Cross Direct Effects in Settings with Two Mediators

Erin E. Gabriel[1] Arvid Sjölander[1] Dean Follmann[2]
Michael C. Sachs[1]

1:Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

2:Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA.

corresponding author: erin.gabriel@ki.se

May 5, 2022

Abstract

When multiple mediators are present, there are additional effects that may be of interest beyond the well-known natural and controlled direct effects. These effects cross the type of control on the mediators holding one to a constant level and one to its natural level, which differs across subjects, under some intervention. We introduce five such estimands for the cross controlled and natural direct effects when two mediators are measured and the intervention of interest is randomized. Such estimands may be of interest in immunology, as we discuss in relation to measured immunological responses to SARS-CoV-2 vaccination. We provide identifying expressions for the estimands in settings where there is no unmeasured confounding and the intervention, outcome and mediators are of arbitrary type. We further provide tight symbolic bounds for the estimands in settings where there may be unmeasured confounding and the intervention, mediators and outcome are binary. *Keywords:* Causal pathways; Multiple mediation; Symbolic bounds

1 Introduction

In causal research in medicine, biology, economics or agriculture it is often of interest to estimate the effect of an intervention on an outcome removing the effect of a possible mediator. There are several such effects that have been discussed in the literature. For example, there is the controlled direct effect (CDE), where the mediator is held to a constant level for all subjects, and the natural direct effect (NDE), where the mediator is held to the

'natural' level it would attain, for each subject, had the intervention been set to a particular level. Arguably, the NDE is less intuitive than the CDE. An example which may help to clarify the NDE concerns the direct effect of smoking on lung cancer, where one wishes to hold all other co-morbidities to the level they would have been for a person given that they had not smoked. This is clearly of interest, and may in fact be of greater interest than the direct effect of holding, for example, all subjects' blood pressure to 120/80, as this might make the subjects healthier than they would be even if they did not smoke.

Daniel et al. [2015] and Steen et al. [2017] introduce estimands with two mediators, discuss estimation of natural direct, controlled direct, and indirect effects, and show the decompositions of the total effect into the natural effects. However, there are several other direct, and indirect, effect estimands that are crosses of natural and controlled effects that can be considered when there are multiple mediators. We define the cross natural and controlled direct effects in the setting of two potentially sequential mediators. No work, to our knowledge, has introduced these estimands previously.

Although it is not immediately clear if these estimands are part of some form of decomposition of the total effect, we believe they are of interest in any setting where one might consider holding one mediator to a constant level while holding the other mediator its natural level under treatment or control. Although we focus on conceptual examples from immunology in vaccination trials in COVID-19, one could imagine that these effects would be of interest in many other scenarios such as political science, sociology and engineering. For example, crosses of controlled and natural effects might be of interest when considering the direct effect of a law that reduces school funding on graduation rates while

holding infrastructure decline to zero and allowing teacher to student ratios to be held at the natural level they would be under the new law.

Regardless of the interest in these estimands, when a mediator is confounded, i.e. having a common cause, with the outcome of interest then conditioning on the mediator opens a non-causal pathway from the intervention to the outcome, even if the intervention is randomized. When these confounders can be controlled for these cross direct effects can be identified under certain assumptions. We provide a set of set of assumptions that we prove allow for the identification and point estimation of our proposed cross direct effects. Identification requires the assumption of no unmeasured or uncontrolled confounders between the mediator and the outcome, which is strong and untestable.

Instead of, or in addition to, providing a point estimate under this strong assumption, nonparametric bounds can provide a range guaranteed to include the unidentifiable causal effect of interest in the presence of uncontrolled confounding. Although potentially informative, i.e., not ranging from [-1,1], numeric bounds are often possible regardless of the type of variables in your data, tight symbolic bounds can be derived using linear programming when all measure variables are binary. Symbolic bounds have the advantage of not needing to be re-derived for each new dataset as them apply to all data conforming to the causal model. Cai et al. [2007] used the linear programming technique of Balke and Pearl [1994] to derive tight symbolic bounds for the CDE in a setting with one binary mediator and binary outcome and intervention. Sjölander [2009] extended the bounds of Cai et al. [2007] to the natural effects in the same setting also using linear programming. Gabriel et al. [2021] derived tight symbolic bounds for a large number of decomposition effects as

well as the CDE in the setting of two sequential mediators, where all measured variables are binary using the linear programming method of Sachs et al. [2021]. We use this same method to provide tight symbolic bounds for out proposed cross direct effects in settings where all measured variables are binary.

The paper is organized as follows. In Section 2 we provide our notation and outline our proposed estimands and settings of interest. In Section 3 we discuss some conceptual examples of scenarios where these estimands would be of interest. In Section 4 we provide a set of sufficient assumptions and equations for identification of our proposed cross direct effects. In Section 5 we provide the bounds for each of the settings and estimands of interest. In Section 6 we illustrate our estimands and calculate the point estimates and bounds in two datasets: a real data example from a framing experiment conducted in psychology, and a synthetic data example from an mRNA vaccine trial to prevent SARS-CoV-2 infection. Finally, in Section 7 we outline the limitations and future areas of research.

2 Preliminaries

2.1 Notation and Settings

Let X and Y be the intervention and outcome of interest, respectively. Let M_1 and M_2 be two mediators on the path from X to Y. Let U be an unmeasured set of confounders between Y, M_1 and M_2. Let C be a measured set of confounders between Y, M_1 and M_2. Let Ω_{Q_j} denote the support of the arbitrary variable Q_j.

