Research Article

Jose M. Peña*, Sourabh Balgi, Arvid Sjölander, and Erin E. Gabriel

On the bias of adjusting for a non-differentially mismeasured discrete confounder

<https://doi.org/10.1515/jci-2021-0033>

received July 05, 2021; accepted August 23, 2021

Abstract: Biological and epidemiological phenomena are often measured with error or imperfectly captured in data. When the true state of this imperfect measure is a confounder of an outcome exposure relationship of interest, it was previously widely believed that adjustment for the mismeasured observed variables provides a less biased estimate of the true average causal effect than not adjusting. However, this is not always the case and depends on both the nature of the measurement and confounding. We describe two sets of conditions under which adjusting for a non-differentially mismeasured proxy comes closer to the unidentifiable true average causal effect than the unadjusted or crude estimate. The first set of conditions apply when the exposure is discrete or continuous and the confounder is ordinal, and the expectation of the outcome is monotonic in the confounder for both treatment levels contrasted. The second set of conditions apply when the exposure and the confounder are categorical (nominal). In all settings, the mismeasurement must be non-differential, as differential mismeasurement, particularly an unknown pattern, can cause unpredictable results.

Keywords: bias, causal inference, confounding

MSC 2020: 62D20

1 Introduction

In observational studies, it is often of interest to estimate the average causal effect of an exposure on an outcome when this relationship is confounded. The confounder may be non-differentially mismeasured, resulting in the observation of a proxy. Non-differentially mismeasurement means that the proxy is conditionally independent of exposure and the outcome given the true confounder.

For binary true confounder and proxy, Greenland [1] argued that adjusting for the proxy produces a partially adjusted measure of the average causal effect of the exposure on the outcome that is between the crude (i.e., unadjusted) and the true (i.e., adjusted for the true confounder) measures. Thus, the partially adjusted measure comes closer to the unidentifiable true measure than the crude one. Ogburn and VanderWeele [2] showed that this result does not always hold. For binary exposure, true confounder, and proxy, they also showed that the result holds if the conditional expectation of the outcome is monotonic in the true confounder, i.e., it is either non-decreasing or non-increasing in the confounder for both exposure

* **Corresponding author: Jose M. Peña**, Department of Computer and Information Science, Linköping University, Sweden, e-mail: jose.m.pena@liu.se

Sourabh Balgi: Department of Computer and Information Science, Linköping University, Sweden, e-mail: sourabh.balgi@liu.se

Arvid Sjölander: Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Sweden, e-mail: arvid.sjolander@ki.se

Erin E. Gabriel: Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Sweden, e-mail: erin.gabriel@ki.se

levels contrasted. Unfortunately, the condition cannot be tested explicitly because the true confounder is unobserved. Ogburn and VanderWeele [3] extended these results to binary exposure and ordinal true confounder and proxy, still under the monotonicity assumption. For binary exposure, true confounder, and proxy, Peña [4] showed that monotonicity in the true confounder holds if monotonicity in the proxy exists, which can be tested empirically.

Although monotonicity may seem like a strong condition, Ogburn and VanderWeele [2] argued that it is likely to hold in most applications in epidemiology. However, it is not difficult to come up with examples that are not monotonic. Ogburn and VanderWeele [2] gave an example of the causal effect of type 2 diabetes on hypertension when confounded by the use of thiazide, a drug to treat hypertension. The drug is known to lower the risk of hypertension in the general population, but it is also known to increase blood glucose levels increasing the risk of hypertension for patients with diabetes. Thus, the drug may have a non-monotone effect by causing harm in patients with diabetes, while being protective for all other patients. Monotonicity is not a necessary condition for improved partial adjustment. Peña [4] characterized non-monotonic cases for binary exposure, true confounder, and proxy, where the partially adjusted average causal effect is still between the crude and the true ones.

We provide new results for both monotonic and non-monotonic settings. For the monotonic settings, we show that the proof of the result in ref. [3] also applies when the exposure is non-binary under additional assumptions.

2 Preliminaries

Let A be the exposure of interest, and Y the outcome, which are confounded by the unobserved C . Let D be the non-differentially mismeasured proxy of C , where $(A, Y) \perp D|C$. The causal diagram depicting this setting is in the left most panel of Figure 1. We use upper-case letters to denote random variables, and the same letters in lower-case to denote fixed (e.g., by observation or by intervention) values. For simplicity, we use $p()$ to denote both probability distributions and density functions.

We can now more clearly define monotonicity. $E[Y|A, C]$ is monotone in a binary confounder C for the exposure levels a and $\bar{a}$ if

- $E[Y|A = a', C = 1] \geq E[Y|A = a', C = 0]$ for $a' \in (a, \bar{a})$, or
- $E[Y|A = a', C = 1] \leq E[Y|A = a', C = 0]$ for $a' \in (a, \bar{a})$.

This is easily extended to ordinal C . Given an exposure level a and $\bar{a}$, and for all level of C , such that $i < j$ if:

- $E[Y|A = a', C = i] \geq E[Y|A = a', C = j]$ for $a' \in (a, \bar{a})$, or
- $E[Y|A = a', C = i] \leq E[Y|A = a', C = j]$ for $a' \in (a, \bar{a})$,

then $E[Y|A, C]$ is monotonic in C , and likewise for $p(a|C)$.

For ordinal C and D , Ogburn and VanderWeele [3] introduced the concept of tapered misclassification probabilities: If $p(D = i|C = j) \leq p(D = i|C = k)$ and $p(D = j|C = i) \leq p(D = k|C = i)$ for all $j < k \leq i$, and $p(D = i|C = j) \geq p(D = i|C = k)$ and $p(D = j|C = i) \geq p(D = k|C = i)$ for all $i \leq j < k$, then $p(D|C)$ is said to be a tapered distribution. Intuitively, this means that, for any level i of C or D , the probability of correct

Figure 1: Causal graphs where C and U are unobserved.

classification into level i is equal or greater than the probability of misclassification into any other level and, moreover, the latter is non-increasing in each direction away from i .

Similar to a tapered misclassification, for categorical C and D , $p(D|C)$ is a preferential distribution if there is a permutation π_D of the levels of D such that

$$p(d|c) = \begin{cases} p_D & \text{if } d = \pi_D(c) \\ q_D & \text{otherwise} \end{cases}$$

with $p_D > 1/K$ and $q_D = (1 - p_D)/(K - 1)$. Note that p_D and q_D do not depend on either c or d . If $p(d|c) = p_D$, then we say that the level c prefers the level d or, equivalently, that the level d is preferred by the level c . We denote such levels of C and D by c_d and d_c , respectively. That $p(A|C)$ is preferential is defined likewise, with π_A , p_A , q_A , c_a , and a_c defined accordingly. Preferential distributions resemble tapered distributions, but they differ in that they apply to categorical variables, and the misclassification probabilities are all equal. For example, say that we are interested in the average causal effect of neighborhood (A) on math achievements (Y) when confounded by parental education (C). Although we do not know the parents' education, we do know their occupation (D). If individuals whose parents have education c prefer to live in the neighborhood a_c and have no preferences for the other neighborhoods and parents with education c prefer the occupation d_c and have no preferences for the other occupations, then $p(A|C)$ and $p(D|C)$ are preferential distributions.

When C is sufficient for confounding control (as in the left most panel of Figure 1), the true causal average treatment effect, or risk difference, between two exposure levels a and $\bar{a}$ is identifiable if C is observed. Let Y_a denote the potential outcome for a given subject, had that subject received exposure level $A = a$. The true causal risk difference is then defined as:

$$RD_{\text{true}} = E[Y_a] - E[Y_{\bar{a}}] = \sum_c E[Y|a, c]p(c) - \sum_c E[Y|\bar{a}, c]p(c).$$

Thus, $E[Y_a]$ is a potential outcome measure that is obtained by standardizing $E[Y|a, C]$ to $p(C)$. When C is unobserved, RD_{true} is not identifiable. However, it can be approximated by the unadjusted or crude risk difference:

$$RD_{\text{crude}} = E[Y|a] - E[Y|\bar{a}]$$

and by the partially adjusted or observed risk difference:

$$RD_{\text{obs}} = S_a - S_{\bar{a}} = \sum_d E[Y|a, d]p(d) - \sum_d E[Y|\bar{a}, d]p(d).$$

Note that just like $E[Y_a]$, S_a and $E[Y|a]$ are obtained by standardizing $E[Y|a, D]$ to $p(D)$ and $p(D|a)$, respectively.

We may be interested in standardizing other C - and D -conditional association measures between A and Y than the expectation. For instance, we may be interested in the probability that the potential outcome is greater than a given value. That is, we may be interested in the cumulative distribution function (CDF) $F_{Y_a}(y) = p(Y_a \leq y)$ at a fixed value y , which is obtained by standardizing $F_{Y|a,C}(y) = p(Y \leq y|a, C)$ to $p(C)$. When C is unobserved, it can be approximated by standardizing $F_{Y|a,D}(y)$ to $p(D)$ or $p(D|a)$.

3 Monotonic settings

We wish to know which of the two approximations comes closer to the unidentifiable true causal quantity. This section answers the question under some assumptions. The quality of RD_{crude} and RD_{obs} as approximations of RD_{true} is related to the quality of $E[Y|a]$ and S_a as approximations of $E[Y_a]$. Thus, we study the latter question first. Specifically, the following theorem shows that S_a comes closer to $E[Y_a]$ than $E[Y|a]$. Moreover, the theorem allows us to conclude whether S_a is an upper or a lower bound of $E[Y_a]$.

Theorem 1. Let A and Y be discrete or continuous random variables. Let C and D be ordinal random variables with levels $1, 2, \dots, K$. Let $p(A, D)$ be strictly positive. Let $A = a$ be any exposure level such that $p(a|C)$ and $E[Y|a, C]$ are both monotonic in C . Let $p(D|C)$ be a tapered distribution. Then, $S_a \geq E[Y_a]$ if and only if $E[Y|a] \geq S_a$.

The previous theorem carries over to standardized conditional CDFs if $F_{Y|a,C}(y)$ is monotonic in C , because the proof does not depend on the association measure being standardized. That $F_{Y|a,C}(y)$ is monotonic in C means that either $F_{Y|a,C=i}(y) \leq F_{Y|a,C=j}(y)$ for all $i < j$ or $F_{Y|a,C=i}(y) \geq F_{Y|a,C=j}(y)$ for all $i < j$. When the former inequality applies to any fixed value y , it is known as first-order stochastic dominance of $F_{Y|a,C=i}(y)$ over $F_{Y|a,C=j}(y)$, and it is then a sufficient (but not necessary) condition for $E[Y|a, C = i] \geq E[Y|a, C = j]$ [5, Section 3.2].

