Zhichao Jiang, Yasutaka Chiba, and Tyler J. VanderWeele*

Monotone Confounding, Monotone Treatment Selection and Monotone Treatment Response

Abstract: Manski (Monotone treatment response. Econometrica 1997;65:1311–34) and Manski and Pepper (Monotone instrumental variables: with an application to the returns to schooling. Econometrica 2000;68:997–1010) gave sharp bounds on causal effects under the assumptions of monotone treatment response (MTR) and monotone treatment selection (MTS). VanderWeele (The sign of the bias of unmeasured confounding. Biometrics 2008;64:702–6) provided bounds for binary treatment under an assumption of monotone confounding (MC). We discuss the relation between MC and MTS and provide bounds under various combinations of these assumptions. We show that MC and MTS coincide for a binary treatment, but MC does not imply MTS for a treatment variable with more than two levels.

Keywords: bias, endogeneity, nonparametric bounds, partial identification

*Corresponding author: Tyler J. VanderWeele, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA, USA, E-mail: tvanderw@hsph.harvard.edu
Zhichao Jiang, Peking University, Beijing, China
Yasutaka Chiba, Division of Biostatistics, Clinical Research Center, Kinki University School of Medicine, Osaka, Japan, E-mail: chibay@med.kindai.ac.jp

1 Introduction

We consider bounds on causal effects for an ordinal treatment variable under treatment selection. The main purpose of this work is to consider the relationship between the assumptions of monotone treatment selection (MTS) [1] and monotone confounding (MC) [2]. We will show MC and MTS coincide for a binary treatment, but MC does not imply MTS for a treatment variable with more than two levels. We also derive new bounds under different combinations of MTS or MC and an assumption of monotone treatment response (MTR).

We assume a treatment response framework identical to Manski [3] and Manski and Pepper [1]. Let A_ω denote the treatment assigned to subject ω: $A_\omega = t$ if ω is assigned to the treatment t, where $t = 0, 1, ..., k$ is an ordered categorical variable. Y_ω denotes the outcome for subject ω, and $Y_\omega(t)$ denotes the potential outcome of Y for subject ω if ω is assigned to treatment t.

We use the notation $B \coprod C | D$ to denote that B is independent of C conditional on D [4]. The assumption of "unconfoundedness" or "weak ignorability" or "selection on observables" [5] for observed covariates X is that $Y(t) \coprod A | X$, where X denotes a covariate or a set of covariates. The assumption allows for identification of treatment effects by

$$E[Y(t)] - E[Y(0)]$$
$$= \sum_x E(Y|A = t, X = x)P(X = x) - \sum_x E(Y|A = 0, X = x)P(X = x).$$

However, the assumption itself is often not plausible. The literature on partial identification allows for more credible inference for treatment effects because it makes much weaker assumptions [1, 2, 6–8]. Some of this literature on partial identification derives bounds by imposing various monotonicity assumptions which are considerably weaker than "selection on observables." This paper advances this literature further by considering the relationships that hold between different monotonicity assumptions and by deriving new bounds under differing monotonicity assumptions that will sometimes be tighter than those currently offered in the literature.

The paper is organized as follows. In Section 2, we review the MTR, MTS, and MC assumptions and the bounds under them. In Section 3, we give results concerning relations between MTS and MC. In Section 4, we provide new results on bounds under MTR and a reversed conditional analog of MTS. Section 5 offers some concluding remarks.

2 Review of existing assumptions and bounds

Without monotonicity assumptions, the nonparametric bounds for an ordered categorical treatment are as follows [1]:

$$E(Y|A = t)P(A = t) + K_0 \sum_{s \neq t} P(A = s)$$
$$\leq E[Y(t)] \leq E(Y|A = t)P(A = t) + K_1 \sum_{s \neq t} P(A = s),$$

where K_0 and K_1 are the lower and upper bounds of Y, respectively.

The nonparametric bounds can be improved by adding certain monotonicity assumptions. In this section, we review assumptions introduced in the previous literature and the bounds under them. We review the MTR and MTS assumptions in Section 2.1. In Section 2.2, we review the MC assumption.

2.1 Monotone treatment response and monotone treatment selection

Manski [1, 6] introduced the following MTR assumption:

ASSUMPTION 1 (MTR assumption). *Let T be an ordered set, and t_1 and t_2 be elements of T. Then, $t_2 \geq t_1$ $\Rightarrow Y_\omega(t_2) \geq Y_\omega(t_1)$ for each ω.*

The MTR assumption is simply that for each individual an increase in the treatment variable A will increase or leave unchanged the outcome that would be observed.

