ($X = 1$ for smoking cessation and $X = 0$ for continued smoking), and Y is the incidence of CHD deaths ($Y = 1$ for dead and $Y = 0$ for alive).

Table 2. The status of cigarette smoking and the incidence of mortality due to CHD in the MRFIT during a 7-year follow-up period.

Group	No. of participants	CHD deaths	Smoking status at 1 year	No. of participants	CHD deaths
Test	3833	69	Quit	991	11
			Not quit	2842	58
Control	3830	74	Quit	374	4
			Not quit	3456	70
Totals	7663	143			

To derive the ACE bounds, it is necessary to discuss whether the assumptions in Section 3 hold. In the MRFIT, subjects would have been aware of their assigned group, since this was an unmasked trial, so the intervention itself might have evoked a psychological response. Furthermore, in addition to smoking cessation counseling, the intervention consisted of dietary advice for the reduction of blood cholesterol and hypertension medication. These interventions might have directly influenced the incidence of CHD independent of smoking status. Therefore, on average, the potential incidence of CHD for subjects assigned to the test group might have been reduced as compared with subjects assigned to the control group. This observation implies that the validity of the IV assumption is questionable, but that the IMIV assumption, $E(Y_{X=x} | R = 1) \leq E(Y_{X=x} | R = 0)$, is reasonable. It is also obvious that cessation of cigarette smoking can prevent death from CHD. Thus, the IMTR assumption, $Y_{X=1}(\omega) \leq Y_{X=0}(\omega)$ for each ω , holds.

In general, health-conscious individuals may tend not to die from CHD and to quit smoking as compared with persons who are not oriented towards health considerations. Trial subjects would likely have had similar tendencies, and subjects who did quit smoking would logically tend not to have died from CHD. Therefore, it is considered that the IMTS assumption, $E(Y_{X=x} | X = 1, R = r) \leq E(Y_{X=x} | X = 0, R = r)$ for $x = 0, 1$ and $r = 0, 1$, is valid.

The arguments presented above demonstrate that the IMIV, IMTR, and IMTS can be assumed. Therefore, from Table 1, the bounds on the ACE become $-0.92\% \leq \text{ACE} \leq 0\%$, as $\text{ITT} = 69/3833 - 74/3830 = -0.13\%$ and $\text{PP} = 11/991 - 70/3456 = -0.92\%$.

5. Discussion

This report proposes bounds on ACE in randomized trials with noncompliance. Whereas previous studies have required the IV assumption, the analytical approach outlined above does not require this assumption. Instead, calculations utilize the MIV assumption, which replaces equality in the IV assumption with inequality. Although the results presented above are relevant to the causal differences, they can also be readily applied to the causal risk ratio when the outcome is binary.

It is generally thought that the ITT analysis is likely to yield a downwardly biased estimate

of causal effects (Sheiner and Rubin, 1995), whereas the PP analysis is likely to yield an upwardly biased estimate (Lewis and Machine, 1993). Thus, the ACE probably exists between the results of the ITT and PP analyses. While this is true under IV + MTR + MTS or under IV + IMTR + IMTS (Chiba, 2009), we cannot be certain that it is true when the IV assumption does not hold. As is evident from the results presented in Table 1, one bound is the ITT or PP effect, and another bound is not one of these effects. Therefore, investigators should not simply conclude that the ACE exists between the results of the ITT and PP analyses.

No standard method for estimating the ACE in randomized trials with noncompliance issues currently exists. Investigators should consider whether the assumptions presented in this report are valid and then yield the bounds on the ACE using the methodology described herein.

A recent interest in causal inference is statistical analysis concerning the role of an intermediate variable between a particular treatment and outcome. Investigators are often interested in understanding how the effect of a treatment on outcome may be mediated through an intermediate variable. For example, in the MRFIT, this implies that investigators are interested in how the effect of a multifactor intervention program on CHD mortality may be mediated through the smoking status 1 year after entry, rather than the effect of the smoking status 1 year after entry on CHD mortality.

Several ways to conceptualize the mediatory role of an intermediate variable in the treatment-outcome relationship have been proposed in the causal inference literature (Joffe et al., 2007). One such approach considers what would happen to the treatment-outcome relationship under interventions on the intermediate variable (Robins and Greenland, 1992; Pearl, 2001). Another approach focuses on the treatment-outcome relation for strata defined by potential outcomes for the treatment-mediator relationship (Frangakis and Rubin, 2002; Rubin, 2004). VanderWeele (2008) has clarified the relations between these two approaches.