The DAG in Figure 1a encodes the following nonparametric structural equation models

(NPSEM) [Pearl, 2009].

$$
\begin{aligned}
y &= g_Y(u, m_1, m_2, x, c, \epsilon_Y) \\
m_1 &= g_{M_1}(u, x, c, \epsilon_{M_1}) \\
m_2 &= g_{M_2}(u, x, c, \epsilon_{M_2}) \\
x &= g_X(\epsilon_X) \\
c &= g_C(\epsilon_C),
\end{aligned}
$$

for some response functions $g_Y, g_{M_1}, g_{M_2}, g_X, g_C$. The unmeasured variables U and the set of ϵ represent errors due to omitted factors. Given the values of the errors and the values of a variable's parents in the graph, the value of the variable is determined by its response function. The errors determine the manner in which the variable is determined from its parents, and we note that Y, M_1, M_2 have a common error variable U. In Figure 1b we additionally allow for M_1 to affect M_2, so that the structural equation for m_2 is instead:

$$
m_2 = g_{M_2}(u, x, m_1, c, \epsilon_{M_2});
$$

all other equations are the same.

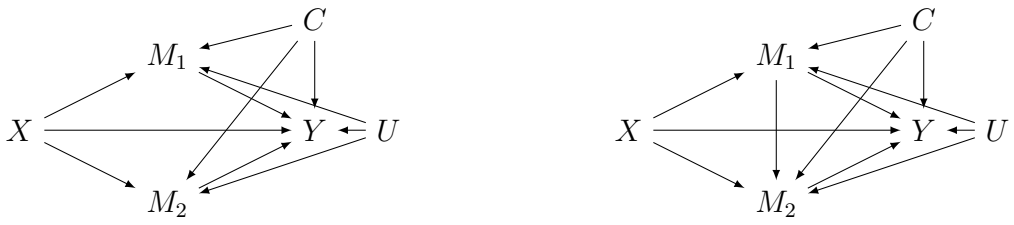

Figure 1: Causal diagrams of the settings of interest. X is the intervention, M_1, M_2 the two mediators, and Y the outcome. C and U denoted measured and unmeasured confounders, respectively.

2.2 Estimands

We let $Y(x)$ denote the potential outcome Y under the intervention which sets X to level x, and $Y(x, m_1, m_2)$ be the equivalent under an intervention which sets X to level x, M_1 to m_1, and M_2 to m_2. These potential outcomes are the same under the settings of Figure 1a and 1b. There are several other potential outcomes of relevance in the setting with two mediators, depending on whether there is a causal effect of one of the mediators M_1 on the other M_2, as in Figure 1b, or whether such effect is absent, as in Figure 1a. Define $Y(x, m_1, M_2(x_2))$ to be the potential outcome Y, under an intervention that holds$X$ to x, M_1 to m_1, and M_2 to the level it would take under an intervention holding X to x_2; similarly, define $Y(x, M_1(x_1), m_2)$. Define $Y(x, M_1(X = x_1), M_2(x_2))$ to be the potential outcome Y, under the interventions that set X to x, and M_1 and M_2 to the level they would take under the intervention that sets X to x_1, and x_2, respectively.

When M_1 has an effect on M_2, as in Figure 1b, there are novel and additional estimands to consider, based on different potential outcomes. Let $Y(x, m_1, M_2(x_2, m_1'))$ be the potential outcome Y, under the intervention that sets X to x, and M_1 to m_1 and M_2 to the level it would take on under the intervention that sets X to x_2 and M_1 in the path to

M_2 to m_1'. Similarly, let $Y(x, M_1(x_1), M_2(x_2, M_1(x_3)))$ be the potential outcome Y, under the interventions that set X to x, and M_1 to the level it would take on under under the intervention that sets X to x_1, and M_2 to the level it would take on under the intervention that sets X to x_2 and M_1 in the path to M_2 to the level it would take if X was set to x_3. Finally, define $Y(x, m_1, M_2(x_2), M_1(x_3)))$ to be the potential outcome of Y, under the interventions that set X to x, M_1 to m_1 and M_2 to the level it would take on under under the intervention that sets X to x_2 and M_1 in the path to M_2 to the level it would take had X been set to x_3.

When the intervention is randomized the total effect of X on Y, which we define in counterfactual terms as,

$$\text{TE-}xx' = E\{Y(x)\} - E\{Y(x')\}$$

is identified. We are interested in the direct effect of X on Y holding each mediator to either a constant level, or the level they would have taken for each subject had X been set to some level x. Daniel et al. [2015] describe the controlled direct effects (CDE) and the natural direct effects (NDE) for two or more mediators, as well as indirect effects and decompositions of the TE in these settings.

Although this was not discussed in Daniel et al. [2015], when there are two mediators, one mediator could be held or controlled to a constant level, while the other was set to the level it would take on for each subject if the intervention were set to a constant level. We call these the cross natural and controlled direct effect (CNCDE). Under Figure 1a for two

mediators there are two relevant estimands of this type:

$$\text{CN}_1\text{CD}_2\text{E-}xx' : x_1m_2 = E\{Y(x, M_1(x_1), M_2 = m_2)\} - E\{Y(x', M_1(x_1), M_2 = m_2)\}$$

and,

$$\text{CN}_2\text{CD}_1\text{E-}xx' : m_1x_2 = E\{Y(x, M_1 = m_1, M_2(x_2))\} - E\{Y(x', M_1 = m_1, M_2(x_2)\}.$$

Under Figure 1b there are three additional relevant estimates of this cross natural and controlled direct effect type:

$$\text{CN}_2\text{CD}_{11}\text{E-}xx' : m_1x_2m_3 = E\{Y(x, M_1 = m_1, M_2(x_2, m_3))\} - E\{Y(x', M_1 = m_1, M_2(x_2, m_3))\},$$

$$\begin{aligned}\text{CN}_{21}\text{CD}_1\text{E-}xx' : m_1x_2x_3 &= E\{Y(x, M_1 = m_1, M_2(x_2, M_1(x_3)))\} \\ &- E\{Y(x', M_1 = m_1, M_2(x_2, M_1(x_3)))\} \text{ and,}\end{aligned}$$

$$\text{CN}_{12}\text{CD}_1\text{E-}xx' : x_1x_2m_1 = E\{Y(x, M_1(x_1), M_2(x_2, m_1))\} - E\{Y(x', M_1(x_1), M_2(x_2, m_1))\}.$$

Within the settings depicted in Figure 1, we will focus on the cross direct effects introduced above, although similar indirect effects are likely also of interest. One can also see that there are direct extensions of these cross direct effects in settings with three or more, possibly sequential, mediators.

