Nonparametric Bounds for Causal Effects in Imperfect Randomized Experiments

Erin E. Gabriel, Arvid Sjölander, and Michael C. Sachs
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

ABSTRACT
Nonignorable missingness and noncompliance can occur even in well-designed randomized experiments, making the intervention effect that the experiment was designed to estimate nonidentifiable. Nonparametric causal bounds provide a way to narrow the range of possible values for a nonidentifiable causal effect with minimal assumptions. We derive novel bounds for the causal risk difference for a binary outcome and intervention in randomized experiments with nonignorable missingness that is caused by a variety of mechanisms, with both perfect and imperfect compliance. We show that the so-called worst-case imputation, whereby all missing subjects on the intervention arm are assumed to have events and all missing subjects on the control or placebo arm are assumed to be event-free, can be too pessimistic in blinded studies with perfect compliance, and is not bounding the correct estimand with imperfect compliance. We illustrate the use of the proposed bounds in our motivating data example of peanut consumption on the development of peanut allergies in infants. We find that, even accounting for potentially nonignorable missingness and noncompliance, our derived bounds confirm that regular exposure to peanuts reduces the risk of development of peanut allergies, making the results of this study much more compelling.

1. Introduction
The goal of randomized experiments is to estimate the causal effect of an intervention such as a medical treatment, vaccine, or social program. However, when the sample arrived upon at the end of the study is missing information on the outcome and/or the intervention, the causal intervention effect may be nonidentifiable. When there is no missing data, randomization allows for the identification of the effect of being assigned to the intervention, sometimes called the intent to treat (ITT) effect. The assignment effect is only equivalent to the intervention effect if all subjects comply with their assigned intervention as directed. When complianc is imperfect, the intervention effect can also be nonidentifiable, even with no missing data.

There are few articles that focus on bounding nonidentifiable causal effects in randomized experiments with missing data. One such article is Horowitz and Manski (2000), in which the authors derived bounds for the risk difference conditional on a measured baseline covariate, making no assumptions about the missingness mechanism. Lee (2009) also derived bounds in this setting, but the proposed bounds are not assumption-free as they impose a model on the missingness process. Marden et al. (2018) derived bounds for population proportions under nonignorable missing data, but not bounds for causal contrasts. Practitioners almost always calculate an assumption-free lower bound when outcome data are missing in a trial by imputing missing data in the least favorable way for the intervention.

Specifically, if the intervention is expected to reduce the probability of the outcome being equal to 1, missing outcomes in the intervention arm would be imputed as 1, and in the control arm as 0, which is recommended as a sensitivity analysis by European Medicines Agency: CPMP/EWP/1776/99 (2010). Similarly, one can obtain an assumption-free upper bound by imputing in the most favorable way possible, thus obtaining what we will call the best/worst-case bounds. Noncompliance is a well-known concept in the causal inference literature. Balke and Pearl (1997) developed nonparametric bounds for the causal risk difference when subjects may not comply with the assigned intervention, and this approach has been extended to multi-arm trials (Cheng and Small 2006). When missing data are compounded by noncompliance, the standard method of best/worst-case imputation produces bounds for the ITT effect rather than the intervention effect.

Much of the literature on nonparametric causal bounds uses the method developed in Balke and Pearl (1994) for deriving valid and tight bounds. "Valid" means that there are no possible values of the true causal effect outside the bounds, while "tight" means that there are no impossible values of the true causal effect inside the bounds, given the available information and the causal assumptions. To use this method, the causal effect of interest and the constraints implied by the causal assumptions must be stated as a linear optimization problem. For this reason, much of the literature on nonparametric bounds for causal effects has focused on simple random sampling in observational...
studies and completely observed data in randomized experiments, which can be easily stated as linear programming problems provided that the causal effect is linear. Kuroki, Cai, and Geng (2010) and Gabriel, Sachs, and Sjölander (2020) were exceptions who derive bounds in settings that are nonlinear. Kuroki, Cai, and Geng (2010) derived bounds for the risk ratio under case-control and cohort sampling with and without missing non-randomized exposure data. Gabriel, Sachs, and Sjölander (2020) derived bounds under more general outcome-dependent observational studies. Although nonignorable missingness can be considered a form of outcome-dependent sampling, Gabriel, Sachs, and Sjölander (2020) and Kuroki, Cai, and Geng (2010) did not consider settings with randomized intervention.

