**Research Article**

Arvid Sjölander* and Ingeborg Waernbaum

# Improved sensitivity bounds for mediation under unmeasured mediator–outcome confounding

https://doi.org/10.1515/jci-2023-0070
received October 19, 2023; accepted May 10, 2024

**Abstract:** It is often of interest to decompose a total effect into an indirect effect, relayed through a particular mediator, and a direct effect. However, these effect components are not identified if there are unmeasured confounding of the mediator and the outcome. We derive nonparametric bounds on the natural direct and indirect effects, and Cornfield inequalities that the unmeasured confounders must satisfy to explain away an "observed" effect. We demonstrate, analytically and by simulation, that these bounds and Cornfield inequalities are sharper than those previously proposed in the literature. We illustrate the methods with an application to cholestyramine treatment for coronary heart disease.

**Keywords:** bounds, causal inference, mediation

**MSC 2020:** 92B15

## 1 Introduction

Mediation analysis is central in biostatistics, epidemiology, and related disciplines. The aim of such analysis is to decompose a total effect of an exposure on an outcome into an indirect effect, relayed through a particular mediator, and a direct (not through the mediator) effect. A common problem in mediation analysis is the presence of unmeasured confounding (common causes) of the mediator and the outcome, which makes direct and indirect effects non-identifiable, even if the exposure is not confounded with the mediator and the outcome [1,2].

Ding and Vanderweele [3], henceforth DV, proposed to handle this non-identifiability problem by computing bounds on the direct and indirect effects, i.e., ranges guaranteed to include the true values of the effects. In their proposed sensitivity analysis, the analyst specifies two sensitivity parameters, which measure the "maximal" conditional association between the confounders and the outcome and between the confounders and the exposure, given the mediator. Together with the observed data distribution, these sensitivity parameters translate into bounds on the direct and indirect effects according to simple analytic formulas.

We extend DV's results in three directions. First, we derive the feasible space (i.e., the space of logically possible values) for the sensitivity parameters, given the observed data distribution. This space determines what values an analyst may, and may not, consider in the computation of the bounds. Second, we derive novel bounds on the direct and indirect effects. Even though DV referred to their bounds as "sharp," they are not sharp in the absolute sense of being the tightest possible, given their assumptions. Our novel bounds are at least as sharp, and sometimes sharper, than DV's bounds. We show by a numerical example that the

*** Corresponding author: Arvid Sjölander**, Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Solna, Stockholm, Sweden, e-mail: arvid.sjolander@ki.se
**Ingeborg Waernbaum:** Department of Statistics, Uppsala University, Uppsala, Sweden, e-mail: ingeborg.waernbaum@ifau.uu.se

improvement over DV's bounds can be substantial. Furthermore, whereas DV only derived a lower bound for the direct effect and an upper bound for the indirect effect, we derive lower and upper bounds on both effects. Third, we derive Cornfield inequalities that the unmeasured confounders must satisfy to explain away an "observed" effect. These inequalities are at least as sharp, and sometimes sharper, than similar Cornfield inequalities for the direct effect derived by DV. Furthermore, whereas DV's Cornfield inequalities only apply to direct effects of positive magnitude, our Cornfield inequalities apply to both direct and indirect effects, having either positive or negative magnitudes.

In line with previous literature, we ignore uncertainty due to sampling variability, and assume that the exposure and outcome are binary. However, whereas DV allowed for a mediator of arbitrary type, we restrict attention to binary mediators. Proofs of all theorems, and R code for all analyses, are provided in the Appendix.

## 2 Notation, definitions, and assumptions

Let $A$, $M$, and $Y$ denote the binary exposure, mediator, and outcome, respectively. In addition to these, data may be available on a set of covariates $C$. However, since all DV's results, and ours, are derived within strata of covariates, we keep the conditioning on $C$ implicit in the notation. Therefore, the observed data distribution is given by $p(Y, M, A)$. Let $M_a$ and $Y_a$ be the counterfactual mediator and outcome, respectively, where $A$ set to $a$, and let $Y_{aM_{a'}}$ be the counterfactual outcome where $A$ set to $a$ and $M$ set to $M_{a'}$. Finally, define $p_{aa'} = p(Y_{aM_{a'}} = 1)$.

Like DV, we allow for unmeasured confounding of the mediator and the outcome, but we assume that the exposure is not confounded with the mediator and the outcome (given $C$). Specifically, we assume that data were generated by a nonparametric structural equation model represented by the causal diagram in Figure 1. In the Appendix, we provide an SWIG illustrating the setting [4]. Here, $U$ is the whole set of (unmeasured) common causes of $M$ and $Y$. This model implies the following three assumptions. Consistency of counterfactuals:

$$\begin{aligned} &A = a \Rightarrow M_a = M, \\ &A = a \Rightarrow Y_a = Y, \\ &A = a,\ M = m \Rightarrow Y_{am} = Y. \end{aligned} \tag{1}$$

Conditional independence between the set of mediator counterfactuals and the set of outcome counterfactuals, given $U$:

$$\{M_0, M_1\} \perp \{Y_0, Y_1, Y_{00}, Y_{01}, Y_{10}, Y_{11}\} | U. \tag{2}$$

Independence between the set of counterfactuals and the exposure and between $U$ and the exposure:

$$\{M_0, M_1, Y_0, Y_1, Y_{00}, Y_{01}, Y_{10}, Y_{11}, U\} \perp A. \tag{3}$$

DV made essentially the same assumptions but phrased slightly differently.

The target parameters are the natural direct effect (NDE) and natural indirect effect (NIE), which are defined as some contrasts

$$\text{NDE} = g(p_{10}) - g(p_{00}),$$

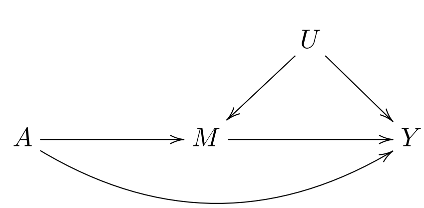


**Figure 1:** Assumed causal diagram.

and

$$\mathrm{NIE} = g(p_{11}) - g(p_{10}),$$

where $g$ is a link function. The identity link, log link, and logit link give the effects as risk differences, log risk ratios, and log odds ratios, respectively. DV only considered the risk difference and risk ratio, but their results do in fact apply more generally.

DV showed that $p_{00}$ and $p_{11}$ are identified and equal $p(Y = 1|A = 0)$ and $p(Y = 1|A = 1)$, respectively. Thus, identification of NDE and NIE crucially relies on identification of $p_{10}$. Define

$$p_{10}^{\mathrm{obs}} = \sum_m p(Y = 1|A = 1, M = m)p(M = m|A = 0),$$

and the "observed" NDE and NIE

$$\mathrm{NDE}^{\mathrm{obs}} = g(p_{10}^{\mathrm{obs}}) - g\{p(Y = 1|A = 0)\}$$

and

$$\mathrm{NIE}^{\mathrm{obs}} = g\{p(Y = 1|A = 1)\} - g(p_{10}^{\mathrm{obs}}).$$

DV showed that, in the absence of unmeasured confounding $U$, $p_{10}$ is identified and equal to $p_{10}^{\mathrm{obs}}$, so that $\mathrm{NDE}^{\mathrm{obs}} = \mathrm{NDE}$ and $\mathrm{NIE}^{\mathrm{obs}} = \mathrm{NIE}$. However, in the presence of unmeasured confounding, $p_{10}$, NDE and NIE are not identified without further assumptions.

## 3 DV's lower bound on $p_{10}$

DV defined the sensitivity parameters $\mathrm{RR}_{UY} = \max_m \mathrm{RR}_{UY|m}$ and $\mathrm{RR}_{AU} = \max_m \mathrm{RR}_{AU|m}$, where

$$\mathrm{RR}_{UY|m} = \frac{\max_u p(Y = 1|A = 1, M = m, U = u)}{\min_u p(Y = 1|A = 1, M = m, U = u)}$$

and

$$\mathrm{RR}_{AU|m} = \max_u \left\{\frac{p(U = u|A = 1, M = m)}{p(U = u|A = 0, M = m)}\right\}.$$

They defined the bounding factor

$$\mathrm{BF} = \frac{\mathrm{RR}_{AU} \times \mathrm{RR}_{UY}}{\mathrm{RR}_{AU} + \mathrm{RR}_{UY} - 1},$$

which is a monotonically increasing function of $\mathrm{RR}_{UY}$ and $\mathrm{RR}_{AU}$. They showed that, given BF, $p_{10}$ is bounded from below by

$$p_{10} \geq l_{DV} = p_{10}^{\mathrm{obs}} \mathrm{BF}^{-1}.$$

Once the analyst has specified $\mathrm{RR}_{UY}$ and $\mathrm{RR}_{AU}$, these can be used to compute BF and the lower bound on $p_{10}$. This bound can then be converted into a lower bound on NDE and an upper bound on NIE. For instance, a lower bound on NDE and an upper bound on NIE as risk ratios are given by $l_{DV}/p(Y = 1|A = 0)$ and $p(Y = 1|A = 1)/l_{DV}$, respectively.

## 4 Feasible space of DV's sensitivity parameters

To use DV's bounds in practice, it is important to know the feasible space of their sensitivity parameters, given the observed data distribution. This space is established by the following theorem.

**Theorem 1.** (i) $\{\mathrm{RR}_{AU}, \mathrm{RR}_{UY}\}$ *are restricted by their definitions to values* $\geq 1$. (ii) $\{\mathrm{RR}_{AU}, \mathrm{RR}_{UY}\}$ *are variation independent of each other and of the observed data distribution* $p(Y, M, A)$.

Part (ii) of the theorem implies that the feasible space of one of the sensitivity parameters is neither restricted by the value of the other sensitivity parameter, nor by the observed data distribution. This means that, regardless of what data distribution one may observe, one should consider all values $\geq 1$, as stated by part (i) of the theorem, as logically possible for both sensitivity parameters. This does not mean that all values $\geq 1$ are plausible for any given application; hopefully, one may be able to use subject matter knowledge to significantly reduce this space when computing the bounds.