These estimands may seem like pure thought experiments, they may be useful for as-

sessing mechanistic pathways and therefore potentially mechanistic surrogate evaluation [Plotkin and Gilbert, 2012]. Additionally, although it is not immediately clear how or if these effects are part of a decomposition of the total effect, we believe they are of interest individually, particularly in immunology.

3 Conceptual Examples

Goel et al. [2021] found that one week after the second dose of vaccination with an mRNA vaccine (Pfizer BNT162b2 or Moderna mRNA-127) SARS-CoV-2 infection naive subjects had seeming independence of the proportion of memory B-cells that recognize spike protein (M_1) and the amount of circulating anti-spike protein IGg antibody (M_2). Thus, Figure 1a may be a depiction of this setting.

There are other mechanisms of vaccine protection beyond B-cells and antibody. To better understand the suite of mechanisms, it would be of interest to estimate the direct effect of vaccination by holding one of these mediators to an undetectable level and the other to the natural level that would occur under vaccination, or non-vaccination; for example, the percentage of mediation by anti-bodies was of interest in Gilbert et al. [2021]. In this case the estimands $\mathrm{CN}_1\mathrm{CD}_2\mathrm{E}\text{-}10:10$ or $\mathrm{CN}_2\mathrm{CD}_1\mathrm{E}\text{-}10:01$ are of interest.

Goel et al. [2021] also found that baseline memory B-cells prior to vaccine with an mRNA vaccine in SARS-CoV-2 recovered individuals was predictive of circulating antibody levels post vaccination. Although baseline memory B-cell percentage cannot be a mediator, the percentage of memory B-cells post dose one of mRNA vaccination is likely highly correlated with baseline levels. Thus, Figure 1b might be a depiction of the percentage of

spike positive memory B-cells after the first dose of the vaccine as mediator one, or (M_1) in Figure 1b, and the circulating anti-spike IGg post dose two of the vaccine as mediator two, or (M_2) in Figure 1b. Therefore, in previously infected individuals, CN_2CD_{11}E-10 : 010, $CN_{12}CD_1$E-10 : 110, may be of interest to help tease out the primary route of protection.

Another manipulation of interest would be to hold antibodies to their natural level following vaccination (ab=M2(1)) but holding the memory B-cells to undetectable (0=m1) thus eliminating any further production of antibody. We denote this by CN_2CD_{11}E-10 : 01ab. This estimand is useful for understanding the effect of infusing exogenous antibodies to levels achieved by vaccination in immuno-compromised individuals who cannot induce a vaccinal immune response. If the direct effect remains via a notably high CN_2CD_{11}E-10 : 01ab, passive transfer may not be effective. Looking at these estimates may motivate the use of passive transfer or the investigation of other options in immuno-compromised individuals.

4 Sufficient assumptions for identification

All cross direct effects are nonparametrically identified under the assumptions labeled in Daniel et al. [2015] as MC.1-MC.2 and MCN.3-MCN.5. These are implied by our NPSEMs

above in the absence of U. Under Figure 1a the NPSEMs become

$$\begin{aligned} y &= g_Y(x, c, m_1, m_2, \epsilon_Y) \\ m_1 &= g_{M_1}(x, c, \epsilon_{M_1}) \\ m_2 &= g_{M_2}(x, c, \epsilon_{M_2}) \\ x &= g_X(\epsilon_X), \\ c &= g_C(\epsilon_C), \end{aligned}$$

and under Figure 1b the equation for m_2 becomes

$$m_2 = g_{M_2}(x, c, m_1, \epsilon_{M_2})$$

with all others remaining the same. Note that, given the control on M_1 or M_2 and/or the lack of an effect on M_1 on M_2, it may be possible to relax these assumptions to allow for intermediate confounders that are affected by X. This was investigated in Daniel et al. [2015] for the NDE and CDE, but we do not offer further investigation here for our proposed cross direct effects.

Proposition 1: Under the above NPSEMs we have:

$$\begin{aligned}
\text{a.) } \mathrm{CN_1CD_2E}\text{-}xx' : x_1m_2 &= E\{Y(x, M_1(x_1), m_2)\} - E\{Y(x', M_1(x_1), m_2)\} \\
E\{Y(x, M_1(x_1), m_2)\} &= \int_{\Omega_C}\int_{\Omega_{M_1}} \mathrm{E}\{Y|C=c, X=x, M_1=m_1, M_2=m_2\} \\
&\quad \times f_{M_1|C,X}(m_1|c, x_1) f_C(c) dm_1 dc.
\end{aligned}$$

By symmetry, when there is no effect of M_1 on M_2, as in Figure 1a, one can simply switch the M_1 and M_2, in the above. When there is an effect of M_1 on M_2, this is not the case and the other cross direct effects become more relevant.