The following theorem answers our original question by showing that RD_{obs} is a better approximation to RD_{true} than RD_{crude} . The theorem extends a similar result in ref. [3, Theorem 1] for binary A .

Theorem 2. Let A and Y be discrete or continuous random variables. Let C and D be ordinal random variables with levels $1, 2, \dots, K$. Let $p(A, D)$ be strictly positive. Let $A = a$ and $A = \bar{a}$ be two exposure levels such that one of $p(a|C)$ and $p(\bar{a}|C)$ is non-decreasing and the other is non-increasing in C . Let $E[Y|a, C]$ and $E[Y|\bar{a}, C]$ be both monotonic in C . Let $p(D|C)$ be a tapered distribution. Then, RD_{obs} lies between RD_{true} and RD_{crude} .

Note that the theorem allows us to conclude whether RD_{obs} is an upper or a lower bound of the unidentifiable RD_{true} . That is, $RD_{\text{obs}} \geq RD_{\text{true}}$ if and only if $RD_{\text{crude}} \geq RD_{\text{obs}}$, and the latter inequality is empirically testable. Note the assumption about $p(\bar{a}|C)$ in the theorem. This is unnecessary in the result in ref. [3] because, when A is binary, the assumption about $p(a|C)$ implies the assumption about $p(\bar{a}|C)$. Although the proofs of both results are similar, we include the proof of our result in the Appendix because the proof in ref. [3] is rather brief and lacks important details, such as the requirement that $p(A, D)$ is strictly positive.

Like Theorems 1 and 2 carries over to differences between standardized conditional CDFs if $F_{Y|a,C}(y)$ is monotonic in C , which allows us to sort the true, crude and observed CDF differences between two exposure levels. Finally, Theorem 2 holds in fact for any effect measure that can be written as $h(g(E[Y_a]) - g(E[Y_{\bar{a}}]))$, where $g()$ and $h()$ are monotonic functions. Therefore, the theorem holds for, for instance, the causal risk ratio and causal odds ratio.

4 Non-monotonic settings

In this section, we do not assume that $E[Y|A, C]$ is monotonic in C for the two exposure levels contrasted. Instead, we identify alternative assumptions that still allow us to conclude that RD_{obs} lies between RD_{true} and RD_{crude} and, thus, that RD_{obs} comes closer to the unidentifiable RD_{true} than RD_{crude} . Specifically, consider again the causal graph to the left in Figure 1.

Theorem 3. Let Y be a discrete or continuous random variable. Let A, C , and D be categorical random variables with levels $1, 2, \dots, K$. Let $p(C)$ be a discrete uniform distribution, and let $p(A|C)$ and $p(D|C)$ be preferential distributions. Then, for any two exposure levels a and $\bar{a}$, RD_{obs} lies between RD_{true} and RD_{crude} .

Note that the theorem also allows us to conclude whether RD_{obs} is an upper or a lower bound of the unidentifiable RD_{true} . That is, $RD_{\text{obs}} \geq RD_{\text{true}}$ if and only if $RD_{\text{crude}} \geq RD_{\text{obs}}$, and the latter inequality is empirically testable.

The following corollary shows that, as expected, the bias of RD_{obs} decreases with increasing quality of D as a proxy of C (except when $p_A = 1$ because, in that case, Y is conditionally independent of D given A and, thus, RD_{obs} reduces to RD_{crude}).

Corollary 4. *Under the conditions of Theorem 3, if $p_A < 1$, then the difference between RD_{obs} and RD_{true} decreases as p_D increases.*

Counterexamples to Theorem 3 can be constructed when the conditions are violated, via the code provided at: <https://www.dropbox.com/s/pv2y84f4w38be8i/necessityBinary.R?dl=0> for binary variables, and <https://www.dropbox.com/s/5psjqijfqj6v5d9/necessityTernary.R?dl=0> for ternary variables. Note that one must rely on knowledge external to the observed data to verify the conditions in the theorem, since C is unobserved. The following corollary partially alleviates this, since the uniformity of $p(D)$ is empirically testable.

Corollary 5. *Let Y be a discrete or continuous random variable. Let A , C , and D be categorical random variables with levels $1, 2, \dots, K$. Then, $p(C)$ is a discrete uniform distribution and $p(D|C)$ is a preferential distribution if and only if $p(D)$ is a discrete uniform distribution and $p(C|D)$ is a preferential distribution. Likewise replacing D with A .*

As in the monotonic case, an analogous result to that in Theorem 3 holds for $E[Y|a]$, S_a , and $E[Y_a]$.

Corollary 6. *Under the conditions of Theorem 3, $S_a \geq E[Y_a]$ if and only if $E[Y|a] \geq S_a$.*

As in the monotonic case, the results above carry over to standardizing other association measures than the conditional expectation, e.g., the conditional CDF. The only difference being that the previous requirement of $F_{Y|A,C}(y)$ being monotonic in C is now replaced by the requirements of Theorem 3.

Although we have restricted consideration to the causal graph to the left in Figure 1, these results apply in a greater space of diagrams and settings. For instance, consider the causal graph in the center of the Figure 1. Let A , B , and Z be categorical random variables with equal cardinality. Let U be unobserved. The average causal effect of the exposure level a on Y can be computed as

$$E[Y_a] = \sum_z p(z|a) E[Y_z] = \sum_z p(z|a) \sum_{\bar{a}} E[Y|z, \bar{a}] p(\bar{a}) \quad (1)$$

by the front-door criterion [6, Section 3.3.2]. Now, suppose that A is not observed but B is observed. Then, $E[Y_z]$ is not computable but it can be approximated by $E[Y|z]$ or $S_z = \sum_b E[Y|z, b] p(b)$. Let S_a^{crude} and S_a^{obs} denote the results of plugging these two approximations in equation (1). Under some conditions, we can conclude that S_a^{obs} comes closer to $E[Y_a]$ than S_a^{crude} , and whether S_a^{obs} is an upper or a lower bound of $E[Y_a]$. Specifically, suppose that $p(Z|A)$ is known, e.g., from substantive knowledge or from a previous study. Suppose also that $p(A)$ is a discrete uniform distribution, and $p(Z|A)$ and $p(B|A)$ are preferential distributions. Now, note that if $E[Y|z] \geq S_z$ for all z , then $S_z \geq E[Y_z]$ for all z by Corollary 6 and, thus, $S_a^{\text{crude}} \geq S_a^{\text{obs}} \geq E[Y_a]$. Likewise, if $E[Y|z] \leq S_z$ for all z , then $S_a^{\text{crude}} \leq S_a^{\text{obs}} \leq E[Y_a]$.

4.1 Binary exposure and confounder

In this section, we consider binary A , C , and D and replace the previous assumptions that $p(A|C)$ and $p(D|C)$ are preferential distributions. The new assumptions yield weaker but still useful results. For instance, the following two theorems do not determine the order between RD_{crude} and RD_{obs} , but they may determine whether RD_{true} is positive or negative. More specifically, the first result below allows us to conclude that $RD_{\text{true}} < 0$ whenever $RD_{\text{crude}} < 0$ or $RD_{\text{obs}} < 0$.

Theorem 7. *Let Y be a discrete or continuous random variable. Let A , C , and D be binary random variables. Let $p(c) = 0.5$, $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$, and $p(\bar{d}|\bar{c}) \geq p(d|c) \geq 0.5$. If $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$, then $RD_{\text{crude}} \geq RD_{\text{true}}$ and $RD_{\text{obs}} \geq RD_{\text{true}}$. If $E[Y|a, c] - E[Y|a, \bar{c}] \leq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \leq 0$, then $RD_{\text{crude}} \leq RD_{\text{true}}$ and $RD_{\text{obs}} \leq RD_{\text{true}}$.*

Theorem 8. Let Y be a discrete or continuous random variable. Let A, C , and D be binary random variables. Let $p(c) = 0.5$, $p(a|c) \leq p(\bar{a}|\bar{c}) \leq 0.5$, and $p(d|c) \leq p(\bar{d}|\bar{c}) \leq 0.5$. If $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$, then $RD_{\text{crude}} \leq RD_{\text{true}}$ and $RD_{\text{obs}} \leq RD_{\text{true}}$. If $E[Y|a, c] - E[Y|a, \bar{c}] \leq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \leq 0$, then $RD_{\text{crude}} \geq RD_{\text{true}}$ and $RD_{\text{obs}} \geq RD_{\text{true}}$.

Note that $p(A|C)$ is not necessarily a preferential distribution in the theorems above, because may be $p(a|c) > p(\bar{a}|\bar{c}) > 0.5$, i.e., the probabilities of the preferred levels are not equal. Likewise for $p(D|C)$.

For example, let A, D , and Y represent three diseases, and C a gene variant that affects the three of them. Moreover, suppose that suffering A affects the risk of suffering Y . The first result in Theorem 7 holds if (i) half of the population carry the gene variant C , i.e., $p(c) = 0.5$, (ii) not carrying C protects against A and D more than carrying it predisposes to suffer the diseases, i.e., $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$ and $p(\bar{d}|\bar{c}) \geq p(d|c) \geq 0.5$, and (iii) carrying C increases the average severity of Y for the individuals suffering A more than it decreases the average severity for the rest, i.e., $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$.

The previous two theorems can be strengthened for RD_{crude} by dropping the assumption that $p(c) = 0.5$. However, analogous results do not hold for RD_{obs} .

Theorem 9. Let Y be a discrete or continuous random variable. Let A and C be binary random variables. Let $p(c) \leq 0.5$ and $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$. If $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$, then $RD_{\text{crude}} \geq RD_{\text{true}}$. If $E[Y|a, c] - E[Y|a, \bar{c}] \leq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \leq 0$, then $RD_{\text{crude}} \leq RD_{\text{true}}$.

Theorem 10. Let Y be a discrete or continuous random variable. Let A and C be binary random variables. Let $p(c) \geq 0.5$ and $p(a|c) \leq p(\bar{a}|\bar{c}) \leq 0.5$. If $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$, then $RD_{\text{crude}} \leq RD_{\text{true}}$. If $E[Y|a, c] - E[Y|a, \bar{c}] \leq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \leq 0$, then $RD_{\text{crude}} \geq RD_{\text{true}}$.