## 5 A sharper lower bound on $p_{10}$

It follows from results by Sjölander [5] that $p_{10}$ is bounded from below by

$$p_{10} \geq l_{AF} = \max\{0, q_0, q_1\},$$

where

$$q_m = p(Y = 1, M = m|A = 1) - p(M = 1 - m|A = 0).$$

This bound is "assumption-free" in the sense that it does not require specification of BF (or any other sensitivity parameter). It follows from Theorem 1 that, for any given observed data distribution $p(Y, M, A)$, BF can be arbitrarily large. Furthermore, DV's lower bound on $p_{10}$ bound approaches 0 as BF goes to infinity. Thus, it is possible to construct a scenario (e.g., a joint distribution $p(Y, M, A, U)$) where DV's lower bound is arbitrarily close to 0, whereas the assumption-free bound is substantially larger than 0, i.e., a scenario where the assumption-free bound is larger than DV's bound.

A simple way to improve DV's lower bound is to replace it with the assumption-free bound, whenever the latter is larger. However, the bound can be improved even further, as stated in the following theorem.

**Theorem 2.**

$$p_{10} \geq l^{\dagger} = \max\{l_{DV}, v, s_0, s_1\},$$

*where*

$$\begin{aligned} v = {} & \max\{0, p(Y = 1, M = 0|A = 1) + p(M = 0|A = 0) - 1\} \\ & + \max\{0, p(Y = 1, M = 1|A = 1) + p(M = 1|A = 0) - 1\} \\ & + \max\{0, p(Y = 1|A = 1, M = 0)p(M = 0|A = 0)\mathrm{BF}^{-1} + p(M = 1|A = 1) - 1\} \\ & + \max\{0, p(Y = 1|A = 1, M = 1)p(M = 1|A = 0)\mathrm{BF}^{-1} + p(M = 0|A = 1) - 1\}, \\ s_m = {} & \max(0, q_m + r_m) \end{aligned}$$

*and*

$$\begin{aligned} r_m = {} & p(Y = 1|A = 1, M = 1 - m)p(M = 1 - m|A = 0)\mathrm{BF}^{-1} \\ & + \max\{0, p(M = 1 - m|A = 0) + p(M = 1 - m|A = 1) - 1\} \\ & + \max\{0, p(Y = 0, M = m|A = 1) + p(M = 1 - m|A = 0) - 1\} \\ & + \max\{0, p(Y = 1|A = 1, M = m)p(M = m|A = 0)\mathrm{BF}^{-1} + p(M = 1 - m|A = 1) - 1\}. \end{aligned}$$

As an example, suppose that $U$ is binary with $p(U = 1) = 0.54$. Suppose further that $p(M = 1|A, U)$ and $p(Y = 1|A, M, U)$ are as in Tables 1 and 2, respectively. For this example, the true values of $p_{10}$ and BF are 0.81 and 1.71, respectively. Suppose now that only the marginal (over $U$) distribution $p(Y, M, A)$ is given, but that we are able to correctly specify the true value of BF. From these, we obtain $l_{DV} = 0.44$, $l_{AF} = 0.59$, and $l^{\dagger} = s_1 = 0.65$.

**Table 1:** Example distribution of $p(M=1|A,U)$

| $A$ | $U$ | $p(M=1\|A,U)$ |
|---|---|---|
| 0 | 0 | 0.83 |
| 0 | 1 | 0.57 |
| 1 | 0 | 0.94 |
| 1 | 1 | 0.97 |

**Table 2:** Example distribution of $p(Y=1|A,M,U)$

| $A$ | $M$ | $U$ | $p(Y=1\|A,M,U)$ |
|---|---|---|---|
| 0 | 0 | 1 | 0.80 |
| 0 | 1 | 0 | 0.20 |
| 0 | 1 | 1 | 0.91 |
| 1 | 0 | 0 | 0.19 |
| 1 | 0 | 1 | 0.62 |
| 1 | 1 | 0 | 0.93 |
| 1 | 1 | 1 | 0.95 |

Thus, for this example, the assumption-free lower bound offers substantial improvement over DV's lower bound, and the lower bound in Theorem 2 offers a substantial improvement over the assumption-free bound.

## 6 An upper bound on $p_{10}$

To construct an upper bound on $p_{10}$, we need to define a new sensitivity parameter relating the exposure to the confounders. In analogy with $\mathrm{RR}_{AU}$, we define the sensitivity parameter $\widetilde{\mathrm{RR}}_{AU}=\max_m\widetilde{\mathrm{RR}}_{AU|m}$, where

$$\widetilde{\mathrm{RR}}_{AU|m}=\max_u\left\{\frac{p(U=u|A=0,M=m)}{p(U=u|A=1,M=m)}\right\},$$

and we define the corresponding bounding factor

$$\widetilde{\mathrm{BF}}=\frac{\widetilde{\mathrm{RR}}_{AU}\times\mathrm{RR}_{UY}}{\widetilde{\mathrm{RR}}_{AU}+\mathrm{RR}_{UY}-1}.$$

The following two theorems establish the feasible space for $\{\widetilde{\mathrm{RR}}_{AU},\mathrm{RR}_{UY}\}$, given the observed data distribution, and an upper bound on $p_{10}$, given $\widetilde{\mathrm{BF}}$.

**Theorem 3.** (i) $\{\widetilde{\mathrm{RR}}_{AU},\mathrm{RR}_{UY}\}$ *are restricted by their definitions to values* $\geq 1$. (ii) $\{\widetilde{\mathrm{RR}}_{AU},\mathrm{RR}_{UY}\}$ *are variation independent of each other and of the observed data distribution* $p(Y,M,A)$.

**Theorem 4.**

$$p_{10}\leq u^{\dagger}=\min\{u_{DV},\tilde{v},\tilde{s}_0,\tilde{s}_1,u_{AF}\},$$

*where*

$$u_{DV}=\min(1,p_{10}^{\mathrm{obs}}\widetilde{\mathrm{BF}}),$$
$$\tilde{s}_m=\min(1,q_m+\tilde{r}_m),$$

$$
\begin{aligned}
u_{AF} &= \min(1, \tilde{q}_0, \tilde{q}_1), \\
\tilde{q}_m &= p(Y=1|A=1) + p(M=m|A=0) + p(Y=0, M=m|A=1), \\
\tilde{v} &= \min[1, \min\{p(Y=1, M=0|A=1), p(M=0|A=0)\} \\
&\quad + \min\{p(Y=1, M=1|A=1), p(M=1|A=0)\} \\
&\quad + \min\{p(Y=1|A=1, M=0)p(M=0|A=0)\widetilde{\mathrm{BF}}, p(M=1|A=1)\} \\
&\quad + \min\{p(Y=1|A=1, M=1)p(M=1|A=0)\widetilde{\mathrm{BF}}, p(M=0|A=1)\}]
\end{aligned}
$$

*and*

$$
\begin{aligned}
\tilde{r}_m &= \min\{1, p(Y=1|A=1, M=1-m)p(M=1-m|A=0)\widetilde{\mathrm{BF}}\} \\
&\quad + \min\{p(M=1-m|A=0), p(M=1-m|A=1)\} \\
&\quad + \min\{p(Y=0, M=m|A=1), p(M=1-m|A=0)\} \\
&\quad + \min\{p(Y=1|A=1, M=m)p(M=m|A=0)\widetilde{\mathrm{BF}}, p(M=1-m|A=1)\}.
\end{aligned}
$$

In Theorem 4, we have used the notation $u_{DV}$ to indicate that this term is “analogous” to $l_{DV}$ in Theorem 2, in the sense that $u_{DV}$ is derived by using the same arguments as in the proof for the term $l_{DV}$ in Theorem 2. In the same sense, the terms $\tilde{v}$, $\tilde{s}_0$, and $\tilde{s}_1$ are analogous to the terms $v$, $s_0$, and $s_1$, respectively, in Theorem 2. The term $u_{AF}$ is an assumption-free upper bound $p_{10}$, which follows from results by Sjölander [5]. The reason why the corresponding term $l_{AF}$ does not appear in Theorem 2 is that this term is redundant, given $\{l_{DV}, s_0, s_1\}$, since $l_{DV} \geq 0$, $s_0 \geq q_0$, and $s_1 \geq q_1$; hence, $l_{AF} \leq \max\{l_{DV}, s_0, s_1\}$.

Continuing on the example from Section 5, we have that $u^{\dagger} = \tilde{s}_1 = 0.90$. Thus, for this example, we can tell that true value of $p_{10}$ is somewhere in the range (0.65, 0.90), provided that we are able to correctly specify the true value of BF.

# 7 Cornfield inequalities

DV used their lower bound $l_{DV}$ to derive Cornfield inequalities for NDE. These inequalities are lower bounds on the common value $\mathrm{RR}_{AU} = \mathrm{RR}_{UY} = \delta$ required to “explain away” an observed NDE with a positive magnitude. Here, we say that an observed NDE has positive magnitude if $p_{10}^{\mathrm{obs}} > p(Y=1|A=0)$, and that the observed NDE is explained away by $\delta$ if the lower bound for the true NDE includes the null value, e.g., 0 for the risk difference and 1 for the risk ratio. For the risk ratio, DV’s Cornfield inequality has a simple analytic form, given by $\delta \geq \mathrm{NDE}^{\mathrm{obs}} + \sqrt{\mathrm{NDE}^{\mathrm{obs}}(\mathrm{NDE}^{\mathrm{obs}} - 1)}$. This Cornfield inequality is valid in the sense that $\delta$ must indeed exceed this lower bound in order to explain away an observed NDE. However, since this Cornfield inequality is derived from the non-sharp bound $l_{DV}$, it is also not sharp and may thus give an unnecessarily pessimistic assessment of the study at hand.