PROPOSITION 1 (Manski [1, Corollary M1.2]). *Under the MTR assumption, bounds on E[Y(t)] are*

$$\sum_{s \leq t} E(Y|A = s)P(A = s) + K_0 P(A > t)$$
$$\leq E[Y(t)] \leq \sum_{s \geq t} E(Y|A = s)P(A = s) + K_1 P(A < t).$$

A characteristic of the MTR bounds is that the upper bound for treatment category of $A = 0$ is equal to the lower bound for that of $A = k$. Thus, the lower bound on the causal effect, $E[Y(k)] - E[Y(0)]$, under the MTR assumption is always zero.

Manski and Pepper [2] introduced the following MTS assumption:

ASSUMPTION 2 (MTS assumption). *Let T be an ordered set, and t_1 and t_2 be elements of T. Then, $t_2 \geq t_1 \Rightarrow E[Y(t) | A = t_2] \geq E[Y(t) | A = t_1]$.*

The interpretation of the MTS assumption is that if we compare two groups, those who actually received treatment level t_1 and those who actually received some higher level of treatment t_2, then, for each fixed treatment level t, the average outcomes that would have occurred if treatment had been fixed to t, is at least as high in the group that actually received t_2 as it is in the group that actually received t_1.

PROPOSITION 2 (Manski and Pepper [2, Proposition 1, Corollary 2]). *Under the MTS assumptions, bounds on* $E[Y(t)]$ *are*

$$E(Y|A = t)P(A \geq t) + K_0 P(A < t) \leq E[Y(t)] \leq E(Y|A = t)P(A \leq t) + K_1 P(A > t).$$

Bounds with narrower width follow by combining the MTR and MTS assumptions.

PROPOSITION 3 (Manski and Pepper [2, Proposition 2, Corollary 2]). *Under the MTR and MTS assumptions, bounds on* $E[Y(t)]$ *are*

$$\sum_{s \leq t} E(Y|A = s)P(A = s) + E(Y|A = t)P(A > t)$$
$$\leq E[Y(t)] \leq \sum_{s \geq t} E(Y|A = s)P(A = s) + E(Y|A = t)P(A < t).$$

Note that even when the MTS assumption is added to the MTR assumption, the upper bound for treatment category of $A = 0$ is still equal to the lower bound for that of $A = k$, and thus the lower bound on $E[Y(k)] - E[Y(0)]$ is always zero.

2.2 Monotone confounding

For observed covariates X, VanderWeele [3] introduced the following MC assumption:

ASSUMPTION 3 (MC assumption). *There exists univariate U such that* $Y(t) \amalg A | \{X, U\}$ *and both* $E(Y | A = t, X = x, U = u)$ *and* $E(A | X = x, U = u)$ *are either non-decreasing or non-increasing in u for all t and x.*

The MC assumption would be plausible if there were a single unmeasured covariate U that affected both A and Y such that controlling for the measured covariates X and the unmeasured covariate U sufficed to control for confounding and such that U affected treatment A and outcome Y in the same direction. The interpretation of MC in the presence of multiple unmeasured variables, however, is more complicated. Consider for example the relations in Figure 1 in which both U and V are unmeasured.
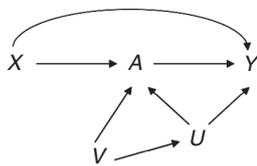

Figure 1 Example in which the effect of A on Y is unconfounded conditional on (X, U) but for which the MC assumption is difficult to interpret because of V

In this diagram, controlling for just U along with the measured covariates X would suffice to control for confounding for the effect of A on Y [9]. However, in Figure 1, the effect of U on A is itself confounded and the MC assumption that $E(A | X = x, U = u)$ is monotonic in u would not simply correspond to the effect of U on A being monotonic but would require that the association between U and A due to V was such that it was still the case that $E(A | X = x, U = u)$ was monotonic. The MC assumption may thus be difficult to establish or argue for on substantive grounds. It is, however, how epidemiologists often conceive of confounding [10] and for this reason it is of interest, from a theoretical perspective at least, to compare this conceptualization to that of MTS above. As we have seen, the interpretation of the MC assumption is only straightforward when there is a single unmeasured variable and appeal should probably only be made

to the assumption in such a context. With a single unmeasured variable, the monotonicity conditions can be interpreted simply as that U affects A and Y in the same direction. While it is impossible to know certainly that there is a single unmeasured confounding variable, epidemiologic studies often have very rich and extensive data on confounding variables which may make such an assumption more plausible in some epidemiologic studies than it would in many economic research contexts wherein data on important confounding variables is sometimes more limited.

For example, many clinical studies have data on nearly all of the information that is available to a physician who makes decisions concerning the treatments of patients. Not infrequently the only information that is not available in a clinical study is the physician's subjective assessment of the patient's prognosis. This subjective assessment could potentially serve as the unmeasured confounding variable in the MC assumption. For instance, Shahinian et al. [11] conducted a cohort study of prostate cancer patients to measure the effect of androgen deprivation therapy (ADT, the standard treatment for prostate cancer, which has fatigue, weakness, and frailty as side effects) on the occurrence of fractures. Control was made for a number of demographic and clinical variables including age, race, grade of prostate cancer, other cancer treatments received, and the occurrence of a fracture or the diagnosis of osteoporosis during the 12 months preceding the diagnosis of cancer. The study did not, however, measure the physicians' subjective assessment of the patients' maneuverability. Those with a high level of maneuverability will in general be less likely to be frail and experience a fracture and thus also more likely to receive ADT. The MC assumption may thus be plausible in this context.