4.2 Sensitivity to assumption violations

In this section, we study how sensitive the results of Theorem 3 are to small departures from the assumptions. Specifically, we study how the ordering of RD_{true} , RD_{crude} , and RD_{obs} changes when $p(C)$, $p(A|C)$, and $p(D|C)$ are replaced by distributions that slightly deviate from the assumptions in the theorem. We use the superscript $*$ to denote the deviating probability distributions and their corresponding causal risk differences. The following definition formalizes what we mean by probabilities and risk differences deviating from the ones in Theorem 3. We say that a quantity r is an ε -approximation to a quantity s if

$$-\varepsilon \leq \log \frac{r}{s} \leq \varepsilon$$

or, in other words, $\exp(-\varepsilon)s \leq r \leq \exp(\varepsilon)s$. Note that r is an ε -approximation to s if and only if s is an ε -approximation to r . Note also that r and s must be of the same sign. Therefore, in order for the ε -approximation definition to be well-defined for $r = RD_{\text{true}}, RD_{\text{obs}}, RD_{\text{crude}}$ and $s = RD_{\text{true}}^*, RD_{\text{obs}}^*, RD_{\text{crude}}^*$, respectively, we limit our analysis to those cases where $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c and the given exposure levels a and $\bar{a}$. This implies that $RD_{\text{true}}, RD_{\text{obs}}, RD_{\text{crude}}, RD_{\text{true}}^*, RD_{\text{obs}}^*$, and RD_{crude}^* are all non-negative. Specifically, note that $RD_{\text{true}}, RD_{\text{true}}^* \geq 0$ follows from $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c . Likewise, $RD_{\text{obs}}, RD_{\text{obs}}^* \geq 0$ follows from $E[Y|a, d] \geq 0 \geq E[Y|\bar{a}, d]$ for all d , which follows by noting that $E[Y|a, d] = \sum_c E[Y|a, c]p^*(c|a, d) \geq 0$ and similarly $E[Y|\bar{a}, d] \leq 0$. Finally, $RD_{\text{crude}}, RD_{\text{crude}}^* \geq 0$ follows from $E[Y|a] \geq 0 \geq E[Y|\bar{a}]$, which follows by noting that $E[Y|a] = \sum_c E[Y|a, c]p^*(c|a) \geq 0$ and similarly $E[Y|\bar{a}] \leq 0$.

The following results indicate that Theorem 3 is not too sensitive to small violations of the assumptions.

Theorem 11. Assume that the conditions of Theorem 3 hold. Moreover, let $p(c)$, $p(a|c)$, and $p(d|c)$ be ε -approximations to $p^*(c)$, $p^*(a|c)$, and $p^*(d|c)$ for all a, c , and d . Given two exposure levels a and $\bar{a}$, if

$E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c and $\left| \log \frac{RD_{true}}{RD_{obs}} \right| \geq 3\epsilon$ and $\left| \log \frac{RD_{obs}}{RD_{crude}} \right| \geq 6\epsilon$, then the ordering among RD_{true}^* , RD_{crude}^* , and RD_{obs}^* is the same as the ordering among RD_{true} , RD_{crude} , and RD_{obs} .

Corollary 12. Assume that the conditions of Theorem 3 hold. Moreover, let $p(c)$, $p(a|c)$, and $p(d|c)$ be ϵ -approximations to $p^*(c)$, $p^*(a|c)$, and $p^*(d|c)$ for all a , c , and d . Given two exposure levels a and $\bar{a}$, if $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c and $\left| \log \frac{RD_{true}}{RD_{obs}} \right| \geq 3\epsilon$ and the ordering between RD_{crude}^* and RD_{obs}^* is the same as the ordering between RD_{crude} and RD_{obs} , then the ordering among RD_{true}^* , RD_{crude}^* , and RD_{obs}^* is the same as the ordering among RD_{true} , RD_{crude} , and RD_{obs} .

We performed simulations to investigate the conclusion above that Theorem 3 is not too sensitive to small departures from the assumptions. Specifically, we let A , C , and D be ternary random variables, and Y be binary. Therefore, these simulations somehow complement the sensitivity analysis above, in the sense that we do not assume that $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$. We consider $p(C) = (1/3, 1/3, 1/3)$, and $p(Y = 1|a, c) \sim \text{Uniform}(0, 1)$ for all a and c . Moreover, we consider

- $p(A|C = 1) = (p_A, (1 - p_A)/2, (1 - p_A)/2)$,
- $p(A|C = 2) = ((1 - p_A)/2, p_A, (1 - p_A)/2)$, and
- $p(A|C = 3) = ((1 - p_A)/2, (1 - p_A)/2, p_A)$,

with $p_A \sim \text{Uniform}(1/3, 3/4)$ to avoid almost deterministic relations, and likewise for $p(D|C)$. We consider four values for ϵ in the interval $[0.03, 0.15]$. For each value of ϵ , we sample 10,000 sets of distributions $\{p^*(C), p^*(A|C), p^*(D|C), p^*(Y|A, C)\}$ such that $p^*(c)$, $p^*(a|c)$, and $p^*(d|c)$ are ϵ -approximations to $p(c)$, $p(a|c)$, and $p(d|c)$ for all a , c , and d . The code to generate the simulations is available at https://www.dropbox.com/s/oh9pazehqkp8ty3/necessityCategorical_experiments_random_pref_distWeb.zip?dl=0.

The top plot in Figure 2 summarizes the distributions $p^*(C)$ sampled, demonstrating how much they deviate from $p(C)$. Similar deviations are observed for $p^*(A|c)$ and $p^*(D|c)$ with respect to $p(A|c)$ and $p(D|c)$ for all c . For each set of distributions sampled, we checked whether the conclusion of Theorem 3 holds or not. The bottom plot in Figure 2 shows the percentage of violations. Although the percentage of violations increases with increasing ϵ , it does not do so abruptly. Thus, the theorem is likely to hold when the conditions are only approximately met. In further simulations we found, not presented here, that the percentage of violations grows linearly up to $\epsilon = 1.1$ where it plateaus at 22%.

5 Discussion

It may seem that adjusting for a proxy of a latent confounder is always superior to not adjusting. Unfortunately, this is true only in some cases. In this article, we have characterized two such cases: one for monotonic settings and the other for non-monotonic settings. For each case, we have described sufficient conditions under which adjusting for a proxy of a latent confounder comes closer to the unidentifiable true average causal effect than not adjusting at all. We have also shown that the result for non-monotonic settings may continue to hold under small violations of the assumptions. Thus, it is likely our suggested set of assumptions is not the weakest possible set of assumptions that guarantees the result. However, in this same non-monotonic setting we have argued that the assumptions are not excessive, as we easily found counterexamples where the result did not hold when the assumptions were violated.

The assumptions in this work are not empirically testable in most settings, and thus, substantive knowledge is needed to confirm that they hold. It is of note that when A , C , D , and Y are all continuous and follow the linear structural equation model represented by the path diagram to the right in Figure 1, it is possible to construct testable hypothesis for the conditions, as we demonstrate in Appendix E. It is a future area of research to investigate if the assumptions can be replaced by empirically testable assumptions that are still realistic in other settings.

Figure 2: Top: Summary of the 10,000 distributions $p^*(C)$ sampled for different values of ϵ . Bottom: Percentage of the sampled distributions for which the conclusion of Theorem 3 does not hold.

Acknowledgements: The authors gratefully acknowledge financial support from the Swedish Research Council (VR SWE-REG grants 2019-00245 and 2019-00227).

Conflict of interest: Authors state no conflict of interest.

References

- [1] Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol.* 1980;112:564–9.
- [2] Ogburn EL, VanderWeele TJ. On the nondifferential misclassification of a binary confounder. *Epidemiology.* 2012;23:433–9.
- [3] Ogburn EL, VanderWeele TJ. Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. *Biometrika.* 2013;100:241–8.
- [4] Peña JM. On the monotonicity of a nondifferentially mismeasured binary confounder. *J Causal Infer.* 2020;8:150–63.
- [5] Wellman MP. Fundamental concepts of qualitative probabilistic networks. *Artif Intell.* 1990;44:257–303.
- [6] Pearl J. Causality: models, reasoning, and inference. Cambridge, UK: Cambridge University Press; 2009.
- [7] Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
- [8] Pearl J. Linear models: a useful “Microscope” for causal analysis. *J Causal Infer.* 2013;1:155–70.
- [9] Wright S. Correlation and causation. *J Agricul Res.* 1921;20:557–85.

Appendix

A Proofs for Section 3

Lemma 13. *Let A and Y be discrete or continuous random variables. Let C and D be ordinal random variables with levels $1, 2, \dots, K$. Let $p(A, D)$ be strictly positive. Let $A = a$ be any exposure level. If $p(a|C)$ and $E[Y|a, C]$ are both non-decreasing or non-increasing in C , then $S_a \geq E[Y_a]$. If one of $p(a|C)$ and $E[Y|a, C]$ is non-decreasing and the other non-increasing in C , then $S_a \leq E[Y_a]$.*

Proof. We prove that $S_a \geq E[Y_a]$ when $p(a|C)$ and $E[Y|a, C]$ are both non-decreasing in C . The other cases can be proven analogously. Specifically, we want to show that

$$\begin{aligned} S_a &= \sum_d E[Y|a, d]p(d) \\ &= \sum_c \sum_d E[Y|a, c, d]p(c|a, d)p(d) \\ &= \sum_c E[Y|a, c] \sum_d p(c|a, d)p(d) \\ &\geq \sum_c E[Y|a, c]p(c) = E[Y_a]. \end{aligned} \quad (2)$$

Actually, it suffices to show that

$$\sum_{c \geq k} \sum_d p(c|a, d)p(d) \geq \sum_{c \geq k} p(c) \quad (3)$$

for all k . To see it, let $E_c = E[Y|a, c]$, $\alpha_c = \sum_d p(c|a, d)p(d)$, and $\beta_c = p(c)$ for all c . Then, equation (2) can be rewritten as

$$\begin{aligned} S_a &= E_1\alpha_1 + \dots + E_K\alpha_K \\ &= E_1(\alpha_1 + \dots + \alpha_K) + (E_2 - E_1)(\alpha_2 + \dots + \alpha_K) + \dots + (E_K - E_{K-1})\alpha_K \\ &\geq E_1(\beta_1 + \dots + \beta_K) + (E_2 - E_1)(\beta_2 + \dots + \beta_K) + \dots + (E_K - E_{K-1})\beta_K \\ &= E_1\beta_1 + \dots + E_K\beta_K = E[Y_a], \end{aligned}$$

where $E_k - E_{k-1} \geq 0$ for all k because $E[Y|a, C]$ is non-decreasing in C by assumption, and $\alpha_1 + \dots + \alpha_K = \beta_1 + \dots + \beta_K = 1$. The latter is important because E_1 may be negative.