To illustrate this, we again return to the numeric example from Section 5. For this example, $p_{10}^{\mathrm{obs}} = 0.76$, $p(Y=1|A=0) = 0.54$, $\mathrm{NDE}^{\mathrm{obs}} = p_{10}^{\mathrm{obs}}/p(Y=1|A=0) = 1.39$, and $\mathrm{NDE}^{\mathrm{obs}} + \sqrt{\mathrm{NDE}^{\mathrm{obs}}(\mathrm{NDE}^{\mathrm{obs}} - 1)} = 2.13$. Thus, in order to explain away the observed NDE, a common value of $\mathrm{RR}_{AU}$ and $\mathrm{RR}_{UY}$ must be at least 2.13. This is a rather modest value, which may give the impression that the observed NDE is quite sensitive to unmeasured confounding. However, the assumption-free lower bound $l_{AF}$ is equal to 0.59, which is larger than $p(Y=1|A=0)$. Hence, even if only the marginal (over $U$) distribution $p(Y, M, A)$ is known, and we are not able to specify BF, we can still in this example tell that there is a true non-null NDE. Put differently, no degree of unmeasured confounding could ever explain away the observed NDE in this example. This conclusion is much stronger than that obtained from DV’s Cornfield inequality.

DV’s Cornfield inequalities assume that the observed NDE has positive magnitude, and they are not applicable to NIE. Using the bounds $(l^{\dagger}, u^{\dagger})$ we can obtain Cornfield inequalities for both NDE and NIE, having either positive or negative magnitudes.

**Theorem 5.** *Let* $\delta$ *be a common value of* $\mathrm{RR}_{AU}$ *and* $\mathrm{RR}_{UY}$, *and let* $\tilde{\delta}$ *be a common value of* $\widetilde{\mathrm{RR}}_{AU}$ *and* $\mathrm{RR}_{UY}$. *Let* $\mathrm{BF}(a)$ *be the value of* $\mathrm{BF}$ *that solves the equation*

$$p(Y=1|A=a)=l^{\dagger}$$

*and let* $\widetilde{\mathrm{BF}}(a)$ *be the value of* $\widetilde{\mathrm{BF}}$ *that solves the equation*

$$p(Y=1|A=a)=u^{\dagger}.$$

*We then have the following four Cornfield inequalities.*

(i) *In order for* $\delta$ *to explain away an observed NDE, it has to obey the inequality*

$$\delta \geq \mathrm{BF}(0)+\sqrt{\mathrm{BF}(0)\{\mathrm{BF}(0)-1\}}.$$

(ii) *In order for* $\tilde{\delta}$ *to explain away an observed NDE, it has to obey the inequality*

$$\tilde{\delta} \geq \widetilde{\mathrm{BF}}(0)+\sqrt{\widetilde{\mathrm{BF}}(0)\{\widetilde{\mathrm{BF}}(0)-1\}}.$$

(iii) *In order for* $\delta$ *to explain away an observed NIE, it has to obey the inequality*

$$\delta \geq \mathrm{BF}(1)+\sqrt{\mathrm{BF}(1)\{\mathrm{BF}(1)-1\}}.$$

(iv) *In order for* $\tilde{\delta}$ *to explain away an observed NIE, it has to obey the inequality*

$$\tilde{\delta} \geq \widetilde{\mathrm{BF}}(1)+\sqrt{\widetilde{\mathrm{BF}}(1)\{\widetilde{\mathrm{BF}}(1)-1\}}.$$

If the observed NDE has positive magnitude and $l^{\dagger}=l_{DV}$, then the Cornfield inequality in Theorem 5 reduces to DV's Cornfield inequality, but is otherwise sharper. The equations defining $\mathrm{BF}(a)$ and $\widetilde{\mathrm{BF}}(a)$ in Theorem 5 have no solution if $p(Y=1|A=a)$ is outside the range $(l^{\dagger},u^{\dagger})$. This means that no degree of unmeasured confounding can explain away the observed NDE (for $a=0$) or NIE (for $a=1$), as in the numeric example above.

## 8 Simulation

We carried out a simulation to assess the numerical performance of the proposed bounds, and the potential for the relatively complex bounds $(l^{\dagger},u^{\dagger})$ to improve on the simpler bounds $(l_{DV},u_{DV})$. We considered a binary confounder $U$ and generated 1,000,000 distributions $p(Y,A,M,U)$ from the model

$$\begin{aligned}
p(U=1) &\sim \mathrm{Unif}(0,1),\\
p(M=1|A=a,U=u) &\sim \mathrm{Unif}(0,1) \quad \text{for } \{a,u\}\in\{0,1\},\\
p(Y=1|A=a,M=m,U=u) &\sim \mathrm{Unif}(0,1) \quad \text{for } \{a,m,u\}\in\{0,1\}.
\end{aligned}$$

For each distribution $p(Y,A,M,U)$, we computed the probabilities $p_{00}$, $p_{10}$, and $p_{11}$, the bounding factors BF and $\widetilde{\mathrm{BF}}$, and the bounds $l_{DV}, l^{\dagger}, u_{DV}$, and $u^{\dagger}$. When computing the bounds, we used the true bounding factors BF and $\widetilde{\mathrm{BF}}$ multiplied with a constant $k\geq 1$. The value $k=1$ represents the ideal scenario where the analyst uses the true bounding factors. This scenario is somewhat unrealistic, since the analyst would rarely know the true bounding factors in practice. Thus, we computed the bounds on both $k=1$ and $k=1.3$. The latter is intended to represent the possibly more realistic scenario where the analyst provides a conservative guess of the bounding factors, in order not to invalidate the bounds.

We first assessed the distance between the bounds and the true value of $p_{10}$, by constructing the standardized differences

$$\Delta_l=\left|\frac{(l_{DV}-p_{10})-(l^{\dagger}-p_{10})}{(l_{DV}-p_{10})}\right|$$

and

$$\Delta_u = \left| \frac{(u_{DV} - p_{10}) - (u^{\dagger} - p_{10})}{(u_{DV} - p_{10})} \right|.$$

Since $(l^{\dagger}, u^{\dagger})$ are at least as sharp as $(l_{DV}, u_{DV})$, we have that $0 \leq (\Delta_l, \Delta_u) \leq 1$. The plots in Figure 2 show the empirical (over the generated distributions) complementary cumulative distribution functions for $\Delta_l$ (left plot) and $\Delta_u$ (right plot). We observe that even though the bounds $(l^{\dagger}, u^{\dagger})$ are usually not sharper than $(l_{DV}, u_{DV})$, they sometimes are and the improvement can be substantial. For instance, with 4% and 6% probability, $l^{\dagger}$ is more than 20% closer to $p_{10}$ than $l_{DV}$, for $k = 1$ and $k = 1.3$, respectively. The corresponding probabilities for $u^{\dagger}$ are larger; 8% and 13% for $k = 1$ and $k = 1.3$, respectively.

Since $l_{DV}$ and $u_{DV}$ approach 0 and 1 as BF and $\widetilde{\text{BF}}$ approach infinity, one would intuitively expect the performance of $(l^{\dagger}, u^{\dagger})$, relative to $(l_{DV}, u_{DV})$, to increase with $(\text{BF}, \widetilde{\text{BF}})$. Figure 3 shows the same functions as in Figure 2, for $k = 1$, but only based on those distributions $p(Y, A, M, U)$ for which BF (for the lower bounds) and

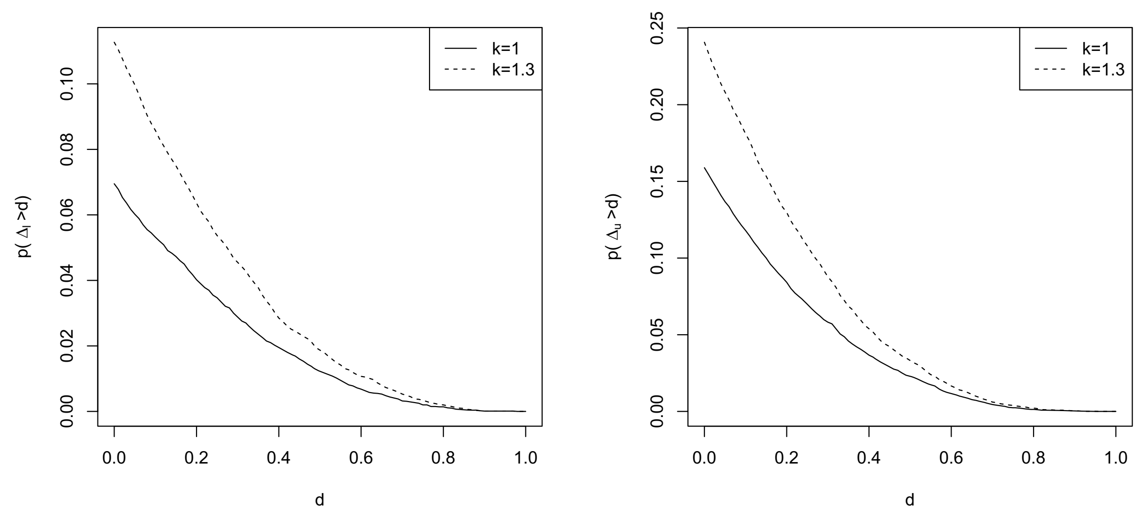


**Figure 2:** Complementary cumulative distribution functions for the standardized differences $\Delta_l$ (left plot) and $\Delta_u$ (right plot), for $k = 1$ (solid lines) and $k = 1.3$ (dashed lines).