As will be seen below, many of the results that follow also hold for multivariate U when the components of U are conditionally independent given X. For multivariate U, we say that $u \geq u'$ whenever each component of u is greater than or equal to the corresponding component of u'. The MC assumption for multivariate U is then once again that $E(Y \mid A = t, X = x, U = u)$ and $E(A \mid X = x, U = u)$ are either both non-decreasing or both non-increasing in u for all t and x. Many of the results below hold with multivariate U under the multivariate MC assumption but also require that the components of U are conditionally independent given X. Such conditional independence is a strong assumption and will be difficult to evaluate since U is unmeasured. Thus, once again, appeal to the MC assumption will in general only be reasonable when there is a single unmeasured covariate. Although we note below when results do hold with multivariate U under the conditional independence assumption, we do this principally for completeness rather than because of the utility of the results for multivariate U, which we believe, in most cases, will be limited. Alternatively, the results would also apply if there were a multivariate unmeasured confounder that could be summarized into a single score U for which the MC assumption held.

PROPOSITION 4 (VanderWeele [3, Theorem 1]). *For binary treatment A ($t_2 > t_1$), under the MC assumption, bounds on E[Y(t)] are*

$$E[Y(t_2)] \leq \sum_{x} E(Y|A = t_2, X = x)P(X = x),$$

$$E[Y(t_1)] \geq \sum_{x} E(Y|A = t_1, X = x)P(X = x).$$

Furthermore, if U is multivariate, the conclusion still holds if the components of U are conditionally independent given X.

The inequalities in Proposition 4 are reversed if, in Assumption 3, one of the conditional expectations is non-increasing in u for all t and x and the other is non-decreasing in u for all t and x.

VanderWeele [3] considered only bounds for a binary treatment. Chiba [12] showed that results similar to Proposition 4 also held for the average treatment effect on the treated employing conditional expectations of the form, $E[Y(t) \mid A = 1]$.

Proposition 4 above was proved using the following lemma that will also be used in developing subsequent results.

LEMMA 1 (Esary et al. [13, Theorem 2.1]). *Let f and g be functions with n real-valued arguments such that both f and g are non-decreasing in each of their arguments. If $X = (X_1,..., X_n)$ is a multivariate random variable with n components such that each component is independent of the other components, then* Cov$\{f(X), g(X)\} \geq 0$.

Construction of confidence intervals for the bounds of a causal effect, sometimes also called "uncertainty regions" [14], is a challenging problem and will not be explored further here as our focus here is principally conceptual, on the relationship between different monotonicity assumptions. The interested reader is referred to recent literature on the topic for further discussion [14–18].

3 Relation between monotone treatment selection and monotone confounding

In this section, we discuss the relation between the MTS and MC assumptions. The discussion is facilitated by considering a conditional version of the MTS assumption. The conditional version of MTS will be useful as it more closely corresponds to the MC assumption. The conditional MTS assumption and the unconditional MTS assumption coincide if the set of conditioning variables is empty. The conditional MTS assumption can be stated follows:

ASSUMPTION 4 (Conditional MTS assumption). *Let T be an ordered set, and t_1 and t_2 be elements of T. Then, $t_2 \geq t_1 \Rightarrow E[Y(t) \mid A = t_2, X = x] \geq E[Y(t) \mid A = t_1, X = x]$ for all x.*

The conditional MTS assumption implies MTS but the converse does not hold. By applying the MTS bounds under each distinct value of x and then summing over x we have the following immediate Corollary to Propositions 2 and 3 above.

COROLLARY 1. *Under conditional MTS, the bounds on $E[Y(t)]$ are*

$$\sum_{s \geq t} \sum_x E(Y|A = t, X = x)P(X = x|A = s)P(A = s) + K_0 P(A < t)$$

$$\leq E[Y(t)] \tag{1}$$

$$\leq \sum_{s \leq t} \sum_x E(Y|A = t, X = x)P(X = x|A = s)P(A = s) + K_1 P(A > t).$$

Under the conditional MTS and MTR assumptions, the bounds on $E[Y(t)]$ are

$$\sum_{s < t} E(Y|A = s)P(A = s) + \sum_{s \geq t} \sum_x E(Y|A = t, X = x)P(X = x|A = s)P(A = s)$$

$$\leq E[Y(t)] \tag{2}$$

$$\leq \sum_{s > t} E(Y|A = s)P(A = s) + \sum_{s \leq t} \sum_x E(Y|A = t, X = x)P(X = x|A = s)P(A = s).$$

Note that the upper bound for $A = 0$ is also equal to the lower bound for $A = k$ under the conditional MTS–MTR assumptions, and the lower bound on $E[Y(k)] - E[Y(0)]$ is always zero.