We derive bounds for the causal risk difference in randomized experiments with nonignorable missingness that is caused by a variety of mechanisms, with both perfect and imperfect compliance. We consider two settings with perfect compliance, with differing forms of nonignorable missing data, and four settings that also have noncompliance. We only consider settings where missingness is confounded with the outcome and where observation of compliance is impossible, such as in our motivating example where the intervention (peanut exposure) may occur at any time over long-term follow-up up to the time of the outcome measurement. To our knowledge, nonparametric bounds have not been derived in any of the settings with perfect compliance that we consider. Two of the four scenarios with imperfect compliance that we consider are equivalent to instrumental variable (IV) scenarios considered in Gabriel, Sachs, and Sjölander (2020). However, the authors only derived bounds that make no assumptions about the relationship between the IV and the nonrandomized exposure. We provide novel bounds under the often plausible assumption of no defiers, that is, the absence of subjects who would take the intervention if and only if they are not assigned it. We find that these bounds are, in some settings, narrower than the bounds derived by Gabriel, Sachs, and Sjölander (2020), which allow for defiers.

Additionally, we map each of the scenarios with imperfect compliance to a corresponding scenario with perfect compliance, to enable a comparison with the best/worst-case bounds. Because these bounds target the assignment effect in the presence of noncompliance, the best/worst-case imputation can actually give much narrower bounds that do not even contain the causal effect of the intervention. For this reason, when noncompliance is possible, we recommend using our proposed bounds for the intervention effect in addition to best/worst-case imputation for the assignment effect.

Our motivating example is a randomized trial of the exposure of infants to peanut products prior to 60 months of life for the outcome of allergic reactions to peanuts at 60 months. In this trial, there was both observed noncompliance and missingness due to dropout. In the primary publication of this trial, the classic worst-case imputation method was used as a sensitivity analysis to nonignorable missingness (Du Toit et al. 2015). We demonstrate that, although the resulting bounds cover the assignment effect, they are not guaranteed to cover the intervention effect. However, as all bounds exclude a null effect, we confirm the findings of the study that exposure of infants to peanut products reduces the risk of peanut allergies later in life.

The article is structured as follows. In Section 2, we define notation and assumptions, describe the causal models of interest, and review the relevant previously derived bounds. In Section 3, we briefly describe the methods that we use before providing the novel bounds and qualitatively comparing them. In Section 4, we carry out a simulation study to assess their performance. In Section 5, we present the results of our motivating example, before providing a summary and discussion in Section 6.

2. Preliminaries

2.1. Notation

We consider a randomized trial where we let \( X \) be the binary intervention, \( Y \) the binary outcome of interest, and \( Y(x) \) be the potential (or counterfactual) outcome (Rubin 1974; Pearl 2009), if the intervention is set to level \( x \). Let \( O \) be an indicator of having observed outcome and compliance/intervention information; \( O = 1 \) for “observed” and \( O = 0 \) for “not observed.” Let \( U \) be a set of unobserved variables that are common causes of, or confounders for, \( Y, O \) and possibly \( X \) with no restrictions on the distribution of \( U \). Thus, the estimable data distribution is given by \( p(X, Y|O = 1) \); where \( p(\cdot) \) is the probability mass function. With perfect compliance we know all subjects’ \( X \) values. We can thus estimate \( p(O = 1|X = x) \), and therefore also the probabilities of the form \( p(Y = y|O = 1, X = x) = p(Y = y|O = 0, X = x)p(O = 1|X = x) \), which we will use in the construction of the bounds.

When there is noncompliance, the randomized assignment and the actual intervention are not the same for all subjects. Let \( R \) be the assignment, with \( R = 1 \) meaning that one was randomized to \( X = 1 \), and \( R = 0 \) to \( X = 0 \). Let \( Y(r) \) be the potential (or counterfactual) outcome for a given subject, if the randomization is set to level \( r \), and let \( X(r) \) be the same for the intervention. In this setting we can estimate \( p(X, Y, R|O = 1) \), but because we are only considering randomized trials, one will also always be able to estimate the marginal probabilities of \( p(R = 1) \) and \( p(O = 1) \) and therefore also \( p(O = 1|R = r) \). We can use this to estimate probabilities of the form

\[
p(X = x, Y = y, O = 1|R = r)
= \frac{p(X = x, Y = y, R = r|O = 1)}{\sum_{x,y} p(R = r, X = x, Y = y|O = 1)p(O = 1|R = r)},
\]

which we will use in the construction of the bounds.