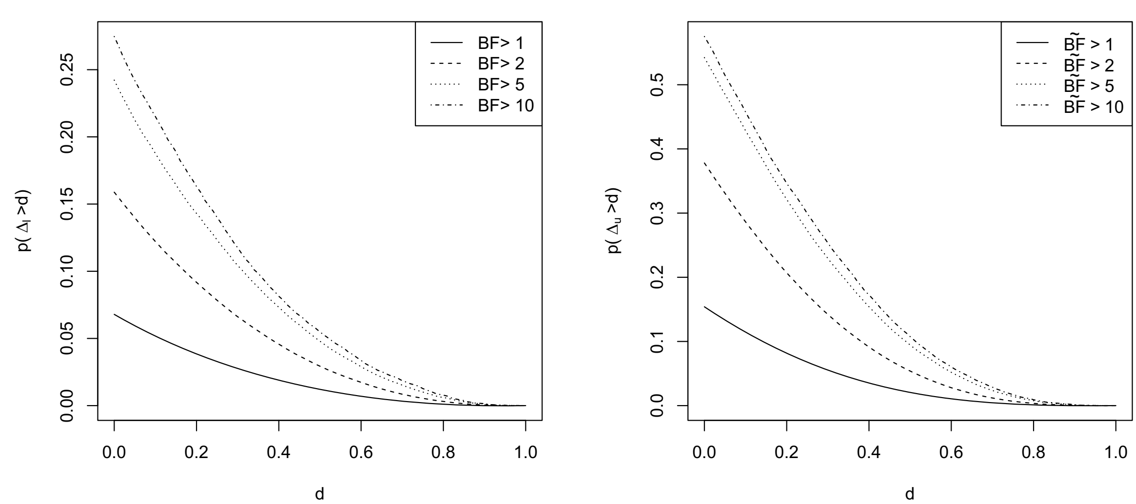


**Figure 3:** Complementary cumulative distribution functions for the standardized differences $\Delta_l$ (left plot) and $\Delta_u$ (right plot), stratified by thresholds for BF and $\widetilde{\text{BF}}$.

**Table 3:** Rejection probabilities for the lower bounds

| | $k = 1$ | $k = 1.3$ |
|---|---|---|
| $p(l_{DV} > p_{00})$ | 0.25 | 0.15 |
| $p(l_{DV} > p_{11})$ | 0.08 | 0.02 |
| $p(l^{\dagger} > p_{00})$ | 0.26 | 0.16 |
| $p(l^{\dagger} > p_{11})$ | 0.08 | 0.02 |

**Table 4:** Rejection probabilities for the upper bounds

| | $k = 1$ | $k = 1.3$ |
|---|---|---|
| $p(u_{DV} < p_{00})$ | 0.22 | 0.12 |
| $p(u_{DV} < p_{11})$ | 0.08 | 0.02 |
| $p(u^{\dagger} < p_{00})$ | 0.23 | 0.13 |
| $p(u^{\dagger} < p_{11})$ | 0.08 | 0.02 |

$\widetilde{\mathrm{BF}}$ (for the upper bounds) exceed a threshold equal to 1, 2, 5, and 10. As expected, we observe that the complementary cumulative distribution functions increase as the bounding factors increase. This indicates that the stronger the unmeasured confounding, the higher the probability that the bounds $(l^{\dagger}, u^{\dagger})$ improves substantially on the bounds $(l_{DV}, u_{DV})$.

We next assessed the ability of the bounds to reject the null hypotheses that NDE = 0 and NIE = 0. The null hypothesis NDE = 0 is rejected by a lower (upper) bound on $p_{10}$ if the bound is larger (smaller) than $p_{00}$. Similarly, the null hypothesis NIE = 0 is rejected by a lower (upper) bound on $p_{10}$ if the bound is larger (smaller) than $p_{11}$. Tables 3 and 4 show the empirical probabilities of these rejections, for $(l_{DV}, l^{\dagger})$ and $(u_{DV}, u^{\dagger})$, respectively. We observe that the rejection probabilities are very similar for $l_{DV}$ and $l^{\dagger}$ and also very similar for $u_{DV}$ and $u^{\dagger}$, for both $k = 1$ and $k = 1.3$, and for both NDE and NIE. This indicates that, even though the bounds $(l^{\dagger}, u^{\dagger})$ are potentially much sharper than the bounds $(l_{DV}, u_{DV})$, their additional impact is modest on the power to reject the null hypotheses.

## 9 Application

We illustrate the proposed bounds by analyzing data from the Lipid Research Clinics Coronary Primary Prevention (LRC) trial [6]. These data were previously used by Cai et al. [7] and Sjölander [5] to illustrate the assumption-free bounds on controlled direct effects and NDEs, respectively. In the LRC trial, 1888 subjects were randomized to cholestyramine treatment and 1918 subjects to placebo. During a follow-up period of 1 year, each coronary heart disease (CHD) event was recorded. At the end of the follow-up, cholesterol levels were recorded for each subject. Both Cai et al. [7] and Sjölander [5] dichotomized cholesterol levels as “low” (<280 mg/dl) or “high” (≥280 mg/dl). The data are displayed in Table 5.

We coded all variables so that the “1” represents the “desirable” level, i.e., treatment ($A = 1$), low cholesterol ($M = 1$), and CHD absent ($Y = 1$). We then computed each element of the min/max expressions for the bounds on $p_{10}$ given in Theorems 2 and 4. We finally constructed bounds on NDE and NIE as risk ratios, for each of these elements separately. For instance, the lower bound on NDE and the upper bound on NIE obtained from the element $l_{DV}$ are given by $l_{DV}/p(Y = 1|A = 0)$ and $p(Y = 1|A = 1)/l_{DV}$, respectively.

Figure 4 shows the lower (left column) and upper (right column) bounds on NDE (top row) and NIE (bottom row), as functions of BF (for NDE) and $\widetilde{\mathrm{BF}}$ (for NIE). In the absence of confounding ($\mathrm{BF} = \widetilde{\mathrm{BF}} = 1$),

**Table 5:** Data from the LRC trial

| | Placebo | | Treatment | |
|---|---|---|---|---|
| | **High chol** | **Low chol** | **High chol** | **Low chol** |
| CHD present | 82 | 86 | 33 | 97 |
| CHD absent | 669 | 1081 | 332 | 1426 |

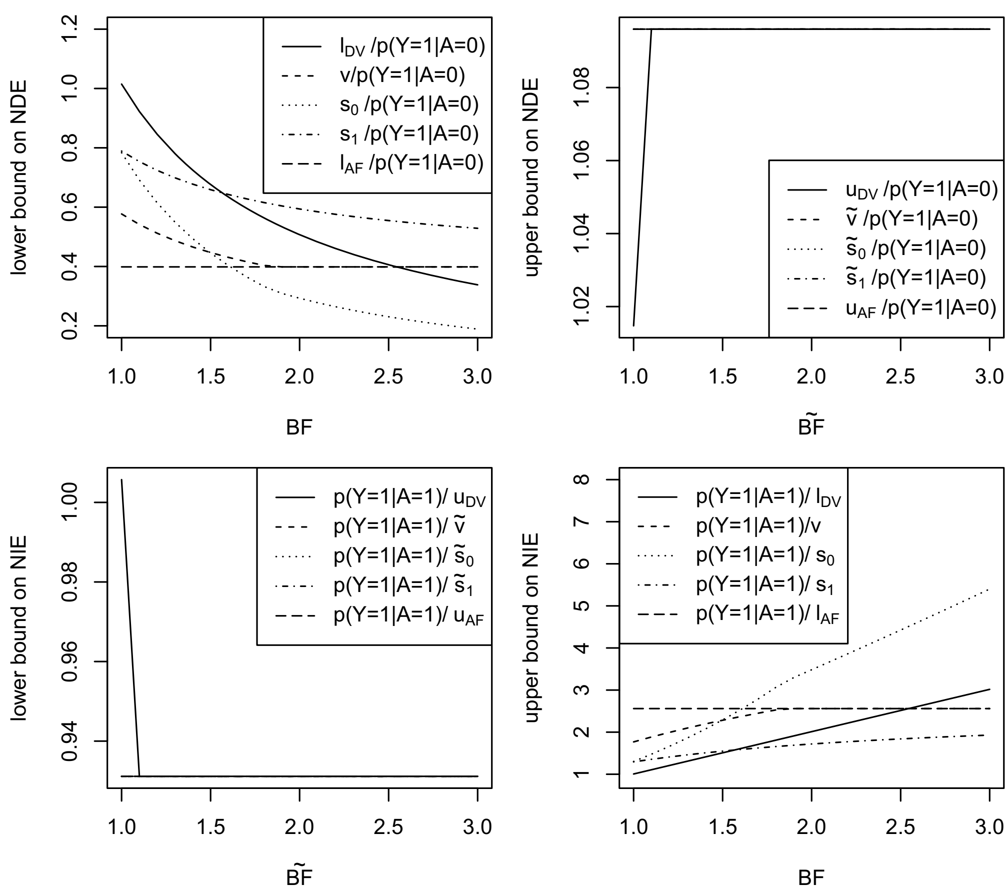


**Figure 4:** Bounds on NDE and NIE, for the LRC data.

the lower and upper bounds collapse at the “observed effects” NDE = 1.02 and NIE = 1.01. These effects are explained away by a tiny amount of confounding; evaluating the Cornfield inequality in Theorem 5 gives that $\delta = 1.01$ for both NDE and NIE. Further, the amount of assumed confounding has quite dramatic effects on some of the bounds. At BF = 1, the bound $l_{DV}/p(Y = 1|A = 0)$ is largest, and thus most informative, among all lower bounds on NDE. However, already at BF = 1.6, $l_{DV}/p(Y = 1|A = 0)$ becomes smaller than $s_1/p(Y = 1|A = 0)$, and at BF = 2.6, $l_{DV}/p(Y = 1|A = 0)$ becomes smaller than the assumption-free bound $l_{AF}/p(Y = 1|A = 0)$. We observe a similar pattern for the upper bounds on NIE, where the bound $p(Y = 1|A = 1)/l_{DV}$ is smallest at BF = 1, but becomes larger than $p(Y = 1|A = 1)/s_1$ at BF = 1.6, and larger than the assumption-free bound $p(Y = 1|A = 1)/l_{AF}$ at BF = 2.6. The upper bounds on NDE are less diverse; apart from $u_{DV}/p(Y = 1|A = 0)$, all bounds are identical (= $1/p(Y = 1|A = 0) = 1.1$) for all values of $\widetilde{\text{BF}}$, and thus not be distinguished from each other in the top-right plot of Figure 4. The bound $u_{DV}/p(Y = 1|A = 0)$ is slightly smaller (= 1.02) at $\widetilde{\text{BF}} = 1$, but becomes equal to all other bounds at $\widetilde{\text{BF}} = 1.1$. We observe a similar pattern for the lower bounds on NIE; apart from $p(Y = 1|A = 1)/u_{DV}$, all bounds are identical (=$p(Y = 1|A = 1)/1 = 0.93$) for all values of $\widetilde{\mathrm{BF}}$. The bound $p(Y = 1|A = 1)/u_{DV}$ is slightly larger (=1.01) at $\widetilde{\mathrm{BF}} = 1$, but becomes equal to all other bounds at $\widetilde{\mathrm{BF}} = 1.1$.