In the case of a binary treatment, the conditional MTS bounds are the same as the MC bounds on $E[Y(t)]$. This follows immediately from inequality (1). Likewise, the conditional MTS bounds on $E[Y(t) \mid A = a]$ ($a = 0, 1$) are also the same as the MC bounds. This is easily verified as follows:

$$E[Y(1)|A=0] = \sum_x E[Y(1)|A=0, X=x]P(X=x|A=0)$$
$$\leq \sum_x E[Y(1)|A=1, X=x]P(X=x|A=0)$$
$$= \sum_x E(Y|A=1, X=x)P(X=x|A=0)$$

by the conditional MTS assumption. Similarly,

$$E[Y(0)|A=1] \geq \sum_x E(Y|A=0, X=x)P(X=x|A=1)$$

is derived.

We now consider the relation between the MTS and MC assumptions themselves (rather than simply the bounds that are obtained under them). The following proposition states that MC and conditional MTS are equivalent for binary treatment. The result also holds under MC with multivariate U, provided the components of U are conditionally independent of one another given X.

PROPOSITION 5. *For binary treatment, the MC assumption implies the conditional MTS assumption. If in addition, the positivity assumption holds, i.e. $\min_{y_1, y_0} P[A=1|Y(1)=y_1, Y(0)=y_0, X=x] > 0$ for all x, then the conditional MTS assumption implies the MC assumption.*

PROOF. If the MC assumption holds, then for $t = 0, 1$,

$$E[Y(t)|A=1, X=x]$$
$$= \sum_u E[Y(t)|A=1, X=x, U=u]P(U=u|A=1, X=x)$$
$$= \sum_u E[Y(t)|A=t, X=x, U=u]P(U=u|A=1, X=x)$$
$$= \sum_u E[Y|A=t, X=x, U=u]\frac{P(A=1|X=x, U=u)P(U=u|X=x)}{P(A=1|X=x)}$$
$$= \frac{E_{F_{U|X=x}}\{E(Y|A=t, X=x, U=u)P(A=1|X=x, U=u)\}}{P(A=1|X=x)}$$
$$\geq \frac{E_{F_{U|X=x}}\{E(Y|A=t, X=x, U=u)\}E_{F_{U|X=x}}\{P(A=1|X=x, U=u)\}}{P(A=1|X=x)}$$
$$= E_{F_{U|X=x}}\{E(Y|A=t, X=x, U=u)\}$$
$$= \frac{E_{F_{U|X=x}}\{E(Y|A=t, X=x, U=u)\}E_{F_{U|X=x}}\{P(A=0|X=x, U=u)\}}{P(A=0|X=x)}$$
$$\geq \frac{E_{F_{U|X=x}}\{E(Y|A=t, X=x, U=u)P(A=0|X=x, U=u)\}}{P(A=0|X=x)}$$
$$= \frac{\sum_u E(Y|A=t, X=x, U=u)P(A=0|X=x, U=u)P(U=u|X=X)}{P(A=0|X=x)}$$
$$= \sum_u E(Y(t)|A=t, X=x, U=u)P(U=u|A=0, X=x)$$
$$= \sum_u E[Y(t)|A=0, X=x, U=u]P(U=u|A=1, X=x)$$
$$= E[Y(t)|A=0, X=x]$$

The first inequality follows because by Lemma 1,

$$E_{\Phi_{Y|\Xi=\xi}}\{\varphi(\xi,Y)\gamma(\xi,Y)\} - E_{\Phi_{Y|\Xi=\xi}}\{\varphi(\xi,Y)\}E_{\Phi_{Y|\Xi=\xi}}\{\gamma(\xi,Y)\}$$

$$= \text{Cov}_{\Phi_{Y|\Xi=\xi}}\{\varphi(\xi,Y),\gamma(\xi,Y)\}$$

$$\geq 0,$$

where both $f(x, U) = E(Y \mid A = t, X = x, U = u)$ and $g(x, U) = P(A = 1 \mid X = x, U = u)$ are non-decreasing in u. Likewise, the second inequality follows because $E(Y \mid A = t, X = x, U = u)$ is non-decreasing in u and $P(A = 0 \mid X = x, U = u)$ is non-increasing in u. A similar calculation holds in the case in which both $f(x, U)$ and $g(x, U)$ are non-increasing in u. We thus see that if the MC assumption holds then the conditional MTS assumption also holds.