Our target parameter is the effect of the intervention, as measured by the causal risk difference

\[
\theta = p(Y(X = 1) = 1) - p(Y(X = 0) = 1).
\]

In most trials, the intervention effect is typically of greatest interest. However, when there is noncompliance it is common to also focus on the assignment (or ITT) effect

\[
\tau = p(Y(R = 1) = 1) - p(Y(R = 0) = 1).
\]
For convenience of notation, we define the following probability abbreviations:

\[
p_{yx,1} = p(Y = y | X = x, O = 1), p_{yx,1,x} = p(Y = y, O = 1 | X = x),
\]

\[
p_{xy, r} = p(X = x, Y = y, O = 1 | R = r),
\]

\[
p_{xy, r} = p(X = x, Y = y, O = 1 | R = r),
\]

\[
p_{o,x} = p(O = 0 | X = x), q_{xy,o} = p(X = x, Y = y | O = 0).
\]

### 2.2. Settings and Assumptions

There are a number of ways to specify causal assumptions and an exhaustive review in the noncompliance setting is provided by Swanson et al. (2018). We proceed by specifying nonparametric structural equations, visualized by the causal diagrams (Pearl 2009) in Figures 1(a)–2(d), to represent possible scenarios in a randomized experiment. These causal diagrams encode the assumptions of the nonparametric structural equation models such that the value of each variable in the diagram is determined from an unspecified function of the parents of the variable and an unknown error term. For example, in Figure 1(b), the value of the variable \( Y \) is determined by \( y = f_Y(x, u, \epsilon_Y) \) given an unspecified function \( f_Y \), values \( x, u \) of \( X, U \) and the \( Y \)-specific error term \( \epsilon_Y \). The explicit structural equations for each causal diagram are given in the supplement.

Figure 1(a), (b) represent randomized experiments with perfect compliance but nonignorable missingness in the outcome. The nonignorable missingness mechanisms we consider are of two types: missingness that is confounded with the outcome, \( Y \), and also causally depends on \( Y \) (Figure 1(a)), and missingness that is confounded with the outcome, \( Y \), and also causally depends on both \( X \) and \( Y \) (Figure 1(b)).

Real-life settings that fit the perfect compliance scenarios are single time-point intervention trials where the intervention is administered at the time of randomization. Some examples are a one dose vaccine, a surgical intervention or a single dose intravenous treatment, where subjects may have previously been screened for entry into the study but are not randomized and therefore not actually enrolled until just before the intervention is performed. Although this type of randomization procedure reduces or even eliminates compliance issues, unless the endpoint is immediate, such trials can still suffer from nonignorable missingness in the outcome. In contrast, any time an intervention requires active participation from the subjects under study, compliance can be an issue.

When noncompliance is possible, so that the actual intervention \( X \) may differ from the randomized assignment \( R \), \( X \) and \( Y \) may be confounded as in all settings of Figure 2. Which variables affect \( O \) determines the missingness mechanism in each setting. The noncompliance settings we consider are again all confounded missingness, but here also dependent on \( Y \) (Figure 2(a)), also dependent on \( X \) and \( Y \) (Figure 2(b)), also dependent on \( R \) and \( Y \) (Figure 2(c)), and also dependent on \( R \), \( X \) and \( Y \) (Figure 2(d)). For all these settings, we assume that \( X \) is missing whenever \( Y \) is missing, and vice versa. Thus, since \( X \) is not observed for all subjects, \( p(X) \) is not estimable.

Real-life trials that fit Figure 2 include any take-at-home medications, diet or physical activity interventions. When such a trial uses an intervention that is available to all participants, and is not blinded to participants, any type of noncompliance is possible. For example, in a randomized trial of diet and exercise it might be the case that being told not to exercise or diet may induce some participants to exercise, while telling those same subjects to exercise might overwhelm them or make them defensive, thus inducing them to not perform the randomized intervention. For this reason, bounds not considering any further assumptions about the type of compliance may be needed in many experimental settings.