$$\begin{aligned}
\text{b.) } \mathrm{CN_2CD_{11}E}\text{-}xx' : m_1x_2m_1' &= E\{Y(x, m_1, M_2(x_2, m_1'))\} - E\{Y(x', m_1, M_2(x_2, m_1'))\} \\
E\{Y(x, m_1, M_2(x_2, m_1'))\} &= \int_{\Omega_C}\int_{\Omega_{M_2}} \mathrm{E}\{Y|C=c, X=x, M_1=m_1, M_2=m_2\} \\
&\quad \times f_{M_2|C,X,M_1}(m_2|c, x_2, m_1') f_C(c) dm_2 dc.
\end{aligned}$$

$$\begin{aligned}
\text{c.) } \mathrm{CN_{21}CD_1E}\text{-}xx' : m_1x_2x_3 &= E\{Y(x, m_1, M_2(x_2, M_1(x_3)))\} - E\{Y(x', m_1, M_2(x_2, M_1(x_3)))\} \\
E\{Y(x, m_1, M_2(x_2, M_1(x_3)))\} &= \int_{\Omega_C}\int_{\Omega_{M_1}}\int_{\Omega_{M_2}} \mathrm{E}\{Y|C=c, X=x, M_1=m_1, M_2=m_2\} \\
&\quad \times f_{M_2|C,X,M_1}(m_2|c, x_2, m_1') f_{M_1|C,X}(m_1'|c, x_3) f_C(c) dm_2 dm_1' dc.
\end{aligned}$$

$$\text{d.) } \mathrm{CN}_{12}\mathrm{CD}_1\mathrm{E}\text{-}xx' : x_1x_2m_1' = E\{Y(x, M_1(x_1), M_2(x_2, m_1'))\} - E\{Y(x', M_1(x_1), M_2(x_2, m_1'))\}$$

$$E\{Y(x, M_1(x_1), M_2(x_2, m_1'))\} = \int_{\Omega_C}\int_{\Omega_{M_1}}\int_{\Omega_{M_2}} \mathrm{E}\{Y|C=c, X=x, M_1=m_1, M_2=m_2\} \times f_{M_2|C,X,M_1}(m_2|c, x_2, m_1')f_{M_1|C,X}(m_1|c, x_1)f_C(c)dm_2dm_1dc.$$

For any Q_j, $\int_{\Omega_{Q_j}}$ is replaced by $\sum_{\Omega_{Q_j}}$ if Q_j is discrete.

The proof of Proposition 1 is in the web Appendix. We note that one of the assumptions required to prove the above is no unmeasured confounding, an untestable assumption.

5 Novel bounds for binary data

When there is unmeasured or uncontrolled confounding, none of the estimands are identified. In settings where Y, X, M_1 and M_2 are all binary we can bound the estimands using the linear programming method [Sachs et al., 2021] implemented in the `R` package `causaloptim` to obtain tight nonparametric bounds.

These bounds apply within levels of a categorical C, or marginally if C is the empty set. We allow here for an arbitrary set of unmeasured confounders U. Define the short hand notation for probabilities as:

$$p_{ym_1m_2.x} = p\{Y=y, M_1=m_1, M_2=m_2|X=x\}.$$

For example, $p_{111.1} = p\{Y = 1, M_1 = 1, M_2 = 1|X = 1\}$.

Result 1:

The bounds given below are valid and tight for $\text{CN}_1\text{CD}_2\text{E-10:00} = E\{Y(1, M_1(0), M_2 = 0)\} - E\{Y(0, M_1(0), M_2 = 0)\}$ under Figure 1a or 1b.

$$\text{CN}_1\text{CD}_2\text{E-10:00} \geq$$
$$\max \left\{ \begin{array}{l} -1 - p_{100\cdot0} + p_{010\cdot0} + p_{110\cdot1} - p_{001\cdot0} - p_{101\cdot0} \\ -2 + 2p_{000\cdot0} + p_{100\cdot0} + p_{100\cdot1} + p_{010\cdot0} + p_{001\cdot0} + p_{101\cdot0} \\ -1 + p_{000\cdot0} + p_{010\cdot0} \end{array} \right\},$$

and

$$\text{CN}_1\text{CD}_2\text{E-10:00} \leq$$
$$\min \left\{ \begin{array}{l} 1 + p_{000\cdot0} - p_{010\cdot1} - p_{110\cdot0} + p_{001\cdot0} + p_{101\cdot0} \\ 1 - p_{100\cdot0} - p_{110\cdot0} \\ 2 - p_{000\cdot0} - p_{000\cdot1} - 2p_{100\cdot0} - p_{110\cdot0} - p_{001\cdot0} - p_{101\cdot0} \end{array} \right\}.$$

Under Figure 1a, the bounds in Result 1 apply directly to the estimand $\text{CN}_2\text{CD}_1\text{E-10} : m_1x_2$ by symmetry as M_1 does not have an effect on M_2. This is not the case in Figure

1b, but it is also not clear if $\text{CN}_2\text{CD}_1\text{E-10} : m_1x_2$ would be of relevance in that setting. If there is an effect of M_1 on M_2, as in Figure 1b, the bounds must account for this effect if M_2 is not being held to a constant level for all individuals.