Recently, several methods for statistical analyses have been developed under both approaches (e.g., Chiba, 2010bde). Such statistical analyses are closely related to issue of inference with a surrogate marker (Tanaka et al., 2010), where a good surrogate outcome serves as an intermediate variable of treatment effect, leaving little effect of the treatment to directly impact the true outcome of interest through other channels. They are also applied to issues of post-randomization selection bias and truncation-by-death (Zhang and Rubin, 2003; Hudgens and Halloran, 2006). Further methodological researches are needed for answering to these issues.

Acknowledgements

The author thanks Tosiya Sato of Kyoto University for his helpful communication.

REFERENCES

- Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444-455.
- Cai, Z., Kuroki, M., and Sato, T. (2007). Non-parametric bounds on treatment effects with non-compliance by covariate adjustment. *Statistics in Medicine* **26**, 3188-3204.
- Chiba, Y. (2009). Bounds on causal effects in randomized trials with noncompliance under monotonicity assumptions about covariates. *Statistics in Medicine* **28**, 3249-3259.
- Chiba, Y. (2010a). Bias analysis of the instrumental variable estimator as an estimator of the average causal effect. *Contemporary Clinical Trials* **31**, 12-17.
- Chiba Y. (2010b). Estimating the principal stratum direct effect when the total effects are consistent between two standard populations. *Statistics and Probability Letters* **80**, 958-961.
- Chiba, Y. (2010c). An approach for estimating causal effects in randomized trials with noncompliance. *Communications in Statistics – Theory and Methods* **39**, 2146-2156.
- Chiba, Y. (2010d). Bias analysis for the principal stratum direct effect in the presence of confounded intermediate variables. *Journal of Biometrics and Biostatistics* **1**, 101.
- Chiba, Y. (2010e). Bounds on controlled direct effects under monotonic assumptions about mediators and confounders. *Biometrical Journal* **52**, 628-637.
- Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21-29.
- Hudgens, M.G. and Halloran, M.E. (2006). Causal vaccine effects on binary postinfection outcomes. *Journal of the American Statistical Association* **101**, 51-64.
- Joffe, M., Small, D., and Hsu, C.-Y. (2007). Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statistical Science* **22**, 74-97.
- Lee, Y., Ellenberg, J., Hirtz, D., and Nelson, K. (1991). Analysis of clinical trials by treatment actually received: Is it really an option? *Statistics in Medicine* **10**, 1595-1605.
- Lewis, J.A. and Machine, D. (1993). Intention to treat – who should use ITT? *British Journal of Cancer* **68**, 647-650.
- Manski, C.F. (1997). Monotone treatment response. *Econometrica* **65**, 1311-1334.
- Manski, C.F. and Pepper, J.V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* **68**, 997-1010.
- Manski, C.F. and Pepper, J.V. (2009). More on monotone instrumental variables. *Econometrics Journal* **12**, S200-S216.
- Matsui, S. (2005). Stratified analysis in randomized trials with noncompliance. *Biometrics* **61**, 816-823.

- Multiple Risk Factor Intervention Trial Research Group (1982). Multiple risk factor intervention trial: Risk factor changes and mortality results. *Journal of the American Medical Association* **248**, 1465-1477.
- Pearl, J. (2001). Direct and indirect effects. In: *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco: Morgan Kaufman, 411-420.
- Robins, J.M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: *Health Service Research Methodology: A Focus on AIDS*, Sechrest, L., Freeman, H., and Mulley, A. (eds). Washington, DC: U.S. Public Health Service, 113-159.
- Robins, J.M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143-155.
- Robins, J.M. and Greenland, S. (1996). Discussion of "Identification of causal effects using instrumental variables" by Angrist, J.D., Imbens, G.W, and Rubin, D.B. *Journal of the American Statistical Association* **91**, 456-458.
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688-701.
- Rubin, D.B. (2004). Direct and indirect effects via potential outcomes. *Scandinavian Journal of Statistics* **31**, 161-170.
- Sato, T. (2000). Sample size calculations with compliance information. *Statistics in Medicine* **19**, 2689-2697.
- Sato, T. (2006). Randomization-based analysis of causal effects. In: *Handbook of Clinical Trials: Design and Analysis*, Tango, T. and Uesaka, H. (eds). Japan: Asakura Publishing, 535-556 (in Japanese).
- Sheiner, L. and Rubin, D.B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* **57**, 6-15.
- Tanaka, S., Oba, K., Yoshimura, K., and Teramukai, S. (2010). Statistical criteria for surrogate endpoint and application: A review. *Japanese Journal of Biometrics* **31**, 23-48 (in Japanese).
- VanderWeele, T.J. (2008). Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters* **78**, 2957-2962.
- Zhang, J.L. and Rubin, D.B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death." *Journal of Educational and Behavioral Statistics* **28**, 353-368.

Appendix

The appendix outlines the derivation of inequalities (1)–(3).