---

**Figure 1.** Causal diagrams for randomized trials with perfect compliance and nonignorable missing data.

**Figure 2.** Causal diagrams for randomized trials with noncompliance and nonignorable missing data.
In any of the settings with noncompliance, it may be of interest to further consider if it is impossible that subjects randomized to a particular intervention would defy it. This no defiers assumption can be stated in terms of counterfactuals as $X(R = 1) \geq X(R = 0)$. Angrist, Imbens, and Rubin (1996) and others have referred to this assumption as monotonicity, but we will use the term no defiers for clarity. The no defiers assumption is justified in settings with experimental intervention only available to those randomized to it. Placebo subjects will not have access to the intervention and therefore $X(R = 0) = 0$. This setting guarantees no defiers, but no defiers may be a plausible assumption in other settings as well.

2.3. Previous Bounds

Robins (1989) derived bounds for the intervention effect in randomized trials with noncompliance, assuming no missingness. However, Balke and Pearl (1997) showed that those bounds are not tight by deriving new bounds using the linear programming method developed in Balke and Pearl (1994). Gabriel, Sachs, and Sjölander (2020) derived bounds for the exposure effect in observational studies with IVs and missingness. Due to the logical equivalence between these designs and randomized trials with noncompliance (Balke and Pearl 1997), the bounds derived by Gabriel, Sachs, and Sjölander (2020) apply in the settings of Figure 2(a) and 2(b), when no assumptions are made about defiers. The bounds derived in Gabriel, Sachs, and Sjölander (2020) for Figure 2(a) are provided in the supplementary materials for reference and comparison. Kuroki, Cai, and Geng (2010) derived bounds for the exposure effect in observational studies with nonrandom (e.g., outcome dependent) sampling and potentially missing exposure information, but these authors did not consider randomized trials.

Ignoring sampling variability, the worst/best-case bounds that are often used in practice can be written in terms of probabilities as:

$$p_{1.11}p_{1.1} - p_{1.01}p_{1.0} - p_{0.0} \leq \theta \leq p_{1.11}p_{1.1} - p_{1.01}p_{1.0} + p_{0.1}. \quad (1)$$

Replacing $x$ in $p_{Y|x}$ and $p_{0,x}$ with $r$ gives the best/worst-case bounds for the assignment effect $\tau$.

Horowitz and Manski (2000) derived bounds for the conditional intervention effect given a baseline covariate in randomized trials with missingness, making no assumptions about the missingness mechanism. In the special case where there is no baseline covariate, it can easily be shown that the bounds given in their corollary 1 of Result 1 simplify to the best/worst-case bounds given in Equation (1).

3. Novel Bounds

We will use a linear programming method to derive bounds; a description of this method is given in the supplementary materials, in addition to detailed derivations for each result.

3.1. Bounds for $\theta$ Under Perfect Compliance

Result 1:

The bounds for $\theta$ given in Equation (2) are valid and tight in the setting of Figure 1(a), perfect compliance with confounded missingness also dependent on $Y$.

$$\max \left\{ \frac{p_{0.0} + p_{1.11} - 1}{p_{1.0} + 2p_{1.11} - 1} \right\} \leq \theta \leq \min \left\{ \frac{-p_{0.1} - p_{1.0} + 1}{p_{0.0} - p_{1.0} + 1} \right\}. \quad (2)$$

Result 2:

The bounds for $\theta$ given in Equation (3) are valid and tight in the settings of Figure 1(b), perfect compliance with confounded missingness also dependent on $X$ and $Y$.

$$\frac{p_{0.0} + p_{1.11} - 1}{p_{0.0} - p_{1.0} + 1} \leq \theta \leq \frac{-p_{0.1} - p_{1.0} + 1}{p_{0.0} - p_{1.0} + 1}. \quad (3)$$

3.2. Bounds for $\theta$ With Noncompliance

Result 3:

The bounds for $\theta$ given in Equations (4) and (5) are valid and tight in the settings of Figures 2(b)–(d), noncompliance with confounded missingness that is also dependent on $X$ and $Y$; $R$ and $Y$; and $R$, $X$, and $Y$; respectively.