On the other hand, if the conditional MTS assumption holds, then we have that the joint distribution of $\{Y(1), Y(0), A, X\}$ are such that $E[Y(t) \mid A = 1, X = x] \geq E[Y(t) \mid A = 0, X = x]$. Let U denote a binary variable such that $P[U = 1|Y(1) = y_1, Y(0) = y_0, A = t, X = x] = \theta_{y_1y_01x}$ and $P[U = 1|Y(1) = y_1, Y(0) = y_0, A = 0, X = x] = \theta_{y_1y_00x}$ with

$$\theta_{y_1y_01x} = \frac{(\frac{1}{e_{0x}} - 1) - R_{y_1y_0x}}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}}, \quad \theta_{y_1y_00x} = \frac{(\frac{1}{e_{1x}} - 1)}{R_{y_1y_0x}}\theta_{y_1y_01x},$$

where $R_{y_1y_0x} = \frac{P[A=0|Y(1)=y_1,Y(0)=y_0,X=x]}{P[A=1|Y(1)=y_1,Y(0)=y_0,X=x]}$ and e_{0x} and e_{1x} are constants satisfying $0 < e_{0x} < e_{1x} \leq \min_{y_1,y_0}\{P[A = 1|Y(1) = y_1, Y(0) = y_0, X = x]\}$. We show that this U satisfies the MC assumption.

First, we have

$$P[A = 1|Y(1) = y_1, Y(0) = y_0, U = 1, X = x]$$

$$= \frac{P[U = 1|Y(1) = y_1, Y(0) = y_0, A = 1, X = x]P[A = 1|Y(1) = y_1, Y(0) = y_0, X = x]}{\sum_{t=0,1} P[U = 1|Y(1) = y_1, Y(0) = y_0, A = t, X = x]P[A = t|Y(1) = y_1, Y(0) = y_0, X = x]}$$

$$= \frac{\theta_{y_1y_01x}P[A = 1|Y(1) = y_1, Y(0) = y_0, X = x]}{\sum_{t=0,1} \theta_{y_1y_0tx}P[A = t|Y(1) = y_1, Y(0) = y_0, X = x]}$$

$$= \frac{1}{1 + \frac{\theta_{y_1y_00x}}{\theta_{y_1y_01x}}R_{y_1y_0x}} = \frac{1}{1 + \left(\frac{1}{e_{1x}} - 1\right)} = e_{1x},$$

$$P[A = 1|Y(1) = y_1, Y(0) = y_0, U = 0, X = x]$$

$$= \frac{P[U = 0|Y(1) = y_1, Y(0) = y_0, A = 1, X = x]P[A = 1|Y(1) = y_1, Y(0) = y_0, X = x]}{\sum_{t=0,1} P[U = 0|Y(1) = y_1, Y(0) = y_0, A = t, X = x]P[A = t|Y(1) = y_1, Y(0) = y_0, X = x]}$$

$$= \frac{(1 - \theta_{y_1y_01x})P[A = 1|Y(1) = y_1, Y(0) = y_0, X = x]}{\sum_{t=0,1} (1 - \theta_{y_1y_0tx})P[A = t|Y(1) = y_1, Y(0) = y_0, X = x]}$$

$$= \frac{1}{1 + \frac{1-\theta_{y_1y_00x}}{1-\theta_{y_1y_01x}}R_{y_1y_0x}} = \frac{1}{1 + \frac{R_{y_1y_0x} - \left(\frac{1}{e_{1x}}-1\right)\theta_{y_1y_01x}}{1-\theta_{y_1y_01x}}}$$

$$= \frac{1}{1 + \left(\frac{1}{e_{0x}} - 1\right)} = e_{0x}.$$

Since e_{1x} and e_{0x} only depend on x, we have

$$P[A = 1|U = 1, X = x] = P[A = 1|Y(1) = y_1, Y(0) = y_0, U = 1, X = x] = e_{1x},$$
$$P[A = 1|U = 0, X = x] = P[A = 1|Y(1) = y_1, Y(0) = y_0, U = 0, X = x] = e_{0x},$$

which means $Y(t) \amalg A \mid \{X, U\}$.

Second, by construction we have $e_{0x} \leq e_{1x}$, thus we $E(A = 1|X = x, U = u)$ is non-decreasing in u. Third, we can get