Proving equation (3) is equivalent to proving

$$\sum_d p(d) \sum_{c \geq k} p(c|a, d) \geq \sum_d p(d) \sum_{c \geq k} p(c|d)$$

and, thus, it suffices to prove that

$$\sum_{c \geq k} p(c|a, d) \geq \sum_{c \geq k} p(c|d)$$

holds for all d since $p(A, D)$ is strictly positive by assumption or, equivalently, that

$$\frac{\sum_{c \geq k} p(a|c)p(d|c)p(c)}{\sum_c p(a|c)p(d|c)p(c)} \geq \frac{\sum_{c \geq k} p(d|c)p(c)}{\sum_c p(d|c)p(c)}$$

holds for all d . We actually prove that

$$\left[\sum_{c \geq k} p(a|c)p(d, c) \right] \left[\sum_c p(d, c) \right] \geq \left[\sum_{c \geq k} p(d, c) \right] \left[\sum_c p(a|c)p(d, c) \right].$$

To do so, we rewrite the previous inequality as

$$\sum_{c \geq k} \sum_{\bar{c}} p(a|c)p(d, c)p(d, \bar{c}) \geq \sum_{c \geq k} \sum_{\bar{c}} p(a|\bar{c})p(d, c)p(d, \bar{c})$$

or, equivalently, as

$$\begin{aligned} & \sum_{c \geq k} \sum_{\bar{c} < k} p(a|c)p(d, c)p(d, \bar{c}) + \sum_{c \geq k} \sum_{c > \bar{c} \geq k} p(a|c)p(d, c)p(d, \bar{c}) \\ & + \sum_{c \geq k} \sum_{\bar{c} > c \geq k} p(a|c)p(d, c)p(d, \bar{c}) + \sum_{c \geq k} \sum_{\bar{c} = c \geq k} p(a|c)p(d, c)p(d, \bar{c}) \\ & \geq \sum_{c \geq k} \sum_{\bar{c} < k} p(a|\bar{c})p(d, c)p(d, \bar{c}) + \sum_{c \geq k} \sum_{c > \bar{c} \geq k} p(a|\bar{c})p(d, c)p(d, \bar{c}) \\ & + \sum_{c \geq k} \sum_{\bar{c} > c \geq k} p(a|\bar{c})p(d, c)p(d, \bar{c}) + \sum_{c \geq k} \sum_{\bar{c} = c \geq k} p(a|\bar{c})p(d, c)p(d, \bar{c}). \end{aligned}$$

Now, note that the second, third, and fourth terms of the left side coincide, respectively, with the third, second, and fourth terms of the right side. Finally, the first term of the left side is equal or greater than the first term of the right side by the assumption that $p(a|C)$ is non-decreasing in C . $\square$

Lemma 14. Let A and Y be discrete or continuous random variables. Let C and D be ordinal random variables with levels $1, 2, \dots, K$. Let $p(D|C)$ be a tapered distribution. Let $A = a$ be any exposure level. If $p(a|C)$ and $E[Y|a, C]$ are both non-decreasing or non-increasing in C , then $E[Y|a] \geq S_a$. If one of $p(a|C)$ and $E[Y|a, C]$ is non-decreasing and the other non-increasing in C , then $E[Y|a] \leq S_a$.

Proof. We prove that $E[Y|a] \geq S_a$ when $p(a|C)$ and $E[Y|a, C]$ are both non-decreasing in C . The other cases can be proven analogously. Specifically, we show that the minimum of $E[Y|a] - S_a$ with respect to $E[Y|a, c]$ is 0 for all c , regardless of $p(a|C)$, $p(C)$, and $p(D|C)$. Specifically, note that

$$\begin{aligned} E[Y|a] - S_a &= \sum_c E[Y|a, c]p(c|a) - \sum_d E[Y|a, d]p(d) \\ &= \sum_c E[Y|a, c]p(c|a) - \sum_d \sum_c E[Y|a, c, d]p(c|a, d)p(d) \\ &= \sum_c E[Y|a, c] \frac{p(a|c)p(c)}{p(a)} - \sum_d \sum_c E[Y|a, c] \frac{p(d|c)p(c|a)}{p(d|a)} p(d) \\ &= \sum_c E[Y|a, c] \frac{p(a|c)p(c)}{p(a)} - \sum_d \sum_c E[Y|a, c] \frac{p(d|c) \frac{p(a|c)p(c)}{p(a)}}{\frac{p(a|d)p(d)}{p(a)}} p(d) \\ &= \sum_c E[Y|a, c] \frac{p(a|c)p(c)}{p(a)} - \sum_d \sum_c E[Y|a, c] \frac{p(d|c)p(a|c)p(c)}{p(a|d)}. \end{aligned} \quad (4)$$

Therefore, the derivative of $E[Y|a] - S_a$ with respect to $E[Y|a, c]$ for any c is

$$p(a|c)p(c) \left(\frac{1}{p(a)} - \sum_d \frac{p(d|c)}{p(a|d)} \right). \quad (5)$$

Moreover, we prove below that $\sum_d p(d|C)/p(a|d)$ is non-increasing in C . Then, because $1/p(a)$ is constant in C , one of the following cases must hold: equation (5) is positive for all c , or it is negative for all c , or it is negative for all c below some cutoff and positive for all c above the cutoff. In all three cases, equation (4) is minimized by setting $E[Y|a, C]$ to be constant in C , because $E[Y|a, C]$ is non-decreasing in C by assumption. This implies that the minimum of equation (4) is 0.

Finally, note that if $p(a|D)$ is non-decreasing in D then, for any c and $i < j$, increasing $p(D = i|c)$ by α while reducing $p(D = j|c)$ by α increases the value of $\sum_d p(d|c)/p(a|d)$ or leaves it unchanged. By the assumption that $p(D|C)$ is tapered, we can transform $p(D|\bar{c})$ into $p(D|c)$ for any $c < \bar{c}$ by moving probability mass as just explained and, moreover, $\sum_d p(d|c)/p(a|d) \geq \sum_d p(d|\bar{c})/p(a|d)$. Therefore, $\sum_d p(d|C)/p(a|d)$ is non-increasing in C . We now argue that $p(a|D)$ is non-decreasing in D or, equivalently, that

$$\frac{\sum_i p(a|C=i)p(C=i)p(d|C=i)}{\sum_j p(C=j)p(d|C=j)} \leq \frac{\sum_i p(a|C=i)p(C=i)p(\bar{d}|C=i)}{\sum_j p(C=j)p(\bar{d}|C=j)}$$

with $d < \bar{d}$. We actually prove that

$$\begin{aligned} 0 &\leq \sum_i \sum_j p(a|C=i)p(C=i)p(\bar{d}|C=i)p(C=j)p(d|C=j) \\ &\quad - \sum_i \sum_j p(a|C=i)p(C=i)p(d|C=i)p(C=j)p(\bar{d}|C=j) \\ &= \sum_i \sum_{j < i} [p(a|C=i) - p(a|C=j)] \times [p(C=i)p(\bar{d}|C=i)p(C=j)p(d|C=j) \\ &\quad - p(C=i)p(d|C=i)p(C=j)p(\bar{d}|C=j)]. \end{aligned}$$

Now, simply note that the differences in square brackets are non-negative, because $p(a|C)$ is non-decreasing in C and $p(D|C)$ is tapered by assumption. $\square$

Proof of Theorem 1. It follows from Lemmas 13 and 14. $\square$

Proof of Theorem 2. Assume that $E[Y|a, C]$ and $E[Y|\bar{a}, C]$ are both non-decreasing in C , and $p(a|C)$ and $p(\bar{a}|C)$ are non-decreasing and non-increasing in C , respectively. Then, $E[Y|a] \geq S_a \geq E[Y_a]$ and $E[Y|\bar{a}] \leq S_{\bar{a}} \leq E[Y_{\bar{a}}]$ by Lemmas 13 and 14. The other cases follow analogously. $\square$

B Proofs for Section 4

Lemma 15. Let Y be a discrete or continuous random variable. Let A , C , and D be categorical random variables with levels $1, 2, \dots, K$. Let $p(C)$ be a discrete uniform distribution, and let $p(A|C)$ and $p(D|C)$ be preferential distributions. Then, $p(A)$ and $p(D)$ are discrete uniform distributions.

Proof. We prove that $p(A)$ is uniform by considering any two levels of A , say a and $\bar{a}$, and noting that

$$\begin{aligned} p(a) &= \sum_c p(a|c)p(c) \\ &= p(a|c_a)p(c_a) + \sum_{c|c \neq c_a} p(a|c)p(c) \\ &= p(\bar{a}|c_{\bar{a}})p(c_{\bar{a}}) + \sum_{c|c \neq c_{\bar{a}}} p(\bar{a}|c)p(c) = p(\bar{a}). \end{aligned}$$

Likewise for $p(D)$. $\square$

Lemma 16. Let Y be a discrete or continuous random variable. Let A , C , and D be categorical random variables with levels $1, 2, \dots, K$. Let $p(C)$ be a discrete uniform distribution, and let $p(A|C)$ and $p(D|C)$ be preferential distributions. Then, for any exposure level a , we have that $p(c|a, d) = p(\bar{c}|a, \bar{d})$ for all $c, \bar{c}, d$, and $\bar{d}$ such that $c = c_a \Leftrightarrow \bar{c} = c_a$ and $d = d_c \Leftrightarrow \bar{d} = d_{\bar{c}}$. Moreover, for any two exposure levels a and $\bar{a}$, we have that $p(c|a, d) = p(\bar{c}|\bar{a}, \bar{d})$ for all $c, \bar{c}, d$, and $\bar{d}$ such that $c = c_a \Leftrightarrow \bar{c} = c_{\bar{a}}$ and $d = d_c \Leftrightarrow \bar{d} = d_{\bar{c}}$.