Result 2:

The bounds given below are valid and tight for $\text{CN}_2\text{CD}_{11}\text{E-10} : 000 = E\{Y(1, M_1 = 0, M_2(X = 0, M_1 = 0))\} - E\{Y(0, M_1 = 0, M_2(X = 0, M_1 = 0))\}$ under Figure 1b.

$$\text{CN}_2\text{CD}_{11}\text{E-10:000} \geq$$
$$\max \left\{ \begin{array}{l} -2 + p_{000\cdot 0} + 2p_{001\cdot 0} + p_{101\cdot 0} + p_{101\cdot 1} \\ -2 + 2p_{000\cdot 0} + p_{100\cdot 0} + p_{100\cdot 1} + p_{001\cdot 0} \\ -1 + p_{000\cdot 0} + p_{001\cdot 0} \end{array} \right\},$$

and

$$\text{CN}_2\text{CD}_{11}\text{E-10:000} \leq$$
$$\min \left\{ \begin{array}{l} 2 - p_{100\cdot 0} - p_{001\cdot 0} - p_{001\cdot 1} - 2p_{101\cdot 0} \\ 1 - p_{100\cdot 0} - p_{101\cdot 0} \\ 2 - p_{000\cdot 0} - p_{000\cdot 1} - 2p_{100\cdot 0} - p_{101\cdot 0} \end{array} \right\}.$$

Result 3:

The bounds given below are valid and tight for $\text{CN}_{12}\text{CD}_1\text{E-10} : 000 = E\{Y(1, M_1(0), M_2(X = 0, M_1 = 0))\} - E\{Y(0, M_1(0), M_2(X = 0, M_1 = 0))\}$ under Figure 1b.

$\text{CN}_{12}\text{CD}_1\text{E-10:000} \geq$

$$\max \left\{ \begin{array}{l} -2 + p_{000\cdot 0} + 2p_{001\cdot 0} + p_{101\cdot 0} + p_{101\cdot 1} \\ -2 + 2p_{000\cdot 0} + p_{100\cdot 0} + p_{100\cdot 1} + p_{001\cdot 0} \\ -1 + p_{000\cdot 0} + p_{001\cdot 0} \\ -1 + p_{000\cdot 0} - p_{000\cdot 1} - p_{100\cdot 1} + p_{010\cdot 0} - p_{010\cdot 1} - p_{110\cdot 1} + p_{001\cdot 0} - p_{001\cdot 1} - p_{101\cdot 1} - p_{011\cdot 1} \\ -2 + p_{000\cdot 0} + p_{110\cdot 1} + p_{001\cdot 0} + p_{011\cdot 0} \end{array} \right\},$$

and

$\text{CN}_{12}\text{CD}_1\text{E-10:000} \leq$

$$\min \left\{ \begin{array}{l} 2 - p_{100\cdot 0} - p_{001\cdot 0} - p_{001\cdot 1} - 2p_{101\cdot 0} \\ 2 - p_{100\cdot 0} - p_{110\cdot 0} - p_{101\cdot 0} - p_{011\cdot 1} \\ 1 + p_{000\cdot 0} + p_{010\cdot 0} - p_{010\cdot 1} + p_{110\cdot 0} + p_{001\cdot 0} + p_{011\cdot 0} \\ 1 - p_{100\cdot 0} - p_{101\cdot 0} \\ 2 - p_{000\cdot 0} - p_{000\cdot 1} - 2p_{100\cdot 0} - p_{101\cdot 0} \end{array} \right\}.$$

Result 4:

The bounds given below are valid and tight for $\text{CN}_{21}\text{CD}_1\text{E-10} : 000 = E\{Y(1, M_1 = 0, M_2(X = 0, M_1(0)))\} - E\{Y(0, M_1 = 0, M_2(X = 0, M_1(0)))\}$ under Figure 1b.

$$\text{CN}_{21}\text{CD}_1\text{E-10:000} \geq$$

$$\max\left\{\begin{array}{l} -1 - p_{100\cdot 0} - p_{010\cdot 0} - p_{110\cdot 0} + p_{001\cdot 0} + p_{101\cdot 1} \\ -2 + 2p_{000\cdot 0} + p_{100\cdot 0} + p_{100\cdot 1} + p_{010\cdot 0} + p_{110\cdot 0} + p_{001\cdot 0} \\ -1 + p_{000\cdot 0} + p_{001\cdot 0} \end{array}\right\},$$

and

$$\text{CN}_{21}\text{CD}_1\text{E-10:000} \leq$$

$$\min\left\{\begin{array}{l} 1 + p_{000\cdot 0} + p_{010\cdot 0} + p_{110\cdot 0} - p_{001\cdot 1} - p_{101\cdot 0} \\ 1 - p_{100\cdot 0} - p_{101\cdot 0} \\ 2 - p_{000\cdot 0} - p_{000\cdot 1} - 2p_{100\cdot 0} - p_{010\cdot 0} - p_{110\cdot 0} - p_{101\cdot 0} \end{array}\right\}.$$

The bounds corresponding to the other levels of $\{x_1, x_2, x_3\}$ are given in the supplementary materials; results are labeled based on the estimand, 1,2,3 or 4, and then b to h.

6 Data Examples

We illustrate our novel estimands using two data sets. The first is a real data example, the framing data from the `mediation` package in R [Tingley et al., 2014]. The framing data contains 265 rows and 15 columns of data from a framing experiment conducted by Brader

et al. [2008]. The experiment was to expose subjects to a news story that is randomized to a tone, positive or negative, and to whether the news story features a Latino or European immigrant. We consider the binary intervention, which is 1 if the news story features a Latino immigrant with negative tone and 0 otherwise (either positive tone or European immigrant), as this was the treatment considered in the original study. These data include covariates and have an ordinal outcome and mediators and we use the outcome of the four-point scale measuring a subject's attitude toward increased immigration, with larger values indicate more negative attitudes, by dichotomizing into 1 or 2 as low, and 3 or 4 as high. Similarly for the mediators, of a subject's negative feeling during the experiment, which was originally measured on the numeric scale ranging between 3 and 12, we dichtomized this as low, below 8 and high above or equal to 8, and a subject's perceived harm caused by increased immigration. This perceived harm was originally on a numeric scale between 2 and 8, which we dichtomize to zero for less than 6, and one for 6 or greater. We consider the dichtomized emotional state as M_1 which may causally impact the perceived harm, M_2. The additional baseline covariates in the dataset we consider are age, highest level of education, gender, and income.