$$\theta \geq \max \left\{ \frac{p_{0.1} + p_{1.11} - 1}{p_{0.0} + p_{1.11} - 1} \right\}; \quad (4)$$

$$\theta \leq \min \left\{ \frac{-p_{0.1} - p_{1.0} + 1}{p_{1.0} + p_{0.1} + p_{1.11} - 2} \right\}. \quad (5)$$

3.3. Bounds for $\theta$ With Noncompliance Under No Defiers

Result 4:

Under the no defiers assumption the bounds for $\theta$ given in Equations (6) and (7) are valid and tight in the setting of Figure 2(a), noncompliance with confounded missingness that is also dependent on $Y$.
\[ \theta \geq \max \left\{ \frac{p_{001.0}}{p_{111.0}} - 1, \frac{p_{001.1} - p_{011.0} + p_{011.1} - p_{111.0} + 2p_{111.1} - 1}{2p_{001.0} - p_{001.1} + p_{101.0} - p_{101.1} + p_{111.0} - 1} \right\}, \tag{6} \]

and

\[ \theta \leq \min \left\{ -\frac{p_{101.1} - p_{011.0} + 1}{-p_{101.0} - 2p_{010.0} + p_{011.1} - p_{111.0} + p_{111.1} + 1}, \frac{p_{001.0} - p_{001.1} + p_{101.0} - 2p_{101.1} - p_{011.1} + 1}{1} \right\}. \tag{7} \]

Result 5:
Under the no defiers assumption the bounds for \( \theta \) given in Equation (8) are valid and tight in the setting of Figure 2(b), noncompliance with confounded missingness that is also dependent on \( X \) and \( Y \).

\[ p_{001.0} + p_{111.0} - 1 \leq \theta \leq 1 - p_{101.1} - p_{011.0} \tag{8} \]

Result 6:
Under the no defiers assumption the bounds for \( \theta \) given in (9) are valid and tight in the settings of Figures 2(c)–(d), noncompliance with confounded missingness that is also dependent on \( R \) and \( Y \); and \( R, X \) and \( Y \), respectively.

\[ \max \left\{ \frac{p_{001.0} + p_{111.0} - 1}{p_{001.1} + p_{111.0} - 1}, \frac{p_{001.1} + p_{111.1} - 1}{p_{001.0} + p_{111.1} - 1} \right\} \leq \theta \leq \min \left\{ 1 - \frac{p_{101.1} - p_{011.0}}{1 - p_{101.0} - p_{011.1}}, \frac{1 - p_{101.0} - p_{011.0}}{1 - p_{101.0} - p_{011.1}} \right\}. \tag{9} \]

### 3.4. Bounds for \( \tau \) With Noncompliance

Under perfect compliance, \( R = X \), and therefore all Figure 1 bounds are for both \( \theta \) and \( \tau \). This is not the case with noncompliance. As randomized trials often target the assignment effect, we map the bounds for Figure 1 for \( \theta \) to the assignment effects bounds for \( \tau \) in Figure 2.

Result 7:
The bounds for \( \theta \) given in Equation (2) for the setting of Figure 1(a) (perfect compliance with confounded missingness also dependent on \( Y \)) are valid and tight for \( \tau \), replacing \( X \) with \( R \), in the setting of Figure 2(a), noncompliance with confounded missingness that is also dependent on \( Y \).

Result 8:
The bounds for \( \theta \) given in Equation (3) for the setting of Figure 1(b) (perfect compliance with confounded missingness dependent on \( X \) and \( Y \)) are valid and tight for \( \tau \), replacing \( X \) with \( R \), in the settings of Figures 2(b)–(d), noncompliance with confounded missingness that is also dependent on \( X \) and \( Y \); \( R \) and \( Y \); and \( R, X \) and \( Y \), respectively.