## 10 Discussion

The proposed bounds are applicable to any direct and indirect effects that can be written as contrasts of probabilities for binary outcomes. A useful extension would be to derive analogous bounds on contrasts of survival probabilities, with time-to-event outcomes.

We have shown that the proposed bounds are at least as sharp as those derived by DV. However, we have not been able to prove (or disprove) that our bounds are sharp in an absolute sense, i.e., that all values inside the bounds are logically possible, given the assumptions. We recognize this as an interesting topic for future research.

**Funding information:** Arvid Sjölander gratefully acknowledges funding from The Swedish Research Council, grant number 2020-01188.

**Author contributions:** All authors have contributed intellectually to the project, and to the drafting of the manuscript.

**Conflict of interest:** Authors state no conflict of interest.

## References

[1] Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992;3(2):143–55.
[2] Pearl J. Direct and indirect effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc.; 2001. p. 411–20.
[3] Ding P, Vanderweele TJ. Sharp sensitivity bounds for mediation under unmeasured mediator–outcome confounding. Biometrika. 2016;103(2):483–90.
[4] Richardson TS, Robins JM. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper. 2013;128(30):1–148.
[5] Sjölander A. Bounds on natural direct effects in the presence of confounded intermediate variables. Stat Med. 2009;28(4):558–71.
[6] Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992;110(2):167–78.
[7] Cai M, Kuroki Z, Pearl J, Tian J. Bounds on direct effects in the presence of confounded intermediate variables. Biometrics. 2008;64(3):695–701.

# Appendix

## A SWIG

Assumed SWIG is displayed in Figure A1.

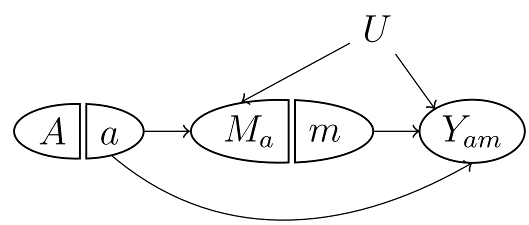


**Figure A1:** Assumed SWIG.

## B Proof of Theorem 1

(i) It is trivially true that $\mathrm{RR}_{UY} \geq 1$, since $\mathrm{RR}_{UY|m} \geq 1$ by definition. Define $g(u,m) = p(U = u|A = 1, M = m)/p(U = u|A = 0, M = m)$, so that $\mathrm{RR}_{AU|m} = \max_u\{g(u,m)\}$. To see that $\mathrm{RR}_{AU|m} \geq 1$, note that $E\{g(U,m)|A = 0, M = m\} = \sum_u g(u,m)p(U = u|A = 0, M = m) = \sum_u p(U = u|A = 1, M = m) = 1$, which implies that $g(u,m) \geq 1$ for at least one value $u$, which in turn implies that $\max_u\{g(u,m)\} \geq 1$, which further implies that $\mathrm{RR}_{AU} \geq 1$ as well.
(ii) To show that $\{\mathrm{RR}_{AU}, \mathrm{RR}_{UY}, p(Y,M,A)\}$ are variation independent, we show that it is possible to construct a distribution $p(Y,M,A,U)$ that marginalizes to any given $\{\mathrm{RR}^*_{AU}, \mathrm{RR}^*_{UY}, p^*(Y,M,A)\}$, such that $U \perp A$. We construct the distribution $p(Y,M,A,U)$ in the following steps.

(1) Set $p(M,A) = p^*(M,A)$, and define $p_a = p(M = 1|A = a)$. Without loss of generality, we assume that

$$p_1 \geq p_0; \tag{A1}$$

this assumption can always be made to hold by swapping the coding of $A$.

(2) Let $U$ be binary, with

$$\begin{aligned} p(U = 1|A = a, M = 0) &= k(p_1 - p_0/\mathrm{RR}^*_{AU}) \quad \text{for } a \in \{0,1\},\\ p(U = 1|A = 0, M = 1) &= k(p_1 - p_0)/\mathrm{RR}^*_{AU},\\ p(U = 1|A = 1, M = 1) &= k(p_1 - p_0), \end{aligned}$$

where

$$k = \min\left\{\frac{1}{p_1 - p_0/\mathrm{RR}^*_{AU}}, \frac{1 - p^*(Y = 1|A = 1, M = 1)}{(p_1 - p_0)(1 - 1/\mathrm{RR}^*_{AU})}\right\}.$$

From part (i) of Theorem 1 and (A1) it follows that $0 \leq p(U = 1|A = a, M = m) \leq 1$ for $(a,m) \in \{0,1\}$. We further have that

$$\begin{aligned} p(U = 1|A = a) &= p(U = 1|A = a, M = 0)(1 - p_a) + p(U = 1|A = a, M = 1)p_a\\ &= k(p_1 + p_0p_1/\mathrm{RR}^*_{AU} - p_0p_1 - p_0/\mathrm{RR}^*_{AU}) \quad \text{for } a \in \{0,1\}; \end{aligned}$$

hence, $U \perp A$. We also have that

$$\frac{p(U = 1|A = 1, M = 0)}{p(U = 1|A = 0, M = 0)} = \frac{p(U = 0|A = 1, M = 0)}{p(U = 0|A = 0, M = 0)} = 1$$

and

$$\frac{p(U=1|A=1,M=1)}{p(U=1|A=0,M=1)}=\mathrm{RR}^*_{AU}\geq\frac{p(U=0|A=1,M=1)}{p(U=0|A=0,M=1)}=\frac{1-k(p_1-p_0)}{1-k(p_1-p_0)/\mathrm{RR}^*_{AU}},$$

so that $\mathrm{RR}_{AU}=\mathrm{RR}^*_{AU}$.

(3) Set

$$\begin{aligned}p(Y=1|A=a,M=m,U=u)&=p^*(Y=1|A=a,M=m)\quad\text{if } a\neq 1 \text{ or } m\neq 1,\\ p(Y=1|A=1,M=1,U=0)&=\frac{p^*(Y=1|A=1,M=1)}{1-k(p_1-p_0)(1-1/\mathrm{RR}^*_{UY})},\\ p(Y=1|A=1,M=1,U=1)&=p(Y=1|A=1,M=1,U=0)/\mathrm{RR}^*_{UY}.\end{aligned}$$

We trivially have that $0\leq p(Y=1|A=a,M=m,U=u)\leq 1$ and $p(Y=1|A=a,M=m)=p^*(Y=1|A=a,M=m)$ if $a\neq 1$ or $m\neq 1$. From part (i) of Theorem 1 and (A1), it follows that $0\leq p(Y=1|A=a,M=m,U=u)\leq 1$ for $a=m=1$. We also have that

$$\begin{aligned}p(Y=1|A=1,M=1)&=p(Y=1|A=1,M=1,U=0)p(U=0|A=1,M=1)\\&\quad+p(Y=1|A=1,M=1,U=0)p(U=1|A=1,M=1)\\&=p^*(Y=1|A=1,M=1)\left[\frac{1-(p_1-p_0)}{1-k(p_1-p_0)(1-1/\mathrm{RR}^*_{UY})}+\frac{(p_1-p_0)/\mathrm{RR}^*_{UY}}{1-k(p_1-p_0)(1-1/\mathrm{RR}^*_{UY})}\right]\\&=p^*(Y=1|A=1,M=1).\end{aligned}$$

We finally have that

$$\frac{p(Y=1|A=1,M=m,U=u)}{p(Y=1|A=1,M=m,U=u')}=1$$

if $a\neq 1$ or $m\neq 1$, and

$$\frac{p(Y=1|A=1,M=m,U=0)}{p(Y=1|A=1,M=m,U=1)}=\mathrm{RR}^*_{UY}\geq\frac{p(Y=1|A=1,M=m,U=1)}{p(Y=1|A=1,M=m,U=0)}=1/\mathrm{RR}^*_{UY},$$

so that $\mathrm{RR}_{UY}=\mathrm{RR}^*_{UY}$.

# C Proof of Theorem 2

**Lemma A1.**

$$p(Y_{1m}=y,M_0=m)\geq p(Y=y|A=1,M=m)p(M=m|A=0)\mathrm{BF}^{-1}.$$

**Proof.**

$$\begin{aligned}p(Y_{1m}=y,M_0=m)&=\sum_u p(Y_{1m}=y|M_0=m,U=u)p(U=u|M_0=m)p(M_0=m)\\&=\sum_u p(Y=y|A=1,M=m,U=u)p(U=u|A=0,M=m)p(M=m|A=0)\\&\geq\sum_u p(Y=y|A=1,M=m,U=u)p(U=u|A=1,M=m)\left(\frac{\mathrm{RR}_{AU|m}\times\mathrm{RR}_{UY|m}}{\mathrm{RR}_{AU|m}+\mathrm{RR}_{UY|m}-1}\right)^{-1}\\&\qquad\times p(M=m|A=0)\\&=p(Y=y|A=1,M=m)\left(\frac{\mathrm{RR}_{AU|m}\times\mathrm{RR}_{UY|m}}{\mathrm{RR}_{AU|m}+\mathrm{RR}_{UY|m}-1}\right)^{-1}p(M=m|A=0)\\&\geq p(Y=y|A=1,M=m)p(M=m|A=0)\mathrm{BF}^{-1}.\end{aligned}$$

The second equality follows from (1)–(3) in the main text. The first inequality follows from Lemma A.3 in the supplementary material in the study of Ding and VanderWeele [3], by setting $r(x) = p(Y = y|A = 1, M = m, U = u)$, $F_0(x) = p(U = u|A = 0, M = m)$, and $F_1(x) = p(U = u|A = 1, M = m)$. □