For point estimation of the effects of interest, we assume that the measured covariates are the only confounders. We then specified logistic regression models for $Y|C, X, M_1, M_2$, $M_1|C, X$, and $M_2|C, X, M_1$ that include all covariates and all two-way interactions among

the covariates, treatment, and mediators. We estimated $E\{Y(x, m_1, M_2(x_2, m_1'))\}$ as

$$\hat{E}\{Y(x, m_1, M_2(x_2, m_1'))\} \quad =$$
$$= n^{-1} \sum_{i=1}^{n} \sum_{j \in \{0,1\}} \hat{p}\{Y = 1 | C = C_i, X = x, M_1 = m_1, M_2 = j\}$$
$$\times \hat{p}\{M_2 = j | C = C_i, X = x_2, M_1 = m_1'\}.$$

In this expression, the estimated probabilities were obtained from the logistic regression models, and the population distribution of C is estimated with its sample distribution. All other terms of our estimands were estimated analogously. For inference we use the nonparametric bootstrap. Although our estimation procedure can be followed directly when the exposure, outcome and mediators are binary, we believe that estimation in other settings that are identified can directly follow that suggested in Daniel et al. [2015].

While some of the confidence intervals exclude the null value of 0, these estimates rely on the untestable assumptions of no unmeasured confounding and correct specification of the models. The bounds do not rely on these assumptions and although all bounds cover zero in this example, the width of the bounds and the causal effects they cover differ greatly over the estimands. As can be seen in Figure 2, $\mathrm{CN_1CD_2E}$-10 : 11, which is the same under either Figure 1a or 1b, have mostly positive values, implying mostly a range of harm or increased risk of negative attitudes. This suggests that a negatively toned news article about a Latino immigrant is most likely to cause negative attitudes, although no change cannot be ruled out, if subjects have a high emotional state and the perceived harm they would have had if they had been exposed to the negatively toned article about a Latino.

Compared with the estimand $CN_{12}CD_1E\text{-}10:110$ which has a smaller point estimate near 0, and wider bounds ranging from around -0.75 to 0.9, we can infer little about the cause of a negatively toned news article about a Latino immigrant controlling the emotional state at 0 in its effect on perceived harm, but otherwise having emotional state and perceived harm at their natural levels under treatment.

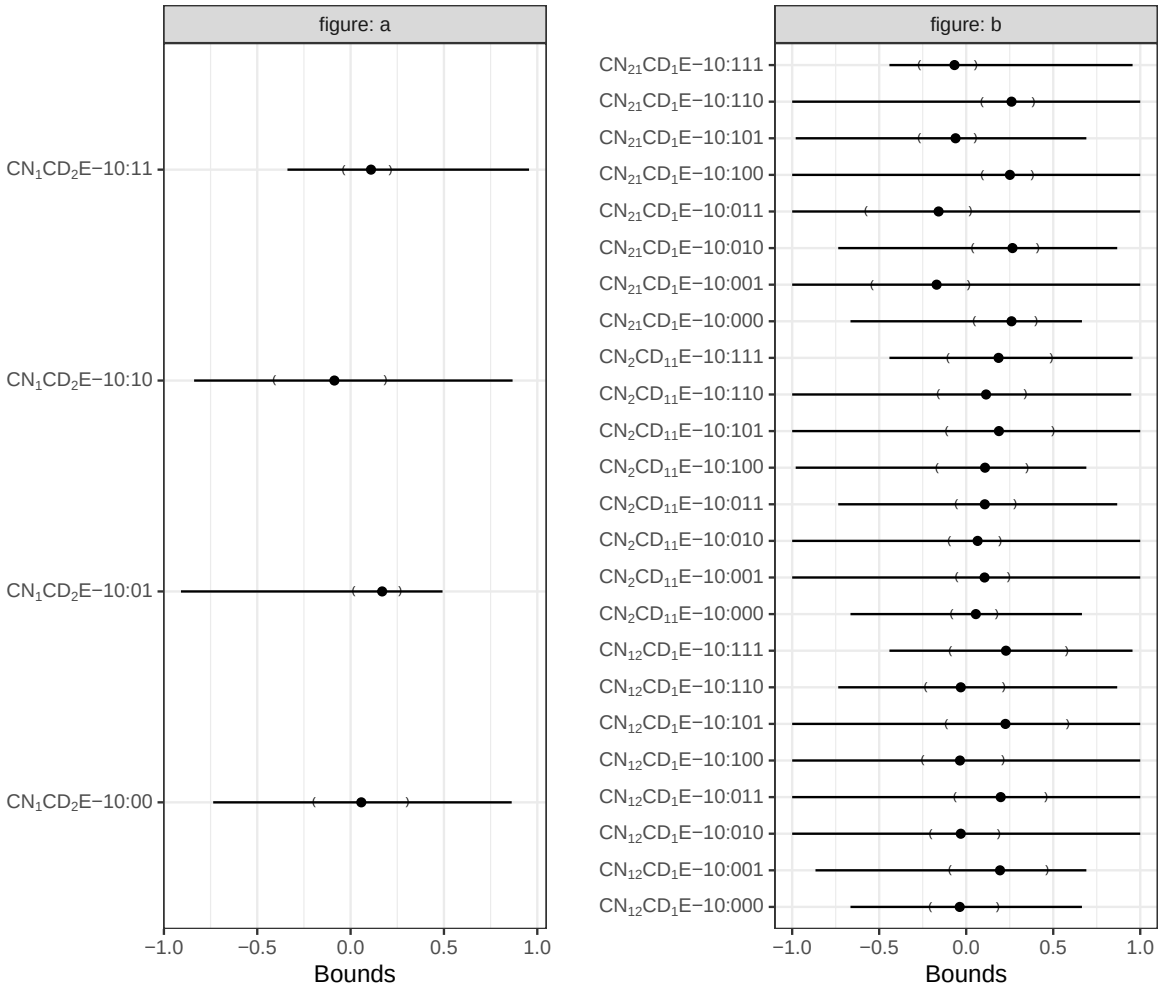

Figure 2: Point estimates (dots) and bootstrap 95% confidence intervals (left and right brackets) under the assumption of no unmeasured confounding and computed bounds (line ranges) for the different effects using the framing dataset.