A. The derivation of inequality (1)

$E(Y_{X=x})$ is transformed as

$$E(Y_{X=x}) = E(Y_{X=x} \mid R=0) \Pr(R=0) + E(Y_{X=x} \mid R=1) \Pr(R=1).$$

Therefore, under the MIV assumption, $E(Y_{X=x} \mid R=1) \geq E(Y_{X=x} \mid R=0)$ for $x=0, 1$, the following inequality holds:

$$\begin{aligned} E(Y_{X=x}) &= E(Y_{X=x} \mid R=0) \Pr(R=0) + E(Y_{X=x} \mid R=1) \Pr(R=1) \\ &\geq E(Y_{X=x} \mid R=0) \Pr(R=0) + E(Y_{X=x} \mid R=0) \Pr(R=1) \\ &= E(Y_{X=x} \mid R=0), \end{aligned}$$

and similarly

$$\begin{aligned} E(Y_{X=x}) &= E(Y_{X=x} \mid R=0) \Pr(R=0) + E(Y_{X=x} \mid R=1) \Pr(R=1) \\ &\leq E(Y_{X=x} \mid R=1) \Pr(R=0) + E(Y_{X=x} \mid R=1) \Pr(R=1) \\ &= E(Y_{X=x} \mid R=1). \end{aligned}$$

$E(Y_{X=x} \mid R=r)$ is further transformed as

$$\begin{aligned} E(Y_{X=x} \mid R=r) &= E(Y_{X=x} \mid X=0, R=r) \Pr(X=0 \mid R=r) + E(Y_{X=x} \mid X=1, R=r) \Pr(X=1 \mid R=r). \end{aligned}$$

When $x=1$, the consistency assumption derives $E(Y_{X=1} \mid X=1, R=r) = E(Y \mid X=1, R=r)$, and finite range $[K_0, K_1]$ of Y derives $K_0 \leq E(Y_{X=1} \mid X=0, R=r) \leq K_1$. Using them, the lower bound of $E(Y_{X=1})$ becomes

$$\begin{aligned} E(Y_{X=1}) &\geq E(Y_{X=1} \mid R=0) \\ &= E(Y_{X=1} \mid X=0, R=0) \Pr(X=0 \mid R=0) \\ &\quad + E(Y_{X=1} \mid X=1, R=0) \Pr(X=1 \mid R=0) \\ &\geq K_0 \Pr(X=0 \mid R=0) + E(Y \mid X=1, R=0) \Pr(X=1 \mid R=0), \end{aligned} \tag{A.1}$$

and the upper bound of $E(Y_{X=1})$ becomes

$$\begin{aligned} E(Y_{X=1}) &\leq E(Y_{X=1} \mid R=1) \\ &= E(Y_{X=1} \mid X=0, R=1) \Pr(X=0 \mid R=1) \\ &\quad + E(Y_{X=1} \mid X=1, R=1) \Pr(X=1 \mid R=1) \\ &\leq K_1 \Pr(X=0 \mid R=1) + E(Y \mid X=1, R=1) \Pr(X=1 \mid R=1). \end{aligned} \tag{A.2}$$

The similar calculations for $E(Y_{X=0})$ yield the lower bounds of

$$\begin{aligned}
 E(Y_{X=0}) &\geq E(Y_{X=0} \mid R=0) \\
 &= E(Y_{X=0} \mid X=0, R=0) \Pr(X=0 \mid R=0) \\
 &\quad + E(Y_{X=0} \mid X=1, R=0) \Pr(X=1 \mid R=0) \\
 &\geq E(Y \mid X=0, R=0) \Pr(X=0 \mid R=0) + K_0 \Pr(X=1 \mid R=0),
 \end{aligned} \tag{A.3}$$

and the upper bound of

$$\begin{aligned}
 E(Y_{X=0}) &\leq E(Y_{X=0} \mid R=1) \\
 &= E(Y_{X=0} \mid X=0, R=1) \Pr(X=0 \mid R=1) \\
 &\quad + E(Y_{X=0} \mid X=1, R=1) \Pr(X=1 \mid R=1) \\
 &\leq E(Y \mid X=0, R=0) \Pr(X=0 \mid R=1) + K_1 \Pr(X=1 \mid R=1),
 \end{aligned} \tag{A.4}$$

because the consistency assumption derives $E(Y_{X=0} \mid X=0, R=r) = E(Y \mid X=0, R=r)$, and finite range $[K_0, K_1]$ of Y derives $K_0 \leq E(Y_{X=0} \mid X=1, R=r) \leq K_1$.

The lower bound on $ACE = E(Y_{X=1}) - E(Y_{X=0})$ is derived from {the lower bound of $E(Y_{X=1})$ } minus {the upper bound of $E(Y_{X=0})$ }, and the upper bound on ACE is derived from {the upper bound of $E(Y_{X=1})$ } minus {the lower bound of $E(Y_{X=0})$ }. These differences derive inequality (1).