Proof. We prove first the result for the exposure level a . Note that $p(C|D) = p(D|C)$ by Lemma 15. Then,

$$p(c|a, d) \propto p(a|c, d)p(c|d) = p(a|c)p(d|c)$$

and likewise $p(\bar{c}|a, \bar{d}) \propto p(a|\bar{c})p(\bar{d}|\bar{c})$. Then, $p(a|c) = p(a|\bar{c})$ if $c = c_a \Leftrightarrow \bar{c} = c_a$, and $p(d|c) = p(\bar{d}|\bar{c})$ if $d = d_c \Leftrightarrow \bar{d} = d_{\bar{c}}$. We now prove the result for two exposure levels a and $\bar{a}$. Note that $p(\bar{c}|\bar{a}, \bar{d}) \propto p(\bar{a}|\bar{c})p(\bar{d}|\bar{c})$. Then, $p(a|c) = p(\bar{a}|\bar{c})$ if $c = c_a \Leftrightarrow \bar{c} = c_{\bar{a}}$, and $p(d|c) = p(\bar{d}|\bar{c})$ if $d = d_c \Leftrightarrow \bar{d} = d_{\bar{c}}$. $\square$

Lemma 17. Let Y be a discrete or continuous random variable. Let A , C , and D be categorical random variables with levels $1, 2, \dots, K$. Let $p(C)$ be a discrete uniform distribution, and let $p(A|C)$ and $p(D|C)$ be preferential distributions. Then, for any exposure level a , we have that $p(d_{c_a}|a) \geq p(d_{c_a})$ and $p(c_a|a, d_{c_a}) \geq p(c_a|a, d)$ for any $d \neq d_{c_a}$.

Proof. First, note that

$$\begin{aligned} p(d_{c_a}|a) &= \sum_c p(d_{c_a}|a, c)p(c|a) \\ &= \sum_c p(d_{c_a}|c)p(a|c) \\ &= p(d_{c_a}|c_a)p(a|c_a) + \sum_{c|c \neq c_a} p(d_{c_a}|c)p(a|c) \\ &= p_D p_A + (K-1)q_D q_A. \end{aligned}$$

Likewise, for any $d \neq d_{c_a}$, note that

$$\begin{aligned} p(d|a) &= \sum_c p(d|c)p(a|c) \\ &= p(d|c_a)p(a|c_a) + p(d|c_d)p(a|c_d) + \sum_{c|c \neq c_a, c_d} p(d|c)p(a|c) \\ &= q_D p_A + p_D q_A + (K-2)q_D q_A. \end{aligned}$$

Next, assume to the contrary that $p(d_{c_a}|a) < p(d_{c_a})$. Recall that $p(d_{c_a}) = 1/K$ by Lemma 15. Then,

$$\begin{aligned} p_D p_A + (K-1)q_D q_A &< \frac{1}{K} \\ p_D p_A + (K-1)\frac{1-p_D}{K-1}\frac{1-p_A}{K-1} &< \frac{1}{K} \\ \frac{Kp_D p_A + 1 - p_D - p_A}{K-1} &< \frac{1}{K} \\ p_A(Kp_D - 1) &< \frac{K-1}{K} - 1 + p_D = \frac{Kp_D - 1}{K} \\ p_A &< \frac{1}{K}, \end{aligned}$$

which is a contradiction.

Finally, assume to the contrary that $p(c_a|a, d_{c_a}) < p(c_a|a, d)$ for any $d \neq d_{c_a}$. Then,

$$\begin{aligned} \frac{p(a|c_a, d_{c_a})p(c_a|d_{c_a})}{p(a|d_{c_a})} &< \frac{p(a|c_a, d)p(c_a|d)}{p(a|d)} \\ p(a|c_a)p(c_a|d_{c_a})p(a|d) &< p(a|c_a)p(c_a|d)p(a|d_{c_a}) \\ p(d_{c_a}|c_a)p(d|a) &< p(d|c_a)p(d_{c_a}|a) \\ p_D(q_D p_A + p_D q_A + (K-2)q_D q_A) &< q_D(p_D p_A + (K-1)q_D q_A) \\ p_D q_D p_A + p_D p_D q_A + (K-2)p_D q_D q_A &< q_D p_D p_A + (K-1)q_D q_D q_A \\ p_D p_D q_A + (K-2)p_D q_D q_A &< (K-1)q_D q_D q_A, \end{aligned}$$

which is a contradiction. The third inequality follows from Bayes' rule. $\square$

Proof of Theorem 3. We start by establishing a relationship between RD_{obs} and RD_{true} . First, note that

$$p(a|d) = \sum_c p(a|d, c)p(c|d) = \sum_c p(a|c)p(c|d) \leq \sum_c p(a|c_a)p(c|d) = p(a|c_a)$$

and, thus,

$$p(c_a|a, d) = \frac{p(a|c_a, d)p(c_a|d)}{p(a|d)} = \frac{p(a|c_a)p(c_a|d)}{p(a|d)} \geq p(c_a|d)$$

and, thus,

$$\sum_d p(c_a|a, d)p(d) \geq \sum_d p(c_a|d)p(d) = p(c_a).$$

Then,

$$\sum_d p(c_a|a, d)p(d) = p(c_a) + \alpha$$

with $\alpha \geq 0$. Likewise, $p(a|d) \geq p(a|c)$ for any $c \neq c_a$ and, thus, $p(c|a, d) \leq p(c|d)$ and, thus, $\sum_d p(c|a, d)p(d) \leq p(c) = p(c_a)$. The last equality follows from the assumption that $p(C)$ is uniform. Moreover, the left side of the last inequality is constant for all $c \neq c_a$ by Lemmas 15 and 16. Then,

$$\sum_d p(c|a, d)p(d) = p(c_a) - \beta$$

for any $c \neq c_a$. Now, note that

$$\begin{aligned} p(c_a) + \alpha + (K-1)(p(c_a) - \beta) &= \sum_d p(c_a|a, d)p(d) + \sum_{c \neq c_a} \sum_d p(c|a, d)p(d) \\ &= \sum_d p(d)p(c_a|a, d) + \sum_d p(d) \sum_{c \neq c_a} p(c|a, d) \\ &= \sum_d p(d) \sum_c p(c|a, d) = 1, \end{aligned}$$

which implies that $\beta = \alpha/(K-1)$, because $p(c_a) = 1/K$ by the assumption that $p(C)$ is uniform. Moreover, Lemmas 15 and 16 imply that

$$\begin{aligned} \sum_d p(c_a|\bar{a}, d)p(d) &= p(c_a|\bar{a}, d_{c_a})p(d_{c_a}) + \sum_{d \neq d_{c_a}} p(c_a|\bar{a}, d)p(d) \\ &= p(c_a|a, d_{c_a})p(d_{c_a}) + \sum_{d \neq d_{c_a}} p(c_a|a, d)p(d) \\ &= \sum_d p(c_a|a, d)p(d) = p(c_a) + \alpha. \end{aligned}$$

Lemmas 15 and 16 also imply that

$$\sum_d p(c|\bar{a}, d)p(d) = \sum_d p(c|a, d)p(d) = p(c_a) - \beta$$

for any $c \neq c_a$ and $\bar{c} \neq c_a$. Consequently,

$$\begin{aligned} \text{RD}_{\text{obs}} &= \sum_d E[Y|a, d]p(d) - \sum_d E[Y|\bar{a}, d]p(d) \\ &= \sum_d \sum_c E[Y|a, d, c]p(c|a, d)p(d) - \sum_d \sum_c E[Y|\bar{a}, d, c]p(c|\bar{a}, d)p(d) \\ &= \sum_d \sum_c E[Y|a, c]p(c|a, d)p(d) - \sum_d \sum_c E[Y|\bar{a}, c]p(c|\bar{a}, d)p(d) \\ &= \sum_c E[Y|a, c] \sum_d p(c|a, d)p(d) - \sum_c E[Y|\bar{a}, c] \sum_d p(c|\bar{a}, d)p(d) \\ &= E[Y|a, c_a](p(c_a) + \alpha) + \sum_{c|c \neq c_a} E[Y|a, c](p(c_a) - \beta) - E[Y|\bar{a}, c_a](p(c_a) + \alpha) \\ &\quad - \sum_{c|c \neq c_a} E[Y|\bar{a}, c](p(c_a) - \beta) \\ &= \text{RD}_{\text{true}} + E[Y|a, c_a]\alpha - \beta \sum_{c|c \neq c_a} E[Y|a, c] - E[Y|\bar{a}, c_a]\alpha + \beta \sum_{c|c \neq c_a} E[Y|\bar{a}, c] \\ &= \text{RD}_{\text{true}} + \alpha \left(E[Y|a, c_a] - \sum_{c|c \neq c_a} E[Y|a, c]/(K-1) - E[Y|\bar{a}, c_a] + \sum_{c|c \neq c_a} E[Y|\bar{a}, c]/(K-1) \right), \end{aligned} \tag{6}$$

where the sixth equality follows from the assumption that $p(C)$ is uniform.

We continue by establishing a similar relationship between RD_{crude} and RD_{true} . First, note that

$$p(c_a|a) = p(a|c_a) = p(\bar{a}|c_{\bar{a}}) = p(c_{\bar{a}}|\bar{a}) = p(c_a) + \gamma$$

with $\gamma \geq 0$, because $p(A|C)$ is a preferential distribution by assumption while $p(C)$ and $p(A)$ are discrete uniform distributions by, respectively, assumption and Lemma 15. Likewise,

$$p(c|a) = p(a|c) = p(\bar{a}|\bar{c}) = p(\bar{c}|\bar{a}) = p(c_a) - \delta$$

for any $c \neq c_a$ and $\bar{c} \neq c_{\bar{a}}$, and $\delta = \gamma/(K-1)$. Consequently,

$$\begin{aligned} RD_{\text{crude}} &= E[Y|a] - E[Y|\bar{a}] \\ &= \sum_c E[Y|a, c]p(c|a) - \sum_c E[Y|\bar{a}, c]p(c|\bar{a}) \\ &= E[Y|a, c_a](p(c_a) + \gamma) + \sum_{c|c \neq c_a} E[Y|a, c](p(c_a) - \delta) \\ &\quad - E[Y|\bar{a}, c_{\bar{a}}](p(c_a) + \gamma) - \sum_{c|c \neq c_{\bar{a}}} E[Y|\bar{a}, c](p(c_a) - \delta) \\ &= RD_{\text{true}} + E[Y|a, c_a]\gamma - \delta \sum_{c|c \neq c_a} E[Y|a, c] - E[Y|\bar{a}, c_{\bar{a}}]\gamma + \delta \sum_{c|c \neq c_{\bar{a}}} E[Y|\bar{a}, c] \\ &= RD_{\text{true}} + \gamma \left(E[Y|a, c_a] - \sum_{c|c \neq c_a} E[Y|a, c]/(K-1) \right. \\ &\quad \left. - E[Y|\bar{a}, c_{\bar{a}}] + \sum_{c|c \neq c_{\bar{a}}} E[Y|\bar{a}, c]/(K-1) \right), \end{aligned} \quad (7)$$

where the fourth equality follows from the assumption that $p(C)$ is uniform.