The second data example is a synthetic data set constructed based on the findings of

Gilbert et al. [2021], who analyzed the effect of circulating antibody post second dose of the Moderna mRNA vaccine on the disease risk and vaccine efficacy. We generate a continuous B-cell response (M1) centered at 1 that would have been just after vaccination. This B-cell response causes, in part, the circulating antibody that is measured two weeks after the second dose (M2). Based on the B-cell and antibody responses we generate a disease indicator. Consistent with Gilbert et al. [2021], the overall vaccine efficacy is 0.92 and VE at M1,M2=(100,1) is 0.91 and at (M1,M2)=(1000,1) is 0.96. We dichotomized these mediators at 1000 and 1, respectively, and the bounds for the synthetic data are given in Figure 3.

There are no covariates in this example, so the point estimates are computed using the same expressions but with C as the empty set. The point estimates are all close to zero, and with quite narrow 95% confidence intervals. All bounds cover zero. Here, CN_1CD_2E-10 : 00 under both panels is nearly always positive while CN_1CD_2E-10 : 10 is nearly always negative. The former can be interpreted as the effect of vaccination holding the antibody response to 0 and the B-cell response to what it would have been under no vaccination while the latter is the effect of vaccination holding the antibody response to 0 and the B-cell response to what it would have been under vaccination. Thus, if this were observed in real data, one might infer that the protection offered by the vaccine may primarily be through the mechanism of B-cell response. Also, the bounds on $\text{CN}_2\text{CD}_{11}$E-10 : 010 rule out large positive risk differences, hence among previously infected individuals, if this were observed in real data, it may suggest that the passive transfer of antibodies alone would not explain the efficacy of vaccination.

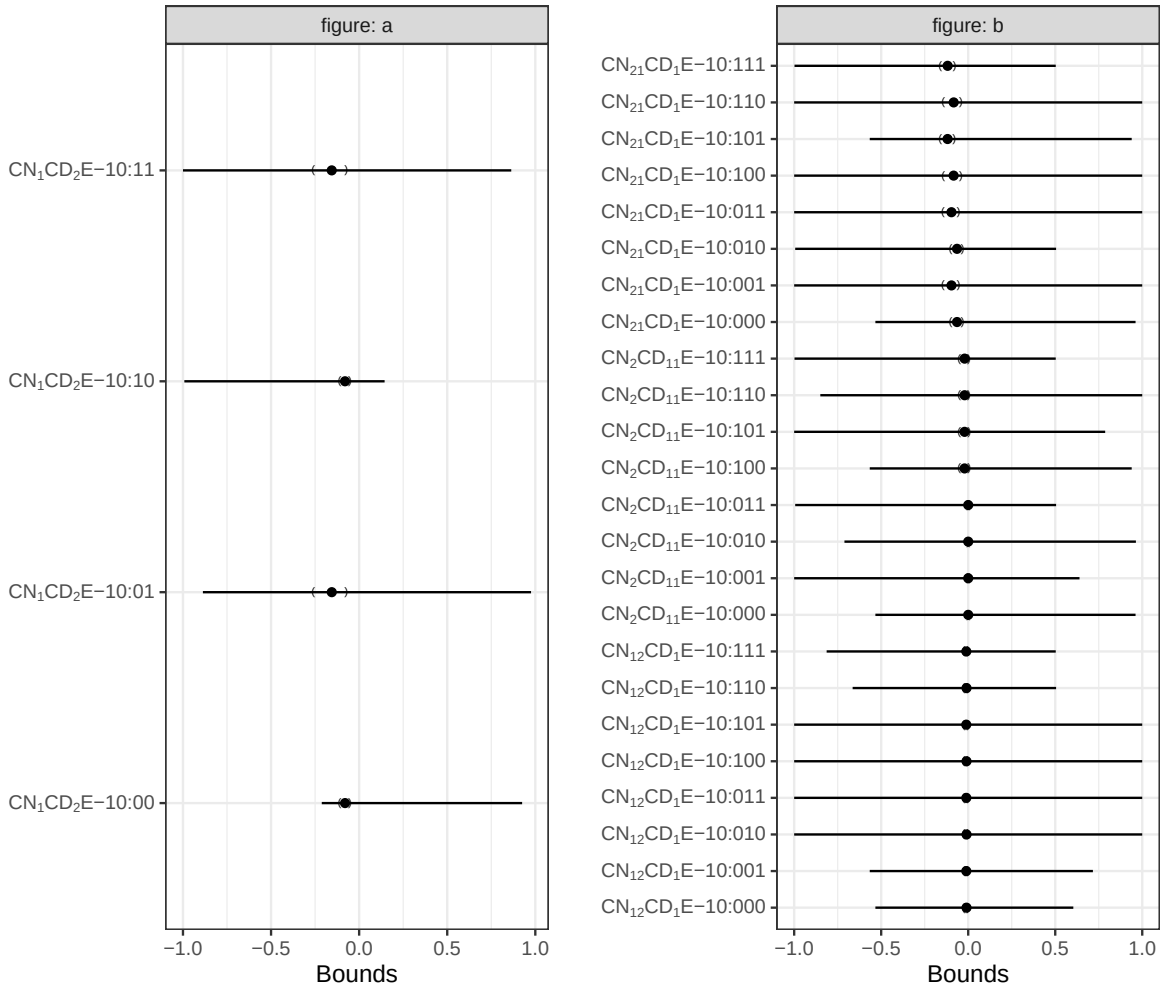

Figure 3: Computed bounds for the different effects using the synthetic covid dataset.