$$P[U = 1|Y(1) = y_1, A = 1, X = x]$$

$$= \sum_{y_0} \theta_{y_1 y_0 1 x} P[Y(0) = y_0|Y(1) = y_1, A = 1, X = x]$$

$$= \sum_{y_0} \frac{\left(\frac{1}{e_{0x}} - 1\right) - R_{y_1 y_0 x}}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} P[Y(0) = y_0|Y(1) = y_1, A = 1, X = x]$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \sum_{y_0} R_{y_1 y_0 x} P[Y(0) = y_0|Y(1) = y_1, A = 1, X = x] \right\}$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \sum_{y_0} \frac{P[A = 0|Y(1) = y_1, Y(0) = y_0, X = x]}{P[A = 1|Y(1) = y_1, Y(0) = y_0, X = x]} P[Y(0) = y_0|Y(1) = y_1, A = 1, X = x] \right\}$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \sum_{y_0} \frac{P[A = 0|Y(1) = y_1, Y(0) = y_0, X = x] P[Y(0) = y_0|Y(1) = y_1, X = x]}{P[A = 1|Y(1) = y_1, X = x]} \right\}$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \sum_{y_0} \frac{P[A = 0, Y(0) = y_0|Y(1) = y_1, X = x]}{P[A = 1|Y(1) = y_1, X = x]} \right\}$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \frac{P[A = 0|Y(1) = y_1, X = x]}{P[A = 1|Y(1) = y_1, X = x]} \right\}$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \frac{P[Y(1) = y_1|A = 0, X = x] P(A = 0|X = x)}{P[Y(1) = y_1|A = 1, X = x] P(A = 0|X = x)} \right\}.$$

Therefore, we have

$$E[Y|A = 1, X = x, U = 1]$$

$$= \sum_{y_1} y_1 P[Y(1) = y_1|A = 1, X = x, U = 1]$$

$$= \sum_{y_1} y_1 \frac{P[U = 1|Y(1) = y_1, A = 1, X = x] P[Y(1) = y_1|A = 1, X = x]}{P(U = 1|A = 1, X = x)}$$

$$= \frac{1}{P(U = 1|A = 1, X = x) \frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \sum_{y_1} y_1 \left\{ \frac{1}{e_{0x}} - 1 - \frac{P[Y(1) = y_1|A = 0, X = x] P(A = 0|X = x)}{P[Y(1) = y_1|A = 1, X = x] P(A = 1|X = x)} \right\} P[Y(1) = y_1|A = 1, X = x]$$

$$= \frac{E[Y|A = 1, X = x]}{P(U = 1|A = 1, X = x) \frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left\{ \frac{1}{e_{0x}} - 1 - \frac{E[Y(1)|A = 0, X = x] P(A = 0|X = x)}{E[Y(1)|A = 1, X = x] P(A = 1|X = x)} \right\},$$

where

$$P(U = 1|A = A, X = x)$$

$$= \sum_{y_1} P[U = 1|Y(1) = y_1, A = 1, X = x] P[Y(1) = y_1|A = 1, X = x]$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \sum_{y_1} \left\{ \frac{1}{e_{0x}} - 1 - \frac{P[Y(1) = y_1|A = 0, X = x] P(A = 0|X = x)}{P[Y(1) = y_1|A = 1, X = x] P(A = 1|X = x)} \right\} P[Y(1) = y_1|A = 1, X = x]$$

$$= \frac{1}{\frac{1}{e_{0x}} - \frac{1}{e_{1x}}} \left[\frac{1}{e_{0x}} - 1 - \frac{P(A = 0|X = x)}{P(A = 1|X = x)} \right].$$

Then, we have

$$E[Y|A = 1, U = 1, X = \mathrm{x}]/E[Y|A = 1, X = \mathrm{x}]$$

$$= \left\{ \frac{1}{e_{0x}} - 1 - \frac{E[Y(1)|A = 0, X = x]P(A = 0|X = \mathrm{x})}{E[Y(1)|A = 1, X = x]P(A = 1|X = \mathrm{x})} \right\} \Big/ \left[\frac{1}{e_{0x}} - 1 - \frac{P(A = 0|X = \mathrm{x})}{P(A = 1|X = \mathrm{x})} \right].$$

Since $E[Y(t) \mid A = 1, X = x] \geq E[Y(t) \mid A = 0, X = x]$ for $t = 0,1$, we have $E[Y| A = 1, X = x, U = 1] \geq E[Y| A = 1, X = x]$, and then $E[Y|A = 1, X = x, U = 1] \geq E[Y|A = 1, X = x, U = 0]$. Similarly, we can get $E[Y| A = 0, X = x, U = 1] \geq E[Y|A = 0, X = x, U = 0]$. Therefore, $E[Y|A = t, X = x, U = u]$ is non-decreasing in u. $\square$

Although MC and conditional MTS are equivalent for binary treatment, the MC assumption does not imply the conditional MTS assumption for a treatment variable with three or more levels, as stated in the next proposition.

PROPOSITION 6. *For treatment with three or more level, the MC assumption does not imply the conditional MTS assumption.*

PROOF. For simplicity, we assume X is empty.