Finally, we show that $\alpha \leq \gamma$ in equations (6) and (7), which implies that RD_{obs} lies between RD_{true} and RD_{crude} . Specifically,

$$\begin{aligned} \alpha - \gamma &= \sum_d p(c_a|a, d)p(d) - p(c_a) - p(c_a|a) + p(c_a) \\ &= \sum_d p(c_a|a, d)p(d) - \sum_d p(c_a|a, d)p(d|a) \\ &= p(c_a|a, d_{c_a})p(d_{c_a}) - p(c_a|a, d_{c_a})p(d_{c_a}|a) \\ &\quad + \sum_{d|d \neq d_{c_a}} p(c_a|a, d)p(d) - \sum_{d|d \neq d_{c_a}} p(c_a|a, d)p(d|a) \\ &= p(c_a|a, d_{c_a})(p(d_{c_a}) - p(d_{c_a}|a)) + p(c_a|a, \bar{d}_{c_a}) \left(\sum_{d|d \neq d_{c_a}} p(d) - \sum_{d|d \neq d_{c_a}} p(d|a) \right) \\ &= p(c_a|a, d_{c_a})(p(d_{c_a}) - p(d_{c_a}|a)) \\ &\quad + p(c_a|a, \bar{d}_{c_a})(1 - p(d_{c_a}) - (1 - p(d_{c_a}|a))) \\ &= p(c_a|a, d_{c_a})(p(d_{c_a}) - p(d_{c_a}|a)) - p(c_a|a, \bar{d}_{c_a})(p(d_{c_a}) - p(d_{c_a}|a)) \\ &= (p(c_a|a, d_{c_a}) - p(c_a|a, \bar{d}_{c_a}))(p(d_{c_a}) - p(d_{c_a}|a)) \leq 0, \end{aligned}$$

where $\bar{d}_{c_a}$ denotes any level of D other than d_{c_a} . The fourth equality follows from Lemma 16, and the final inequality from Lemma 17. $\square$

Proof of Corollary 4. Recall from equation (6) that the difference between RD_{obs} and RD_{true} decreases with decreasing α , where

$$\begin{aligned}
\alpha &= \sum_d p(c_a|a, d)p(d) - p(c_a) \\
&= p(c_a|a, d_{c_a})p(d_{c_a}) + \sum_{d \neq d_{c_a}} p(c_a|a, d)p(d) - p(c_a) \\
&= \frac{p(a|c_a, d_{c_a})p(c_a|d_{c_a})}{p(a|d_{c_a})}p(d_{c_a}) + \sum_{d \neq d_{c_a}} \frac{p(a|c_a, d)p(c_a|d)}{p(a|d)}p(d) - p(c_a) \\
&= \frac{p(a|c_a)p(d_{c_a}|c_a)}{p(d_{c_a}|a)}p(d_{c_a}) + \sum_{d \neq d_{c_a}} \frac{p(a|c_a)p(d|c_a)}{p(d|a)}p(d) - p(c_a) \\
&= \frac{p_A p_D}{p_D p_A + (K-1)q_D q_A} \frac{1}{K} + \frac{(K-1)p_A q_D}{q_D p_A + p_D q_A + (K-2)q_D q_A} \frac{1}{K} - \frac{1}{K},
\end{aligned}$$

where the fourth equality follows from Lemma 15, and the fifth is established in the proof of Lemma 17. After some algebra, we have that

$$\frac{\partial \alpha}{\partial p_D} = \frac{p_A(K-1)(p_A-1)(Kp_A-1)(Kp_D-1)(\lambda-\tau)}{K(K+p_A+p_D-Kp_Ap_D-2)^2(Kp_Ap_D-p_A-p_D+1)^2},$$

where $\lambda = K^2p_Ap_D + 2p_A + 2p_D + 2K$ and $\tau = 2Kp_Ap_D + Kp_A + Kp_D + 3$. Now, assume to the contrary that

$$\begin{aligned}
2p_A + 2p_D + 2K &\leq Kp_A + Kp_D + 3 \\
2K - 3 &\leq (K-2)p_A + (K-2)p_D \\
2K - 3 &\leq 2K - 4,
\end{aligned}$$

which is a contradiction. Moreover, $K^2p_Ap_D \geq 2Kp_Ap_D$ and, thus, $\lambda > \tau$ and, thus, $\frac{\partial \alpha}{\partial p_D} < 0$ if $p_A < 1$, which implies the desired result. $\square$

Proof of Corollary 5. Lemma 15 proves that if $p(C)$ is a discrete uniform distribution and $p(D|C)$ is a preferential distribution, then $p(D)$ is a discrete uniform distribution. Then, $p(C|D) = p(D|C)$ and, thus, $p(C|D)$ is a preferential distribution with $p_C = p_D$, $q_C = q_D$, and $\pi_C = \pi_D^{-1}$. The reverse implication can be proven analogously. Likewise replacing D with AA . $\square$

Proof of Corollary 6. By adapting equations (6) and (7). $\square$

C Proofs for Section 4.1

Proof of Theorem 7. We start by proving the first result in the theorem, specifically that $RD_{\text{crude}} \geq RD_{\text{true}}$. Note that

$$p(c|a, d) = \frac{p(a, d|c)p(c)}{p(a, d|c)p(c) + p(a, d|\bar{c})p(\bar{c})} = \frac{1}{1 + \exp(-\delta(a, d))} = \sigma(\delta(a, d)),$$

where

$$\delta(a, d) = \ln \frac{p(a, d|c)p(c)}{p(a, d|\bar{c})p(\bar{c})}$$

is known as the log odds, and $\sigma(\cdot)$ is known as the logistic sigmoid function [7, Section 4.2]. Then,

$$p(c|a, d) = \sigma\left(\ln \frac{p(a, d|c)p(c)}{p(a, d|\bar{c})p(\bar{c})}\right) = \sigma\left(\ln \frac{p(a|c)p(d|c)}{p(a|\bar{c})p(d|\bar{c})}\right), \quad (8)$$

where the second equality follows from the assumption that $p(c) = 0.5$, and the fact that A and D are conditionally independent given C due to the causal graph under consideration. Likewise,

$$p(c|a, \bar{d}) = \sigma \left(\ln \frac{p(a|c)p(\bar{d}|c)}{p(a|\bar{c})p(\bar{d}|\bar{c})} \right) \quad (9)$$

and

$$p(c|d) = \sigma \left(\ln \frac{p(d|c)}{p(d|\bar{c})} \right) \quad (10)$$

and

$$p(c|\bar{d}) = \sigma \left(\ln \frac{p(\bar{d}|c)}{p(\bar{d}|\bar{c})} \right). \quad (11)$$

Analogously,

$$p(c|a) = \sigma \left(\ln \frac{p(a|c)}{p(a|\bar{c})} \right)$$

and

$$p(\bar{c}|\bar{a}) = \sigma \left(\ln \frac{p(\bar{a}|\bar{c})}{p(\bar{a}|c)} \right).$$

Therefore, $p(c|a) \geq 0.5$ and $p(\bar{c}|\bar{a}) \geq 0.5$ due to the assumption that $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$. Now, consider the function $f(x) = x(1 - x)$. By inspecting the first and second derivatives, we can conclude that $f(x)$ has a single maximum at $x = 0.5$, and that it is increasing in the interval $[0, 0.5]$ and decreasing in the interval $[0.5, 1]$. This implies that $f(p(a|c)) = p(a|c)p(\bar{a}|c) \geq p(a|\bar{c})p(\bar{a}|\bar{c}) = f(p(\bar{a}|\bar{c}))$ due to the assumption that $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$. Then,

$$\frac{p(a|c)}{p(a|\bar{c})} \geq \frac{p(\bar{a}|\bar{c})}{p(\bar{a}|c)}, \quad (12)$$

which together with the fact that $\sigma()$ and $\ln()$ are increasing functions imply that $p(c|a) \geq p(\bar{c}|\bar{a})$.

The results in the previous paragraph allow us to write $p(c|a) = 0.5 + \alpha$ and $p(\bar{c}|\bar{a}) = 0.5 + \beta$ with $\alpha \geq \beta \geq 0$. Therefore,

$$\begin{aligned} \text{RD}_{\text{crude}} &= E[Y|a] - E[Y|\bar{a}] \\ &= E[Y|a, c]p(c|a) + E[Y|a, \bar{c}]p(\bar{c}|a) - E[Y|\bar{a}, c]p(c|\bar{a}) - E[Y|\bar{a}, \bar{c}]p(\bar{c}|\bar{a}) \\ &= E[Y|a, c](0.5 + \alpha) + E[Y|a, \bar{c}](0.5 - \alpha) - E[Y|\bar{a}, c](0.5 - \beta) - E[Y|\bar{a}, \bar{c}](0.5 + \beta) \\ &= \text{RD}_{\text{true}} + \alpha(E[Y|a, c] - E[Y|a, \bar{c}]) - \beta(E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c]), \end{aligned} \quad (13)$$

which implies that $\text{RD}_{\text{crude}} \geq \text{RD}_{\text{true}}$ because $\alpha \geq \beta \geq 0$, and $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$ by assumption.

We continue by proving that $\text{RD}_{\text{obs}} \geq \text{RD}_{\text{true}}$. First, recall equations (8)–(11). Then, $p(c|a, d) \geq p(c|d)$ and $p(c|a, \bar{d}) \geq p(c|\bar{d})$ because $\sigma()$ and $\ln()$ are increasing functions and $p(a|c)/p(a|\bar{c}) \geq 1$ by the assumption that $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$. Then,

$$p(c|a, d)p(d) + p(c|a, \bar{d})p(\bar{d}) \geq p(c|d)p(d) + p(c|\bar{d})p(\bar{d}) = p(c) = 0.5$$

by the assumption that $p(c) = 0.5$. We can analogously prove that $p(\bar{c}|\bar{a}, d)p(d) + p(\bar{c}|\bar{a}, \bar{d})p(\bar{d}) \geq 0.5$. Moreover, it also holds that

$$p(c|a, d)p(d) + p(c|a, \bar{d})p(\bar{d}) \geq p(\bar{c}|\bar{a}, d)p(d) + p(\bar{c}|\bar{a}, \bar{d})p(\bar{d}).$$

To prove this inequality and after failing to do it on our own, we resorted to the function `FindInstance` from Mathematica 12.2.0. Specifically, we used `FindInstance` to find an instance of the probabilities that satisfied the reverse of the inequality above subject to $p(c) = 0.5$, $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$, and $p(\bar{d}|\bar{c}) \geq p(d|c) \geq 0.5$. Since no such instance was found, the inequality above must hold. It is worth mentioning

that `FindInstance` works analytically and not numerically and, thus, its outcome is exact and correct. The code we used to investigate this is available at <https://www.dropbox.com/s/i8pnmm9zz4pqmgrp/inequality.nb?dl=0>.