7 Discussion

We present several novel direct effects when there are two, potentially sequential, mediators that are likely of interest in immunology. We show that the same set of sufficient assumptions for identification given in Daniel et al. [2015] are sufficient for these cross natural and controlled direct effects. When these assumptions do not hold due to unmeasured confounding, we provide a means of bounding these estimands when all measured variables are binary.

The estimands, although to our knowledge novel, have likely been considered previously in immunological settings due to their clear alignment with the goals of understanding mechanisms. These investigations have likely been undertaken without consideration of how or if the estimands are identifiable. We hope that this work provides a means for further investigation of these estimands, and their use, particularly in the settings of vaccine immunology.

This paper is limited in two ways, by the lack of discussion of the corresponding cross indirect effects and a inference procedure for the bounds. Although we believe estimation of the bounds in the unidentified settings to be straightforward we do not further discuss inference. However we suggest the inference about the bounds can be obtained using the bootstrap procedure suggested and investigated in Gabriel et al. [2020]. Consideration of the indirect estimands and if the cross direct and indirect effects form some type of decomposition is an area of future research for the authors.

References

Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, pages 46–54. Morgan Kaufmann Publishers Inc., 1994.

Ted Brader, Nicholas A Valentino, and Elizabeth Suhay. What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. *American Journal of Political Science*, 52(4):959–978, 2008.

Zhihong Cai, Manabu Kuroki, and Tosiya Sato. Non-parametric bounds on treatment effects with non-compliance by covariate adjustment. *Statistics in medicine*, 26(16): 3188–3204, 2007.

Rhian M Daniel, Bianca L De Stavola, SN Cousens, and Stijn Vansteelandt. Causal mediation analysis with multiple mediators. *Biometrics*, 71(1):1–14, 2015.

Erin E. Gabriel, Michael C. Sachs, and Arvid Sjölander. Causal bounds for outcome-dependent sampling in observational studies. *Journal of the American Statistical Association*, In press, 2020.

Erin E Gabriel, Michael C Sachs, and Arvid Sjölander. Sharp nonparametric bounds for decomposition effects with two mediators. *arxiv*, x, 2021.

Peter B. Gilbert, David C. Montefiori, Adrian McDermott, Youyi Fong, David Benkeser, Weiping Deng, Honghong Zhou, Christopher R. Houchens, Karen Martins, Lakshmi Jayashankar, Flora Castellino, Britta Flach, Bob C. Lin, Sarah O'Connell, Charlene McDanal, Amanda Eaton, Marcella Sarzotti-Kelsoe, Yiwen Lu, Chenchen Yu, Bhavesh Borate, Lars W. P. van der Laan, Nima Hejazi, Chuong Huynh, Jacqueline Miller, Hana M. El Sahly, Lindsey R. Baden, Mira Baron, Luis De La Cruz, Cynthia Gay, Spyros Kalams, Colleen F. Kelley, Mark Kutner, Michele P. Andrasik, James G. Kublin, Lawrence Corey, Kathleen M. Neuzil, Lindsay N. Carpp, Rolando Pajon, Dean Follmann, Ruben O. Donis, Richard A. Koup, Inc.; Coronavirus Vaccine Prevention Network (CoVPN)/Coronavirus Efficacy (COVE); on behalf of the Immune Assays; Moderna, and United States Government (USG)/CoVPN Biostatistics Teams. Immune correlates analysis of the mrna-1273

covid-19 vaccine efficacy trial. *medRxiv*, 2021. doi: 10.1101/2021.08.09.21261290. URL https://www.medrxiv.org/content/early/2021/08/15/2021.08.09.21261290.

R. Goel, S. Apostolidis, M. M. Painter, D. Mathew, A. Pattekar, O. Kuthuru, S. Gouma, P. Hicks, W. Meng, Aaron M. Rosenfeld, Sarah Dysinger, K. A. Lundgreen, L. Kuri-Cervantes, S. Adamski, A. Hicks, S. Korte, D. Oldridge, A. Baxter, J. Giles, M. Weirick, C. McAllister, J. Dougherty, S. Long, K. D'Andrea, J. Hamilton, M. Betts, E. L. Luning Prak, P. Bates, S. Hensley, A. Greenplate, and E. Wherry. Distinct antibody and memory b cell responses in sars-cov-2 naïve and recovered individuals following mrna vaccination. *Science Immunology*, 6, 2021.

J. Pearl. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2nd edition, 2009.

Stanley A Plotkin and Peter B Gilbert. Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases*, 54(11):1615–1617, 2012.

Michael C Sachs, Gustav Jonzon, Arvid Sjölander, and Erin E Gabriel. A general method for deriving tight symbolic bounds on causal effects, 2021. URL https://arxiv.org/abs/2003.10702.

A. Sjölander. Bounds on natural direct effects in the presence of confounded intermediate variables. *Statistics in Medicine*, 28(4):558–571, 2009.

Johan Steen, Tom Loeys, Beatrijs Moerkerke, and Stijn Vansteelandt. Flexible mediation analysis with multiple mediators. *American journal of epidemiology*, 186(2):184–193, 2017.

Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5): 1–38, 2014. URL `http://www.jstatsoft.org/v59/i05/`.