**Lemma A2.** *For any two random variables G and H, we have that*

$$p(G = g, H = h) \geq \max\{0, p(G = g) + p(H = h) - 1\}.$$

**Proof.**

$$\begin{aligned} p(G = g) &= p(G = g, H = h) + p(G = g, H \neq h) \\ p(H = h) &= p(G = g, H = h) + p(G \neq g, H = h) \\ 1 &= p(G = g, H = h) + p(G = g, H \neq h) + p(G \neq g, H = h) + p(G \neq g, H \neq h) \end{aligned}$$

so

$$p(G = g) + p(H = h) - 1 = p(G = g, H = h) - p(G \neq g, H \neq h) \leq p(G = g, H = h).$$

We now prove that $p_{10} \geq q_m + r_m$ for $m = 0$. The proof for $m = 1$ is analogous. First, we have that

$$\begin{aligned} q_0 &= p(Y = 1, M = 0|A = 1) - p(M = 1|A = 0) \\ &= p(Y_1 = 1, M_1 = 0|A = 1) - p(M_0 = 1|A = 0) \\ &= p(Y_1 = 1, M_1 = 0) - p(M_0 = 1) \\ &= p(Y_{10} = 1, M_1 = 0) - p(M_0 = 1), \end{aligned}$$

where the second equality follows from (1) in the main text and the third from (3) in the main text. We also have that

$$p_{10} = p(Y_{1M_0} = 1, M_0 = 1) + p(Y_{1M_0} = 1, M_0 = 0) = p(Y_{11} = 1, M_0 = 1) + p(Y_{10} = 1, M_0 = 0). \tag{A2}$$

Combining gives that

$$\begin{aligned} p_{10} &= q_0 + p(Y_{11} = 1, M_0 = 1) + p(Y_{10} = 1, M_0 = 0) + p(M_0 = 1) - p(Y_{10} = 1, M_1 = 0) \\ &= q_0 + p(Y_{11} = 1, M_0 = 1) \\ &\quad + \{\underbrace{p(M_0 = 0, M_1 = 0, Y_{10} = 1, Y_{11} = 0)}_{I} + \underbrace{p(M_0 = 0, M_1 = 0, Y_{10} = 1, Y_{11} = 1)}_{II} \\ &\quad + p(M_0 = 0, M_1 = 1, Y_{10} = 1, Y_{11} = 0) + p(M_0 = 0, M_1 = 1, Y_{10} = 1, Y_{11} = 1)\} \\ &\quad + \{p(M_0 = 1, M_1 = 0, Y_{10} = 0, Y_{11} = 0) + p(M_0 = 1, M_1 = 0, Y_{10} = 0, Y_{11} = 1) \\ &\quad + \underbrace{p(M_0 = 1, M_1 = 0, Y_{10} = 1, Y_{11} = 0)}_{III} + \underbrace{p(M_0 = 1, M_1 = 0, Y_{10} = 1, Y_{11} = 1)}_{IV} \\ &\quad + p(M_0 = 1, M_1 = 1, Y_{10} = 0, Y_{11} = 0) + p(M_0 = 1, M_1 = 1, Y_{10} = 0, Y_{11} = 1) \\ &\quad + p(M_0 = 1, M_1 = 1, Y_{10} = 1, Y_{11} = 0) + p(M_0 = 1, M_1 = 1, Y_{10} = 1, Y_{11} = 1)\} \\ &\quad - \{\underbrace{p(M_0 = 0, M_1 = 0, Y_{10} = 1, Y_{11} = 0)}_{I} + \underbrace{p(M_0 = 0, M_1 = 0, Y_{10} = 1, Y_{11} = 1)}_{II} \\ &\quad + \underbrace{p(M_0 = 1, M_1 = 0, Y_{10} = 1, Y_{11} = 0)}_{III} + \underbrace{p(M_0 = 1, M_1 = 0, Y_{10} = 1, Y_{11} = 1)}_{IV}\} \\ &= q_0 + p(Y_{11} = 1, M_0 = 1) + p(M_0 = 0, M_1 = 1, Y_{10} = 1) + p(M_0 = 1, M_1 = 0, Y_{10} = 0) + p(M_0 = 1, M_1 = 1), \end{aligned} \tag{A3}$$

where we have used roman numerals to indicate terms that cancel with each other. It follows directly from Lemma A1 that

$$p(Y_{11} = 1, M_0 = 1) \geq p(Y = 1|A = 1, M = 1)p(M = 1|A = 0)\mathrm{BF}^{-1}.$$

It follows from Lemma A1 that

$$p(Y_{10} = 1, M_0 = 0) \geq p(Y = 1|A = 1, M = 0)p(M = 0|A = 0)\mathrm{BF}^{-1}$$

and from (1) and (3) in the main text that $p(M_1 = 1) = p(M = 1|A = 1)$. Setting $G = \{Y_{10}, M_0\}$ and $H = M_1$ in Lemma A2 thus gives that

$$p(M_0 = 0, M_1 = 1, Y_{10} = 1) \geq \max\{0, p(Y = 1|A = 1, M = 0)p(M = 0|A = 0)\text{BF}^{-1} + p(M = 1|A = 1) - 1\}.$$

By a similar argument, and using that $p(Y_{10} = 0, M_1 = 0) = p(Y = 0, M = 0|A = 1)$ under (1) and (3) in the main text, we have that

$$p(M_0 = 1, M_1 = 0, Y_{10} = 0) \geq \max\{0, p(Y = 0, M = 0|A = 1) + p(M = 1|A = 0) - 1\}.$$

Finally, using that $p(M_0 = 1) = p(M = 1|A = 0)$ and $p(M_1 = 1) = p(M = 1|A = 1)$ under (1) and (3) in the main text, we have that

$$p(M_0 = 1, M_1 = 1) \geq \max\{0, p(M = 1|A = 1) + p(M = 1|A = 0) - 1\}.$$

Combining gives that $p_{10} \geq q_0 + r_0$.

We now prove that $p_{10} \geq v$. From (A2) we can further decompose $p_{10}$ as

$$\begin{aligned} p_{10} = {} & p(Y_{11} = 1, M_0 = 1, M_1 = 0) + p(Y_{11} = 1, M_0 = 1, M_1 = 1) \\ & + p(Y_{10} = 1, M_0 = 0, M_1 = 0) + p(Y_{10} = 1, M_0 = 0, M_1 = 1). \end{aligned} \tag{A4}$$

Applying Lemmas A1 and A2 to minimize each of the terms on the right-hand side of (A4), similar to how we did for the terms on the right-hand side of (A3), gives that $p_{10} \geq v$. □

# D Proof of Theorem 3

Theorem 3 follows by symmetry from Theorem 1.

# E Proof of Theorem 4

**Lemma A3.**

$$p(Y_{1m} = y, M_0 = m) \leq p(Y = y|A = 1, M = m)p(M = m|A = 0)\widetilde{\text{BF}}.$$

**Proof.**

$$\begin{aligned} p(Y_{1m} = y, M_0 = m) &= \sum_u p(Y_{1m} = y|M_0 = m, U = u)p(U = u|M_0 = m)p(M_0 = m) \\ &= \sum_u p(Y = y|A = 1, M = m, U = u)p(U = u|A = 0, M = m)p(M = m|A = 0) \\ &\leq \sum_u p(Y = y|A = 1, M = m, U = u)p(U = u|A = 1, M = m)\frac{\widetilde{\text{RR}}_{AU|m} \times \text{RR}_{UY|m}}{\widetilde{\text{RR}}_{AU|m} + \text{RR}_{UY|m} - 1}p(M = m|A \\ &\quad = 0) \\ &= p(Y = y|A = 1, M = m)\frac{\widetilde{\text{RR}}_{AU|m} \times \text{RR}_{UY|m}}{\widetilde{\text{RR}}_{AU|m} + \text{RR}_{UY|m} - 1}p(M = m|A = 0) \\ &\leq p(Y = y|A = 1, M = m)p(M = m|A = 0)\widetilde{\text{BF}}. \end{aligned}$$

The second equality follows from (1)–(3) in the main text. The first inequality follows from Lemma A.3 in the supplementary material in Ding and VanderWeele [3], by setting $r(x) = p(Y = y|A = 1, M = m, U = u)$, $F_1(x) = p(U = u|A = 0, M = m)$ and $F_0(x) = p(U = u|A = 1, M = m)$. □

**Lemma A4.** *For any two random variables G and H, we have that*

$$p(G = g, H = h) \leq \min\{p(G = g), p(H = h)\}.$$

The proof is immediate. Applying Lemma A3 to maximize each of the terms on the right-hand side of (A2) gives that $p_{10} \leq u_{DV}$. Applying Lemmas A3 and A4 to maximize each of the terms on the right-hand sides of (A3) and (A4) gives that $p_{10} \leq q_0 + \tilde{r}_0$ and $p_{10} \leq \tilde{v}$, respectively.

## F Proof of Theorem 5

We prove part (i) of the theorem. The other parts of the theorem can be proved analogously.

Setting $\mathrm{RR}_{AU} = \mathrm{RR}_{UY} = \delta$,

$$\mathrm{BF} = \frac{\delta \times \delta}{\delta + \delta - 1}$$

and solving for $\delta$ gives that

$$\mathrm{BF} = \mathrm{BF}(0) \Rightarrow \delta = \mathrm{BF}(0) + \sqrt{\mathrm{BF}(0)\{\mathrm{BF}(0) - 1\}}. \tag{A5}$$

We thus have that

$$\delta < \mathrm{BF}(0) + \sqrt{\mathrm{BF}(0)\{\mathrm{BF}(0) - 1\}} \Rightarrow \mathrm{BF} < \mathrm{BF}(0) \Rightarrow l^{\dagger} > p(Y = 1|A = 0),$$

where the first implication follows from (A5) and the fact that BF is a monotonically increasing function of $\mathrm{RR}_{AU} = \mathrm{RR}_{UY} = \delta$, and the second implication follows from the definition of BF(0) and the fact that $l^{\dagger}$ is a mononically decreasing function of BF. Since an observed NDE is not explained away if $l^{\dagger} > p(Y = 1|A = 0)$, the result follows.