The results in the previous paragraph allow us to write

$$p(c|a, d)p(d) + p(c|a, \bar{d})p(\bar{d}) = 0.5 + \alpha$$

and

$$p(\bar{c}|\bar{a}, d)p(d) + p(\bar{c}|\bar{a}, \bar{d})p(\bar{d}) = 0.5 + \beta$$

with $\alpha \geq \beta \geq 0$ and, thus,

$$p(\bar{c}|a, d)p(d) + p(\bar{c}|a, \bar{d})p(\bar{d}) = 1 - (p(c|a, d)p(d) + p(c|a, \bar{d})p(\bar{d})) = 0.5 - \alpha$$

and

$$p(c|\bar{a}, d)p(d) + p(c|\bar{a}, \bar{d})p(\bar{d}) = 1 - (p(\bar{c}|\bar{a}, d)p(d) + p(\bar{c}|\bar{a}, \bar{d})p(\bar{d})) = 0.5 - \beta.$$

Therefore,

$$\begin{aligned} \text{RD}_{\text{obs}} &= E[Y|a, d]p(d) + E[Y|a, \bar{d}]p(\bar{d}) - E[Y|\bar{a}, d]p(d) - E[Y|\bar{a}, \bar{d}]p(\bar{d}) \\ &= (E[Y|a, c, d]p(c|a, d) + E[Y|a, \bar{c}, d]p(\bar{c}|a, d))p(d) + (E[Y|a, c, \bar{d}]p(c|a, \bar{d}) \\ &\quad + E[Y|a, \bar{c}, \bar{d}]p(\bar{c}|a, \bar{d}))p(\bar{d}) - (E[Y|\bar{a}, c, d]p(c|\bar{a}, d) + E[Y|\bar{a}, \bar{c}, d]p(\bar{c}|\bar{a}, d))p(d) \\ &\quad - (E[Y|\bar{a}, c, \bar{d}]p(c|\bar{a}, \bar{d}) + E[Y|\bar{a}, \bar{c}, \bar{d}]p(\bar{c}|\bar{a}, \bar{d}))p(\bar{d}) \\ &= E[Y|a, c](p(c|a, d)p(d) + p(c|a, \bar{d})p(\bar{d})) + E[Y|a, \bar{c}](p(\bar{c}|a, d)p(d) + p(\bar{c}|a, \bar{d})p(\bar{d})) \\ &\quad - E[Y|\bar{a}, c](p(c|\bar{a}, d)p(d) + p(c|\bar{a}, \bar{d})p(\bar{d})) - E[Y|\bar{a}, \bar{c}](p(\bar{c}|\bar{a}, d)p(d) + p(\bar{c}|\bar{a}, \bar{d})p(\bar{d})) \\ &= E[Y|a, c](0.5 + \alpha) + E[Y|a, \bar{c}](0.5 - \alpha) - E[Y|\bar{a}, c](0.5 - \beta) - E[Y|\bar{a}, \bar{c}](0.5 + \beta) \\ &= \text{RD}_{\text{true}} + \alpha(E[Y|a, c] - E[Y|a, \bar{c}]) - \beta(E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c]), \end{aligned} \quad (14)$$

where the third equality follows from the fact that Y and D are conditionally independent given A and C due to the causal graph under consideration. Then, $\text{RD}_{\text{obs}} \geq \text{RD}_{\text{true}}$ because $\alpha \geq \beta \geq 0$, and $E[Y|a, c] - E[Y|a, \bar{c}] \geq E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c] \geq 0$ by assumption.

Finally, the second result in the theorem follows equations (13) and (14). $\square$

Proof of Theorem 8. Similar to the proof of Theorem 7. Specifically, the assumption that $p(a|c) \leq p(\bar{a}|\bar{c}) \leq 0.5$ implies that $p(a|c)p(\bar{a}|c) \leq p(a|\bar{c})p(\bar{a}|\bar{c})$, which together imply that $p(c|a) \leq p(\bar{c}|\bar{a}) \leq 0.5$, which implies that

$$\text{RD}_{\text{crude}} = \text{RD}_{\text{true}} + \alpha(E[Y|a, c] - E[Y|a, \bar{c}]) - \beta(E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c])$$

with $\alpha \leq \beta \leq 0$. This implies the stated relationships between RD_{crude} and RD_{true} . The assumption that $p(a|c) \leq p(\bar{a}|\bar{c}) \leq 0.5$ also implies that $p(a|c)/p(a|\bar{c}) \leq 1$, which implies that $p(c|a) \leq p(c|d)$ and $p(c|a, \bar{d}) \leq p(c|\bar{d})$, which together with the assumption that $p(d|c) \leq p(\bar{d}|\bar{c}) \leq 0.5$ imply that

$$p(c|a, d)p(d) + p(c|a, \bar{d})p(\bar{d}) \leq p(\bar{c}|\bar{a}, d)p(d) + p(\bar{c}|\bar{a}, \bar{d})p(\bar{d}) \leq 0.5,$$

which implies that

$$\text{RD}_{\text{obs}} = \text{RD}_{\text{true}} + \alpha(E[Y|a, c] - E[Y|a, \bar{c}]) - \beta(E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c])$$

with $\alpha \leq \beta \leq 0$. This implies the stated relationships between RD_{obs} and RD_{true} . $\square$

Proof of Theorem 9. Recall from the proof of Theorem 7 that

$$p(c|a) = \sigma \left(\ln \frac{p(a|c)p(c)}{p(a|\bar{c})p(\bar{c})} \right)$$

and

$$p(\bar{c}|\bar{a}) = \sigma\left(\ln \frac{p(\bar{a}|\bar{c})p(\bar{c})}{p(\bar{a}|c)p(c)}\right)$$

and

$$p(c) = \sigma\left(\ln \frac{p(c)}{p(\bar{c})}\right)$$

and

$$p(\bar{c}) = \sigma\left(\ln \frac{p(\bar{c})}{p(c)}\right).$$

Therefore, $p(c|a) \geq p(c)$ and $p(\bar{c}|\bar{a}) \geq p(\bar{c})$ because $\sigma(\cdot)$ and $\ln(\cdot)$ are increasing functions and $p(a|c)/p(a|\bar{c}) \geq 1$ and $p(\bar{a}|\bar{c})/p(\bar{a}|c) \geq 1$ by the assumption that $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$. Then, we can write $p(c|a) = p(c) + \alpha$ and $p(\bar{c}|\bar{a}) = p(\bar{c}) + \beta$ with $\alpha, \beta \geq 0$. Moreover, $\alpha \geq \beta$. To see it, recall from the proof of Theorem 7 that a function of the form $f(x) = x(1-x)$ has a single maximum at $x = 0.5$, and it is increasing in the interval $[0, 0.5]$ and decreasing in the interval $[0.5, 1]$. Now, note that $\sigma'(z) = \sigma(z)(1 - \sigma(z))$ [7, equation 4.88]. Then, $\sigma'(z)$ has a single maximum at $\sigma(z) = 0.5$ (i.e., at $z = 0$), and it is increasing in the interval $\{\sigma(z) | 0 \leq \sigma(z) \leq 0.5\}$ (i.e., $\{z | -\infty < z \leq 0\}$) and decreasing in the interval $\{\sigma(z) | 0.5 \leq \sigma(z) \leq 1\}$ (i.e., $\{z | 0 \leq z < +\infty\}$). In other words, $\sigma(z)$ increases at an increasing rate in the interval $(-\infty, 0]$ and increases at a decreasing rate in the interval $[0, +\infty)$. Therefore, for any $u, v \geq 0$, $\sigma(z)$ increases faster when going from $z = -u$ to $z = -u + v$ than when going from $z = u$ to $z = u + v$ and, thus, $\sigma(-u + v) - \sigma(-u) \geq \sigma(u + v) - \sigma(u)$. Then,

$$\begin{aligned} \alpha &= p(c|a) - p(c) \\ &= \sigma\left(\ln \frac{p(a|c)}{p(a|\bar{c})} + \ln \frac{p(c)}{p(\bar{c})}\right) - \sigma\left(\ln \frac{p(c)}{p(\bar{c})}\right) \\ &\geq \sigma\left(\ln \frac{p(a|c)}{p(a|\bar{c})} + \ln \frac{p(\bar{c})}{p(c)}\right) - \sigma\left(\ln \frac{p(\bar{c})}{p(c)}\right) \\ &\geq \sigma\left(\ln \frac{p(\bar{a}|\bar{c})}{p(\bar{a}|c)} + \ln \frac{p(\bar{c})}{p(c)}\right) - \sigma\left(\ln \frac{p(\bar{c})}{p(c)}\right) \\ &= p(\bar{c}|\bar{a}) - p(\bar{c}) = \beta, \end{aligned}$$

where the first inequality follows from the fact that $\sigma(-u + v) - \sigma(-u) \geq \sigma(u + v) - \sigma(u)$ with $u = \ln \frac{p(\bar{c})}{p(c)}$ and $v = \ln \frac{p(a|c)}{p(a|\bar{c})}$, and the second inequality follows from equation (12) and the fact that $\sigma(\cdot)$ and $\ln(\cdot)$ are increasing functions. Note that $u, v \geq 0$ by the assumptions that $p(c) \leq 0.5$ and $p(\bar{a}|\bar{c}) \geq p(a|c) \geq 0.5$.