B. The derivation of inequality (2)

The MTR assumption, $Y_{X=1}(\omega) \geq Y_{X=0}(\omega)$ for each ω , implies that

$$E(Y_{X=1}) \geq E(Y_{X=0}) \tag{B.1}$$

and further

$$E(Y_{X=1} \mid X=x, R=r) \geq E(Y_{X=0} \mid X=x, R=r) \tag{B.2}$$

hold. Substituting inequality (B.2) into inequality (A.1) yields

$$\begin{aligned}
 &E(Y_{X=1}) \\
 &\geq E(Y_{X=1} \mid X=0, R=0) \Pr(X=0 \mid R=0) + E(Y_{X=1} \mid X=1, R=0) \Pr(X=1 \mid R=0) \\
 &\geq E(Y_{X=0} \mid X=0, R=0) \Pr(X=0 \mid R=0) + E(Y_{X=1} \mid X=1, R=0) \Pr(X=1 \mid R=0) \\
 &= E(Y \mid X=0, R=0) \Pr(X=0 \mid R=0) + E(Y \mid X=1, R=0) \Pr(X=1 \mid R=0) \\
 &= E(Y \mid R=0)
 \end{aligned}$$

by the consistency assumption. Likewise, using inequality (B.2) and the consistency assumption, from inequality (A.4), the upper bound of $E(Y_{X=0})$ becomes

$$\begin{aligned}
 &E(Y_{X=0}) \\
 &\leq E(Y_{X=0} \mid X=0, R=1) \Pr(X=0 \mid R=1) + E(Y_{X=0} \mid X=1, R=1) \Pr(X=1 \mid R=1)
 \end{aligned}$$

$$\begin{aligned}
&\leq E(Y_{X=0} \mid X=0, R=1) \Pr(X=0 \mid R=1) + E(Y_{X=1} \mid X=1, R=1) \Pr(X=1 \mid R=1) \\
&= E(Y \mid X=0, R=1) \Pr(X=0 \mid R=1) + E(Y \mid X=1, R=1) \Pr(X=1 \mid R=1) \\
&= E(Y \mid R=1).
\end{aligned}$$

The bounds of $E(Y_{X=1})$ and $E(Y_{X=0})$ yield the lower bounds of $ACE \geq -\{E(Y \mid R=1) - E(Y \mid R=0)\}$. Inequality (B.1) yields the lower bound of $ACE = E(Y_{X=1}) - E(Y_{X=0}) \geq 0$. Therefore, inequality (2) is derived.

C. The derivation of inequality (3)

Substituting the MTS assumption, $E(Y_{X=x} \mid X=1, R=r) \geq E(Y_{X=x} \mid X=0, R=r)$ for $x=0, 1$ and $r=0, 1$, into inequality (A.2) yields

$$\begin{aligned}
&E(Y_{X=1}) \\
&\leq E(Y_{X=1} \mid X=0, R=1) \Pr(X=0 \mid R=1) + E(Y_{X=1} \mid X=1, R=1) \Pr(X=1 \mid R=1) \\
&\leq E(Y_{X=1} \mid X=1, R=1) \Pr(X=0 \mid R=1) + E(Y_{X=1} \mid X=1, R=1) \Pr(X=1 \mid R=1) \\
&= E(Y \mid X=1, R=1) \{\Pr(X=0 \mid R=1) + \Pr(X=1 \mid R=1)\} \\
&= E(Y \mid X=1, R=1),
\end{aligned}$$

by the consistency assumption. Likewise, using the MTS and consistency assumptions, from inequality (A.3), the lower bound of $E(Y_{X=0})$ becomes

$$\begin{aligned}
&E(Y_{X=0}) \\
&\geq E(Y_{X=0} \mid X=0, R=0) \Pr(X=0 \mid R=0) + E(Y_{X=0} \mid X=1, R=0) \Pr(X=1 \mid R=0) \\
&\geq E(Y_{X=0} \mid X=0, R=0) \Pr(X=0 \mid R=0) + E(Y_{X=0} \mid X=0, R=0) \Pr(X=1 \mid R=0) \\
&= E(Y \mid X=0, R=0) \{\Pr(X=0 \mid R=0) + \Pr(X=1 \mid R=0)\} \\
&= E(Y \mid X=0, R=0).
\end{aligned}$$

The bounds of $E(Y_{X=1})$ and $E(Y_{X=0})$ derive the upper bound of $ACE \leq E(Y \mid X=1, R=1) - E(Y \mid X=0, R=0)$, *i.e.*, inequality (3).