## G R code

```
rm(list=ls())
    set.seed(1)
    #—Utility functions—
    logit <- function(x)  log(x/(1-x))
    boundsfun <- function(pAMY, BF, BFtilde){

pM1.A0 <- pAMY[1]
pM1.A1 <- pAMY[2]
pY1.A0M0 <- pAMY[3]
pY1.A0M1 <- pAMY[4]
pY1.A1M0 <- pAMY[5]
pY1.A1M1 <- pAMY[6]

pY0M0.A0 <- (1-pY1.A0M0)*(1-pM1.A0)
pY0M1.A0 <- (1-pY1.A0M1)*pM1.A0
pY1M0.A0 <- pY1.A0M0*(1-pM1.A0)
pY1M1.A0 <- pY1.A0M1*pM1.A0

pY0M0.A1 <- (1-pY1.A1M0)*(1-pM1.A1)
pY0M1.A1 <- (1-pY1.A1M1)*pM1.A1
```

```
pY1M0.A1 <- pY1.A1M0*(1-pM1.A1)
pY1M1.A1 <- pY1.A1M1*pM1.A1

pY1.A0 <- pY1.A0M0*(1-pM1.A0)+pY1.A0M1*pM1.A0
pY1.A1 <- pY1.A1M0*(1-pM1.A1)+pY1.A1M1*pM1.A1
p10obs <- pY1.A1M0*(1-pM1.A0)+pY1.A1M1*pM1.A0

q0 <- pY1M0.A1-pM1.A0
q1 <- pY1M1.A1-(1-pM1.A0)

r0 <- pY1.A1M1*pM1.A0/BF+
  max(0,pM1.A1+pM1.A0-1)+
  max(0, pM1.A0+pY0M0.A1-1)+
  max(0, pY1.A1M0*(1-pM1.A0)/BF+pM1.A1-1)

r1 <- pY1.A1M0*(1-pM1.A0)/BF+
  max(0,(1-pM1.A1)+(1-pM1.A0)-1)+
  max(0, (1-pM1.A0)+pY0M1.A1-1)+
  max(0, pY1.A1M1*pM1.A0/BF+(1-pM1.A1)-1)

r0tilde <- min(1, pY1.A1M1*pM1.A0*BFtilde)+
  min(pM1.A1,pM1.A0)+
  min(pM1.A0,pY0M0.A1)+
  min(pY1.A1M0*(1-pM1.A0)*BFtilde,pM1.A1)

r1tilde <- min(1, pY1.A1M0*(1-pM1.A0)*BFtilde)+
  min((1-pM1.A1),(1-pM1.A0))+
  min((1-pM1.A0),pY0M1.A1)+
  min(pY1.A1M1*pM1.A0*BFtilde,(1-pM1.A1))

v <- max(0, pY1.A1M0*(1-pM1.A1)+(1-pM1.A0)-1)+
  max(0, pY1.A1M1*pM1.A1+pM1.A0-1)+
  max(0, pY1.A1M0*(1-pM1.A0)/BF+pM1.A1-1)+
  max(0, pY1.A1M1*pM1.A0/BF+(1-pM1.A1)-1)

vtilde <- min(1, min(pY1.A1M0*(1-pM1.A1), 1-pM1.A0)+
  min(pY1.A1M1*pM1.A1, pM1.A0)+
  min(pY1.A1M0*(1-pM1.A0)*BFtilde, pM1.A1)+
  min(pY1.A1M1*pM1.A0*BFtilde, 1-pM1.A1))

s0 <- max(0, q0+r0)
s1 <- max(0, q1+r1)
s0tilde <- min(1, q0+r0tilde)
s1tilde <- min(1, q1+r1tilde)

lDV <- p10obs/BF
lAF <- max(0, q0, q1)
l <- c(lDV, v, s0, s1, lAF)

uDV <- min(1, p10obs*BFtilde)
uAF <- min(1, pY1.A1+pM1.A0+pY0M1.A1, pY1.A1+(1-pM1.A0)+pY0M0.A1)
```

```
u <- c(uDV, vtilde, s0tilde, s1tilde, uAF)

b <- matrix(c(l, u), nrow=2, ncol=5, byrow=TRUE)
colnames(b) <- c("DV", "v", "s0", "s1", "AF")
b

}

pfun <- function(pAMY){.

pM1.A0 <- pAMY[1]
pM1.A1 <- pAMY[2]
pY1.A0M0 <- pAMY[3]
pY1.A0M1 <- pAMY[4]
pY1.A1M0 <- pAMY[5]
pY1.A1M1 <- pAMY[6]

pY0M0.A0 <- (1-pY1.A0M0)*(1-pM1.A0)
pY0M1.A0 <- (1-pY1.A0M1)*pM1.A0
pY1M0.A0 <- pY1.A0M0*(1-pM1.A0)
pY1M1.A0 <- pY1.A0M1*pM1.A0

pY0M0.A1 <- (1-pY1.A1M0)*(1-pM1.A1)
pY0M1.A1 <- (1-pY1.A1M1)*pM1.A1
pY1M0.A1 <- pY1.A1M0*(1-pM1.A1)
pY1M1.A1 <- pY1.A1M1*pM1.A1

pY1.A0 <- pY1.A0M0*(1-pM1.A0)+pY1.A0M1*pM1.A0
pY1.A1 <- pY1.A1M0*(1-pM1.A1)+pY1.A1M1*pM1.A1

p10obs <- pY1.A1M0*(1-pM1.A0)+pY1.A1M1*pM1.A0

p <- c(pY1.A0, pY1.A1, p10obs)
names(p) <- c("pY1.A0", "pY1.A1", "p10obs")
p

}

truthfun <- function(pUAMY){.

pU <- pUAMY$pU
pM1.A0U <- pUAMY$pM1.A0U
pM1.A1U <- pUAMY$pM1.A1U
pY1.A1M0U <- pUAMY$pY1.A1M0U
pY1.A1M1U <- pUAMY$pY1.A1M1U

pU.A0M0 <- pU*(1-pM1.A0U)/sum(pU*(1-pM1.A0U))
pU.A0M1 <- pU*pM1.A0U/sum(pU*pM1.A0U)
pU.A1M0 <- pU*(1-pM1.A1U)/sum(pU*(1-pM1.A1U))
pU.A1M1 <- pU*pM1.A1U/sum(pU*pM1.A1U)
```

```
RRAU <- max(max(pU.A1M0/pU.A0M0), max(pU.A1M1/pU.A0M1))
RRUY <- max(max(pY1.A1M0U)/min(pY1.A1M0U), max(pY1.A1M1U)/min(pY1.A1M1U))
BF <- RRAU*RRUY/(RRAU+RRUY-1)

RRAUtilde <- max(max(pU.A0M0/pU.A1M0), max(pU.A0M1/pU.A1M1))
BFtilde <- RRAUtilde*RRUY/(RRAUtilde+RRUY-1)

p10 <- sum((pY1.A1M0U*(1-pM1.A0U)+pY1.A1M1U*pM1.A0U)*pU)

truth <- c(RRAU, RRAUtilde, RRUY, BF, BFtilde, p10)
names(truth) <- c("RRAU", "RRAUtilde", "RRUY", "BF", "BFtilde", "p10")
truth

}

margU <- function(pUAMY){.

pU <- pUAMY$pU
pM1.A0U <- pUAMY$pM1.A0U
pM1.A1U <- pUAMY$pM1.A1U
pY1.A0M0U <- pUAMY$pY1.A0M0U
pY1.A0M1U <- pUAMY$pY1.A0M1U
pY1.A1M0U <- pUAMY$pY1.A1M0U
pY1.A1M1U <- pUAMY$pY1.A1M1U

pU.A0M0 <- pU*(1-pM1.A0U)/sum(pU*(1-pM1.A0U))
pU.A0M1 <- pU*pM1.A0U/sum(pU*pM1.A0U)
pU.A1M0 <- pU*(1-pM1.A1U)/sum(pU*(1-pM1.A1U))
pU.A1M1 <- pU*pM1.A1U/sum(pU*pM1.A1U)

pY1.A0M0 <- sum(pY1.A0M0U*pU.A0M0)
pY1.A0M1 <- sum(pY1.A0M1U*pU.A0M1)
pY1.A1M0 <- sum(pY1.A1M0U*pU.A1M0)
pY1.A1M1 <- sum(pY1.A1M1U*pU.A1M1)

pM1.A0 <- sum(pM1.A0U*pU)
pM1.A1 <- sum(pM1.A1U*pU)

c(pM1.A0, pM1.A1, pY1.A0M0, pY1.A0M1, pY1.A1M0, pY1.A1M1)

}

#---Numeric example in Tables 1 and 2---

pU <- c(0.46, 0.54)
pM1.A0U <- c(0.83, 0.57)
pM1.A1U <- c( 0.94, 0.97)
pY1.A0M0U <- c(0.02, 0.80)
pY1.A0M1U <- c(0.20, 0.91)
pY1.A1M0U <- c(0.19, 0.62)
pY1.A1M1U <- c(0.93, 0.95)
```

```
pUAMY <- list(pU=pU, pM1.A0U=pM1.A0U, pM1.A1U=pM1.A1U, pY1.A0M0U=pY1.A0M0U,
  pY1.A0M1U=pY1.A0M1U, pY1.A1M0U=pY1.A1M0U, pY1.A1M1U=pY1.A1M1U)
truth <- truthfun(pUAMY)
BF <- truth[4]
BFtilde <- truth[5]
pAMY <- margU(pUAMY)

bounds <- boundsfun(pAMY, BF, BFtilde)

print(round(truth, 2))
print(round(bounds, 2))
p <- pfun(pAMY)
print(round(p, 2))
p1.0 <- p[1]
p10obs <- p[3]
NDEobs <- p10obs/p1.0
E <- NDEobs+sqrt(NDEobs*(NDEobs-1))
NDE <- c(NDEobs, E)
names(NDE) <- c("NDEobs", "E")
print(round(NDE, 2))

#---Simulation in Section 8---

N <- 1000000
nU <- 2
lall <- uall <- matrix(nrow=N, ncol=2*5)
pall <- matrix(nrow=N, ncol=2)
p10all <- vector(length=N)
truthall <- matrix(nrow=N, ncol=6)


for(i in 1:N){.

pU <- -log(runif(nU))
pU <- pU/sum(pU)
pM1.A0U <- runif(nU)
pM1.A1U <- runif(nU)
pY1.A0M0U <- runif(nU)
pY1.A0M1U <- runif(nU)
pY1.A1M0U <- runif(nU)
pY1.A1M1U <- runif(nU)

pUAMY <- list(pU=pU, pM1.A0U=pM1.A0U, pM1.A1U=pM1.A1U, pY1.A0M0U=pY1.A0M0U,
  pY1.A0M1U=pY1.A0M1U, pY1.A1M0U=pY1.A1M0U, pY1.A1M1U=pY1.A1M1U)