First, suppose $P(U = 1) = P(U = 0) = 0.5$, and A can take a value in $\{0,1,2\}$ with $P(A = 0|U = 1) = 0.3$, $P(A = 0|U = 0) = 0.5$, $P(A = 1|U = 1) = 0.5$, $P(A = 1|U = 0) = 0.2$, $P(A = 2|U = 1) = 0.2$ and $P(A = 2|U = 0) = 0.3$, then we have $E[A|U = 1] = 0.9$ and $E[A|U = 0] = 0.8$. Thus, $E[A|U = u]$ is increasing in u. Suppose $E[Y|A = t, U = 1] > E[Y|A = t, U = 0]$ for $t = 0,1,2$. Therefore, the MC assumption holds. Then we can get

$$E[Y(1)|A = 1] = \sum_{u=0,1} E[Y|A = 1, U = u]P(U = u|A = 1)$$
$$= \frac{5}{7}E[Y|A = 1, U = 1] + \frac{2}{7}E[Y|A = 1, U = 0],$$

$$E[Y(1)|A = 2] = \sum_{u=0,1} E[Y|A = 1, U = u]P(U = u|A = 2)$$
$$= 0.4E[Y|A = 1, U = 1] + 0.6E[Y|A = 1, U = 0].$$

Thus, $E[Y(1)|A = 1] > E[Y(1)|A = 2]$, which means that the MTS assumption does not hold.

However, the problem whether the conditional MTS assumption implies the MC assumption for a treatment variable with three or more levels remains open.

4 New bounds under reverse conditional monotone treatment selection and monotone treatment response

We finally give one more proposition that gives bounds when the MTR and conditional MTS assumptions hold but in opposite directions from each other. We thus further introduce the following reverse conditional MTS assumption:

ASSUMPTION 5 (Reverse conditional MTS assumption). *Let T be an ordered set, and t_1 and t_2 be elements of T. Then, $t_2 \geq t_1 \Rightarrow E[Y(t) \mid A = t_2, X = x] \leq E[Y(t) \mid A = t_1, X = x]$ for all x.*

New bounds can be derived under a combination of the MTR and reverse conditional MTS assumptions.

PROPOSITION 7. *Suppose that the MTR assumption holds and that the reverse conditional MTS assumption holds, then*

$$K_0 P(A > t) + \sum_x \sum_{s \leq t} \max \left\{ \begin{array}{l} E(Y|A = t, X = x) \\ E(Y|A = s, X = x) \end{array} \right\} P(A = s|X = x) P(X = x)$$

$$\leq E[Y(t)]$$

$$\leq K_1 P(A < t) + \sum_x \sum_{s \geq t} \min \left\{ \begin{array}{l} E(Y|A = t, X = x) \\ E(Y|A = s, X = x) \end{array} \right\} P(A = s|X = x) P(X = x).$$

PROOF.

$$E[Y(t)|X = x]$$

$$= \sum_{s \geq t} E[Y(t)|A = s, X = x] P(A = s|X = x)$$

$$+ \sum_{s < t} E[Y(t)|A = s, X = x] P(A = s|X = x)$$

$$\leq \sum_{s \geq t} E[Y(t)|A = s, X = x] P(A = s|X = x) + K_1 \sum_{s < t} P(A = s|X = x).$$

Now, for $s \geq t$, we have that $E[Y(t) | A = s, X = x] \leq E[Y(s) | A = s, X = x]$ by the MTR assumption and $E[Y(t) | A = s, X = x] \leq E[Y(t) | A = t, X = x]$ by the reverse conditional MTS assumption. Thus, we have that

$$E[Y(t)|A = s, X = x]$$

$$\leq \min \left\{ \begin{array}{l} E[Y(t)|A = t, X = x] \\ E[Y(s)|A = s, X = x] \end{array} \right\} = \min \left\{ \begin{array}{l} E(Y|A = t, X = x) \\ E(Y|A = s, X = x) \end{array} \right\}.$$

Consequently,

$$E[Y(t)|X = x]$$

$$\leq \sum_{s \geq t} \min \left\{ \begin{array}{l} E(Y|A = t, X = x) \\ E(Y|A = s, X = x) \end{array} \right\} P(A = s|X = x) + K_1 P(A < t|X = x).$$

Multiplying by $P(X = x)$ and summing over x establishes the right-hand side of the inequality of Proposition 7. The left-hand side of the inequality of Proposition 7 is proved similarly. This completes the proof. □

Note that if A is binary then we have the following bounds for $E[Y(0)]$ and $E[Y(1)]$:

$$K_0 P(A = 1) + \sum_x E(Y|A = 0, X = x) P(A = 0|X = x) P(X = x)$$

$$\leq E[Y(0)]$$

$$\leq \sum_x E(Y|A = 0, X = x) P(A = 0|X = x) P(X = x)$$

$$+ \sum_x \min \left\{ \begin{array}{l} E(Y|A = 0, X = x) \\ E(Y|A = 1, X = x) \end{array} \right\} P(A = 1|X = x) P(X = x),$$

and

$$\sum_x E(Y|A = 1, X = x) P(A = 1|X = x) P(X = x)$$

$$+ \sum_x \max \left\{ \begin{array}{l} E(Y|A = 0, X = x) \\ E(Y|A = 1, X = x) \end{array} \right\} P(A = 0|X = x) P(X = x)$$