Finally, note that

$$\begin{aligned} \text{RD}_{\text{crude}} &= E[Y|a] - E[Y|\bar{a}] \\ &= E[Y|a, c]p(c|a) + E[Y|a, \bar{c}]p(\bar{c}|a) - E[Y|\bar{a}, c]p(c|\bar{a}) - E[Y|\bar{a}, \bar{c}]p(\bar{c}|\bar{a}) \\ &= E[Y|a, c](p(c) + \alpha) + E[Y|a, \bar{c}](p(c) - \alpha) - E[Y|\bar{a}, c](p(\bar{c}) - \beta) - E[Y|\bar{a}, \bar{c}](p(\bar{c}) + \beta) \\ &= \text{RD}_{\text{true}} + \alpha(E[Y|a, c] - E[Y|a, \bar{c}]) - \beta(E[Y|\bar{a}, \bar{c}] - E[Y|\bar{a}, c]), \end{aligned} \quad (15)$$

which implies the desired results because $\alpha \geq \beta \geq 0$. $\square$

Proof of Theorem 10. Similar to the proof of Theorem 9. Specifically, the assumption that $p(a|c) \leq p(\bar{a}|\bar{c}) \leq 0.5$ implies that $p(c|a) \leq p(c)$ and $p(\bar{c}|\bar{a}) \leq p(\bar{c})$, which implies that $p(c|a) = p(c) + \alpha$ and $p(\bar{c}|\bar{a}) = p(\bar{c}) + \beta$ with $\alpha \leq \beta \leq 0$, because $\sigma(-u + v) - \sigma(-u) \leq \sigma(u + v) - \sigma(u)$ for all $u \geq 0$ and $v \leq 0$ and, moreover, the reverse inequality of equation (12) holds now. Finally, equation (15) implies the desired results. $\square$

D Proofs for Section 4.2

Lemma 18. Assume that the conditions of Theorem 3 hold. Moreover, let $p(c)$ be an ε -approximation to $p^*(c)$ for all c . Given two exposure levels a and $\bar{a}$, if $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c , then RD_{true} is an ε -approximation to RD_{true}^* .

Proof. Note that

$$\begin{aligned} \sum_c (E[Y|a, c] - E[Y|\bar{a}, c])p^*(c) \exp(-\varepsilon) &\leq \sum_c (E[Y|a, c] - E[Y|\bar{a}, c])p(c) \\ &\leq \sum_c (E[Y|a, c] - E[Y|\bar{a}, c])p^*(c) \exp(\varepsilon), \end{aligned}$$

which implies the desired result. $\square$

Lemma 19. Assume that the conditions of Theorem 3 hold. Moreover, let $p(c)$ and $p(d|c)$ be ε -approximations to $p^*(c)$ and $p^*(d|c)$ for all c and d . Given two exposure levels a and $\bar{a}$, if $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c , then RD_{obs} is a (2ε) -approximation to RD_{obs}^* .

Proof. First, note that

$$\sum_c p^*(d|c)p^*(c) \exp(-\varepsilon) \exp(-\varepsilon) \leq \sum_c p(d|c)p(c) \leq \sum_c p^*(d|c)p^*(c) \exp(\varepsilon) \exp(\varepsilon),$$

which implies that $p(d)$ is an (2ε) -approximation to $p^*(d)$ for all d . Then, note that

$$E[Y|a, d] = \sum_c E[Y|a, c]p^*(c|a, d) = \sum_c E[Y|a, c]p^*(c|a, d),$$

which implies that $E[Y|a, d] \geq 0$ for all d . Likewise, $E[Y|\bar{a}, d] \leq 0$ for all d . Finally, note that

$$\begin{aligned} \sum_c (E[Y|a, d] - E[Y|\bar{a}, d])p^*(d) \exp(-2\varepsilon) &\leq \sum_c (E[Y|a, d] - E[Y|\bar{a}, d])p(d) \\ &\leq \sum_c (E[Y|a, d] - E[Y|\bar{a}, d])p^*(d) \exp(2\varepsilon) \end{aligned}$$

which implies the desired result. $\square$

Lemma 20. Assume that the conditions of Theorem 3 hold. Moreover, let $p(c)$ and $p(a|c)$ be ε -approximations to $p^*(c)$ and $p^*(a|c)$ for all a and c . Given two exposure levels a and $\bar{a}$, if $E[Y|a, c] \geq 0 \geq E[Y|\bar{a}, c]$ for all c , then RD_{crude} is a (4ε) -approximation to RD_{crude}^* .

Proof. First, note that

$$\sum_c p^*(a|c)p^*(c) \exp(-\varepsilon) \exp(-\varepsilon) \leq \sum_c p(a|c)p(c) \leq \sum_c p^*(a|c)p^*(c) \exp(\varepsilon) \exp(\varepsilon),$$

which implies that $p(a)$ is a (2ε) -approximation to $p^*(a)$ for all a . Then, note that

$$\begin{aligned} \log \frac{p(c|a)}{p^*(c|a)} &= \log p(a|c) + \log p(c) - \log p(a) - \log p^*(a|c) - \log p^*(c) + \log p^*(a) \\ &= \log \frac{p(a|c)}{p^*(a|c)} + \log \frac{p(c)}{p^*(c)} - \log \frac{p(a)}{p^*(a)}, \end{aligned}$$

which implies that $p(c|a)$ is a (4ϵ) -approximation to $p^*(c|a)$ for all a and c . Finally, note that

$$\begin{aligned} & \sum_c E[Y|a, c] p^*(c|a) \exp(-4\epsilon) - \sum_c E[Y|\bar{a}, c] p^*(c|\bar{a}) \exp(-4\epsilon) \\ & \leq \sum_c E[Y|a, c] p(c|a) - \sum_c E[Y|\bar{a}, c] p(c|\bar{a}) \\ & \leq \sum_c E[Y|a, c] p^*(c|a) \exp(4\epsilon) - \sum_c E[Y|\bar{a}, c] p^*(c|\bar{a}) \exp(4\epsilon), \end{aligned}$$

which implies the desired result. $\square$

Proof of Theorem 11. Theorem 3 implies that either $\log \text{RD}_{\text{crude}} \geq \log \text{RD}_{\text{obs}} \geq \log \text{RD}_{\text{true}}$ or $\log \text{RD}_{\text{crude}} \leq \log \text{RD}_{\text{obs}} \leq \log \text{RD}_{\text{true}}$. Assume the latter without loss of generality. Note that $\log \text{RD}_{\text{true}} - \epsilon \leq \log \text{RD}_{\text{true}}^*$ by Lemma 18, and $\log \text{RD}_{\text{obs}}^* \leq \log \text{RD}_{\text{obs}} + 2\epsilon$ by Lemma 19. Finally, note that $\log \text{RD}_{\text{obs}} + 2\epsilon \leq \log \text{RD}_{\text{true}} - \epsilon$ because $\log \text{RD}_{\text{true}} - \log \text{RD}_{\text{obs}} \geq 3\epsilon$ by assumption. That $\log \text{RD}_{\text{crude}}^* \leq \log \text{RD}_{\text{obs}}^*$ can be proven analogously. $\square$

Proof of Corollary 12. Theorem 3 implies that either $\log \text{RD}_{\text{crude}} \geq \log \text{RD}_{\text{obs}} \geq \log \text{RD}_{\text{true}}$ or $\log \text{RD}_{\text{crude}} \leq \log \text{RD}_{\text{obs}} \leq \log \text{RD}_{\text{true}}$. Assume without loss of generality that $\log \text{RD}_{\text{crude}}^* \leq \log \text{RD}_{\text{obs}}^*$, which implies that $\log \text{RD}_{\text{crude}} \leq \log \text{RD}_{\text{obs}} \leq \log \text{RD}_{\text{true}}$. Finally, we can prove that $\log \text{RD}_{\text{obs}}^* \leq \log \text{RD}_{\text{true}}^*$ like in the proof of Theorem 11. $\square$

E Conditions for testable assumptions and proofs

For normally distributed settings, it is possible to find empirically testable assumptions that allow us to sort the true, crude, and partially adjusted average causal effects. To see it, suppose that the variables A , C , D , and Y are all continuous and follow the linear structural equation model represented by the path diagram to the right in Figure 1. The true, crude, and partially adjusted average causal effects of A on Y are given by the partial regression coefficients $\beta_{YA \cdot C}$, β_{YA} , and $\beta_{YA \cdot D}$, respectively. Note that the first cannot be computed because C is unobserved. The following theorem proves that the partially adjusted average causal effect lies between the true and the crude ones and, thus, it comes closer to the unidentifiable true average causal effect than the crude.

Theorem 21. Assume that the variables are standardized. If $\text{sign}(\beta) = \text{sign}(\gamma)$, then $\beta_{YA \cdot C} \leq \beta_{YA \cdot D} \leq \beta_{YA}$, else $\beta_{YA \cdot C} \geq \beta_{YA \cdot D} \geq \beta_{YA}$.

Proof. [8, Section 3.11] shows that $\beta_{YA \cdot C} = \alpha$, $\beta_{YA} = \alpha + \beta\gamma$ and

$$\beta_{YA \cdot D} = \alpha + \frac{\gamma\beta(1 - \delta^2)}{1 - \beta^2\delta^2}. \quad (16)$$

Note that the linear structural equation model corresponding to the path diagram under consideration implies that

$$A = \beta C + \epsilon_A,$$

where ϵ_A is an error term that is independent of C and, thus,

$$\text{var}(A) = \beta^2 \text{var}(C) + \text{var}(\epsilon_A),$$

where $\text{var}(A) = \text{var}(C) = 1$ due to the assumption that the variables are standardized. This implies that $\beta^2 < 1$. Similarly $\delta^2 < 1$. Then, $1 - \delta^2 \leq 1 - \beta^2\delta^2$ in equation (16). The desired result is now immediate. $\square$

Note that unlike in the discrete case, no assumptions about the model's parameter values are required to conclude that the partially adjusted average causal effect lies between the true and the crude ones. Moreover, the empirically untestable sign condition in the theorem above can be replaced by an empirically testable condition.

Corollary 22. *Under the conditions in Theorem 21, $\text{sign}(\beta) = \text{sign}(\gamma)$ if and only if $\text{sign}(\text{cov}(A, D)) = \text{sign}(\text{cov}(D, Y|A))$.*

Proof. Recall (e.g., [8, Equation 5]) that

$$\text{cov}(D, Y|A) = \text{cov}(D, Y) - \frac{\text{cov}(D, A)\text{cov}(A, Y)}{\text{var}(A)}.$$

Note that $\text{var}(A) = 1$ due to the assumption that the variables are standardized. This assumption together with path analysis imply that $\text{cov}(D, Y) = \delta\gamma + \delta\beta\alpha$, $\text{cov}(D, A) = \delta\beta$, and $\text{cov}(A, Y) = \alpha + \beta\gamma$ [8,9]. Then, $\text{cov}(D, Y|A) = \delta\gamma(1 - \beta^2)$. Moreover, $\beta^2 < 1$ as shown in the proof of Theorem 21. Then, $\text{sign}(\text{cov}(D, Y|A)) = \text{sign}(\delta\gamma)$. Finally, $\text{sign}(\text{cov}(A, D)) = \text{sign}(\delta\beta)$ by path analysis. The desired result is now immediate. $\square$