truth <- truthfun(pUAMY)
pAMY <- margU(pUAMY)
pall[i, ] <- pfun(pAMY)[1:2]
truthall[i, ] <- truth

k <- 1
BF <- truth[4]*k
```

```
BFtilde <- truth[5]*k
bounds <- boundsfun(pAMY, BF, BFtilde)
lall[i, 1:5] <- bounds[1, ]
uall[i, 1:5] <- bounds[2, ]

k <- 1.3
BF <- truth[4]*k
BFtilde <- truth[5]*k
bounds <- boundsfun(pAMY, BF, BFtilde)
lall[i, 6:10] <- bounds[1, ]
uall[i, 6:10] <- bounds[2, ]

}

lmat <- umat <- matrix(nrow=N, ncol=8)
for(i in 1:N){.

lmat[i, ] <- c(lall[i, 1]>pall[i, ], max(lall[i, 1:5])>pall[i, ],
  lall[i, 6]>pall[i, ], max(lall[i, 6:10])>pall[i, ])
umat[i, ] <- c(uall[i, 1]<pall[i, ], min(uall[i, 1:5])<pall[i, ],
  uall[i, 6]<pall[i, ], min(uall[i, 6:10])<pall[i, ])

}

lg <- colMeans(lmat)
lg <- cbind(lg[1:4], lg[5:8])
colnames(lg) <- c("k=1", "k=1.3")rownames(lg) <- c("p(l_{DV}>p_{00})", "p(l_{DV}>p_{11})",
  "p(l^{dagger}>p_{00})", "p(l^{dagger}>p_{11})")
print(round(lg, 2))
ug <- colMeans(umat)
ug <- cbind(ug[1:4], ug[5:8])
colnames(ug) <- c("k=1", "k=1.3")
rownames(ug) <- c("p(u_{DV}<p_{00})", "p(u_{DV}<p_{11})",
"p(u^{dagger}<p_{00})", "p(u^{dagger}<p_{11})")
print(round(ug, 2))

p10all <- truthall[, 6]
dl1 <- ((p10all-lall[, 1])-(p10all-apply(X=lall[, 1:5], MARGIN=1, FUN=max)))/
  (p10all-lall[, 1])
dl2 <- ((p10all-lall[, 6])-(p10all-apply(X=lall[, 6:10], MARGIN=1, FUN=max)))/
  (p10all-lall[, 6])
du1 <- ((uall[, 1]-p10all)-(apply(X=uall[, 1:5], MARGIN=1, FUN=min)-p10all))/
  (uall[, 1]-p10all)
du2 <- ((uall[, 6]-p10all)-(apply(X=uall[, 6:10], MARGIN=1, FUN=min)-p10all))/
  (uall[, 6]-p10all)

Fl1 <- ecdf(dl1)
Fl2 <- ecdf(dl2)
Fu1 <- ecdf(du1)
Fu2 <- ecdf(du2)
```

```
d <- seq(0, 1, 0.01)

windows(width=18, height=8)
op <- par(mfrow=c(1,2), mar=c(5, 5, 1, 2))
matplot(d, 1-cbind(Fl1(d), Fl2(d)), type="l", col=1,
  ylab=expression("p("~Delta[l]~">d)"))
legend(x="topright", col=1, lty=1:2, legend=c("k=1", "k=1.3"))
matplot(d, 1-cbind(Fu1(d), Fu2(d)), type="l", col=1,
  ylab=expression("p("~Delta[u]~">d)"))
legend(x="topright", col=1, lty=1:2, legend=c("k=1", "k=1.3"))
par(op)

out <- c(1-Fl1(0.2), 1-Fl2(0.2), 1-Fu1(0.2), 1-Fu2(0.2))
names(out) <- c("l^{dagger}, k=1", "l^{dagger}, k=1.3",
  "u^{dagger}, k=1", "u^{dagger}, k=1.3")
print(round(out, 2))

bf <- c(1, 2, 5, 10)
lbf <- length(bf)
BFall <- truthall[, 4]
BFtildeall <- truthall[, 5]
l <- u <- matrix(nrow=length(d), ncol=lbf)
for(i in 1:lbf){.
  Fl <- ecdf(dl1[BFall>bf[i]])
  l[, i] <- 1-Fl(d)
  Fu <- ecdf(du1[BFtildeall>bf[i]])
  u[, i] <- 1-Fu(d)
}

windows(width=18, height=8)
op <- par(mfrow=c(1,2), mar=c(5, 5, 1, 2))
matplot(d, l, type="l", col=1,
  ylab=expression("p("~Delta[l]~">d)"))
legend(x="topright", col=1, lty=1:lbf, legend=paste("BF>", bf))
matplot(d, u, type="l", col=1,
  ylab=expression("p("~Delta[u]~">d)"))
legend(x="topright", col=1, lty=1:lbf, legend=c(expression(tilde(BF)~">"~1),
  expression(tilde(BF)~">"~2), expression(tilde(BF)~">"~5),
  expression(tilde(BF)~">"~10)))
par(op)

#--- Real data illustration ----

#nAMY
n000 <- 1426
n001 <- 97
n010 <- 332
n011 <- 33
n100 <- 1081
n101 <- 86
```

```
n110 <- 669
n111 <- 82

#switch coding of A, M and Y
pM1.A0 <- (n101+n100)/(n101+n100+n111+n110)
pM1.A1 <- (n001+n000)/(n001+n000+n011+n010)
pY1.A0M0 <- n110/(n110+n111)
pY1.A0M1 <- n100/(n100+n101)
pY1.A1M0 <- n010/(n010+n011)
pY1.A1M1 <- n000/(n000+n001)

pY1.A0 <- pY1.A0M0*(1-pM1.A0)+pY1.A0M1*pM1.A0
pY1.A1 <- pY1.A1M0*(1-pM1.A1)+pY1.A1M1*pM1.A1

pAMY <- c(pM1.A0, pM1.A1, pY1.A0M0, pY1.A0M1, pY1.A1M0, pY1.A1M1)

BF <- BFtilde <- 1

b <- boundsfun(pAMY, BF, BFtilde)
print("NDE")
print(b/pY1.A0)
print("NIE")
print(pY1.A1/b)

#if(0){

bf <- seq(1, 3, 0.1)
l <- u <- matrix(nrow=length(bf), ncol=5)
rownames(l) <- rownames(u) <- bf
colnames(l) <- colnames(u) <- c("DV", "v", "s0", "s1", "AF")
for(i in 1:length(bf)){.
  BF <- BFtilde <- bf[i]
  b <- boundsfun(pAMY, BF, BFtilde)
  l[i, ] <- b[1, ]
  u[i, ] <- b[2, ]
}

lNDE <- l/pY1.A0
uNDE <- u/pY1.A0
lNIE <- pY1.A1/u
uNIE <- pY1.A1/l

windows(width=10, height=8)
op <- par(mfrow=c(2,2), mar=c(5, 4.5, 0.1, 0.5))
matplot(bf, lNDE, type="l", col=1, lty=1:5, ylab="lower bound on NDE",
  xlab="BF", ylim=c(0.2, 1.2))
legend(x="topright", col=1, lty=1:5,
  legend=c(expression(l[DV] "/p(Y=1 A=0)"),
  "v/p(Y=1 A=0)",
  expression(s[0] "/p(Y=1 A=0)"),
```

```
  expression(s[1] "/p(Y=1 A=0)"),
  expression(l[AF] "/p(Y=1 A=0)")))
matplot(bf, uNDE, type="l", col=1, lty=1:5, ylab="upper bound on NDE",
  xlab=expression(tilde(BF)))
legend(x="bottomright", col=1, lty=1:5,
  legend=c(expression(u[DV] "/p(Y=1 A=0)"),
  expression(tilde(v) "/p(Y=1 A=0)"),
  expression(tilde(s)[0] "/p(Y=1 A=0)"),
  expression(tilde(s)[1] "/p(Y=1 A=0)"),
  expression(u[AF] "/p(Y=1 A=0)")))
matplot(bf, lNIE, type="l", col=1, lty=1:5, ylab="lower bound on NIE",
  xlab=expression(tilde(BF)))
legend(x="topright", col=1, lty=1:5,
  legend=c(expression("p(Y=1 A=1)/" u[DV]),
  expression("p(Y=1 A=1)/" tilde(v)),
  expression("p(Y=1 A=1)/" tilde(s)[0]),
  expression("p(Y=1 A=1)/" tilde(s)[1]),
  expression("p(Y=1 A=1)/" u[AF])))
matplot(bf, uNIE, type="l", col=1, lty=1:5, ylab="upper bound on NIE",
  xlab="BF", ylim=c(0.95, 8))
legend(x="topleft", col=1, lty=1:5,
  legend=c(expression("p(Y=1 A=1)/" l[DV]),
  "p(Y=1 A=1)/v",
  expression("p(Y=1 A=1)/" s[0]),
  expression("p(Y=1 A=1)/" s[1]),
  expression("p(Y=1 A=1)/" l[AF])))
par(op)
```