$$\leq E[Y(1)]$$

$$\leq K_1 P(A = 0) + \sum_x E(Y|A = 1, X = x) P(A = 1|X = x) P(X = x).$$

In particular, a lower bound on the causal effect is given by

$$E[Y(1)] - E[Y(0)]$$

$$\geq \sum_x E(Y|A = 1, X = x)P(A = 1|X = x)P(X = x)$$

$$+ \sum_x \max\left\{ \begin{array}{l} E(Y|A = 0, X = x) \\ E(Y|A = 1, X = x) \end{array} \right\} P(A = 0|X = x)P(X = x)$$

$$- \left[\sum_x E(Y|A = 0, X = x)P(A = 0|X = x)P(X = x) \right.$$

$$\left. + \sum_x \min\left\{ \begin{array}{l} E(Y|A = 0, X = x) \\ E(Y|A = 1, X = x) \end{array} \right\} P(A = 0|X = x)P(X = x) \right].$$

This lower bound for the causal effect will often be greater than (and never less than) the lower bound given in VanderWeele [3] under the MC (or conditional MTS) assumption, which was

$$E[Y(1)] - E[Y(0)]$$

$$\geq \sum_x E(Y|A = 1, X = x)P(X = x) - \sum_x E(Y|A = 0, X = x)P(X = x)$$

$$= \sum_x E(Y|A = 1, X = x)P(A = 1, X = x)P(X = x)$$

$$+ \sum_x E(Y|A = 1, X = x)P(A = 0, X = x)P(X = x)$$

$$- \left\{ \sum_x E(Y|A = 0, X = x)P(A = 0, X = x)P(X = x) \right.$$

$$\left. + \sum_x E(Y|A = 0, X = x)P(A = 1, X = x)P(X = x) \right\}.$$

5 Conclusion

When thinking about the direction and extent of bias due to unmeasured confounding, different formal assumptions have been employed. Economists have more frequently appealed to a MTS assumption. Epidemiologists have more frequently appealed to a MC assumption. In this paper, we have shown that the MC assumption and the conditional MTS assumptions are in fact equivalent for binary treatment variables, but the MC assumption does not imply the conditional MTS assumption for treatment variables with three or more levels. We have also presented new bounds under the MTR assumption along with a reversed analog of the conditional MTS assumption. These new bounds will generally be narrower than bounds currently in the literature. The MC assumption is often more difficult to justify on substantive grounds than the conditional MTS assumption. This is because the MC assumption will generally only be reasonable if there is a single unmeasured confounder and it is often difficult to know this in practice. However, again, in the case of a binary treatment, the two assumptions are in fact equivalent and so, with a binary treatment, epidemiologists and economists can reasonably conceptualize unmeasured confounding in either manner.

References

1. Manski CF, Pepper JV. Monotone instrumental variables: with an application to the returns to schooling. Econometrica 2000;68:997–1010.
2. VanderWeele TJ. The sign of the bias of unmeasured confounding. Biometrics 2008;64:702–6.
3. Manski CF. Monotone treatment response. Econometrica 1997;65:1311–34.
4. Dawid AP. Conditional independence in statistical theory (with discussion). J R Stat Soc Ser B 1979;41:1–31.
5. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
6. Manski CF. Nonparametric bounds on treatment effects. Am Econ Rev 1990;80:319–23.
7. Manski CF. Partial identification of probability distributions. New York: Springer, 2003.
8. Manski CF, Pepper JV. More on monotone instrumental variables. Econometrics J 2009;12:S200–16.
9. Pearl J. Causality: models, reasoning, and inference, 2nd ed. Cambridge: Cambridge University Press, 2009.
10. Vander Stoep A, Beresford SA, Weiss NS. A didactic device for teaching epidemiology students how to anticipate the effect of a third factor on an exposure-outcome relation. Am J Epidemiol 1999;150:221.
11. Shahinian VB, Kuo YF, Freeman JL, Goodwin JS. Risk of fracture after androgen deprivation for prostate cancer. New Engl J Med 2005;352:154–64.
12. Chiba Y. The sign of the unmeasured confounding bias under various standard populations. Biom J 2009;51:670–6.
13. Esary JD, Proschan F, Walkup DW. Association of random variables, with applications. Ann Math Stat 1967;38:1466–74.
14. Vansteelandt G, Goetghebuer E, Kenward MG, Molenberghs G. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. Stat Sin 2006;16:953–79.
15. Bugni FA. Bootstrap inference in partially identified models defined by moment inequalities: coverage of the identified set. Econometrica 2010;78:735–53.
16. Imbens GW, Manski CF. Confidence intervals for partially identified parameters. Econometrica 2004;72:1845–57.
17. Romano JP, Shaikh AM. Inference for identifiable parameters in partially identified econometric models. J Stat Plann Inference 2008;138:2786–807.
18. Todem D, Fine J, Peng L. A global sensitivity test for evaluating statistical hypotheses with non-identifiable models. Biometrics 2010;66:558–66.