### 3.5. Discussion and Comparison of Results

Comparing Results 1 and 2 it can be seen that the single term bounds in Equation (3) are a subset of the min/max terms in Equation (2). This means that whenever those terms are active in Equation (2), the bounds in Equations (2) and (3) are the same. Additionally, it is easily shown that the bounds in Equation (3) are equivalent to the best/worst-case bounds in (1). Therefore, whenever one uses the best/worst-case bounds in settings with perfect compliance one is, in effect, allowing for both confounding of the outcome and the missingness, and the missingness to be influenced by both the outcome and the intervention, as in Figure 1(b). Additionally, from Result 1, and the comparison to Result 2, it can be seen that, in settings without a direct effect of \( X \) on missingness, the best/worst-case bounds will often be wider than needed and therefore overly pessimistic.

Although the bounds given in Result 3 are the same as those given in Gabriel, Sachs, and Sjölander (2020) for the setting of Figure 2(b), it has not been previously shown that these bounds are tight and valid for the settings of Figures 2(c) and (d) as well, which is a novel consequence of Result 3. This result implies that the information about the intervention effect in the observed data is the same if the missingness depends on only the randomization as if the missingness depends on only the assignment or both the intervention and the assignment. Hence, the bounds derived by Gabriel, Sachs, and Sjölander (2020) are in fact applicable to a much wider range of scenarios than previously reported.

Balke and Pearl (1997) and Robins (1989) found that, when there is no missingness, assuming no defiers in the noncompliance setting results in bounds that are a subset of the min/max terms in the bounds that allow for defiers. As the former can never be narrower than the latter, nothing is gained by assuming no defiers in this setting. However, as pointed out by Balke and Pearl (1997) nothing is lost either, since, if there are truly no defiers, then the aforementioned subset of min/max terms are guaranteed to be active in the bounds that allow for defiers, thus making the two bounds equal. In fact, this follows logically from the two bounds being both valid and tight.

A similar equality of bounds does not hold, however, when there is missingness, and the missingness mechanism is as in Figure 2(a). This is because, for this setting, the bounds assuming no defiers (in Equations (6) and (7)) are not a subset of the min/max terms in the bounds that allow for defiers (presented in displays (1) and (2) of the supplementary materials). In fact, the former are occasionally narrower when there are truly no defiers, which we demonstrate via simulation.

In the settings of Figures 2(b)–(d), we agree with the findings of Balke (1995), where the bounds assuming no defiers are a subset of the min/max terms in the bounds allowing for defiers. Hence, for these settings, there is again nothing gained (or lost) by assuming no defiers; we also demonstrate this via simulation. These results are interesting, however, because the bounds for Figures 2(b)–(d) are the same when we do not assume no defiers, but different when assuming no defiers, with the bounds under the settings in Figures 2(c) and (d) possibly being narrower than those in Figure 2(b) in some scenarios.

The bounds given in Equations (4) and (5), which are valid and tight in Figures 2(b)–(d), become the bounds in
Equation (3) when \( R = X \). A similar equivalence was observed in the noncompliance setting with no missing data in [Balke and Pearl 1997]. This is also true in the setting of Figure 2(a), although we do not reproduce the bounds not assuming no defiers here. Considering the bounds assuming no defiers given in Equations (6) and (7), it is easy to see that if \( R = X \), then these bounds become those given for Figure 1(b), Equation (2).

4. Simulations

We carried out simulation studies in order to compare the width of the true bounds across the different causal diagrams, assess the impact of the amount of missingness and amount of defiers on the width of the true bounds, and also to assess the performance of estimated bounds based on samples. For the settings in Figure 2 we generated probability distributions \( p(U, R, X, Y, O) \) under the model

\[
\begin{align*}
p(R = 1) & \sim \text{Unif}(0.2, 0.8), (\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4) \sim \text{Dirichlet}(1) \\
U &= (U_1, U_2, U_3, U_4) \sim \text{Multinomial}(1, (\upsilon_1, \ldots, \upsilon_4)) \\
X &= 0 \cdot I(U_1 = 1) + 1 \cdot I(U_2 = 1) + R \cdot I(U_3 = 1) + (1 - R) \cdot I(U_4 = 1) \\
p(Y = 1 | U, X) &= \expit(\beta_1 + \beta_2 U + \beta_3 X) \\
p(O = 1 | U, Y, X, R) &= \expit(\gamma_1 U + \gamma_2 Y + \gamma_3 X + \delta_1 \gamma_3 X + \delta_2 \gamma_4 R) \\
\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2, \gamma_4 &\sim N(0, 4)
\end{align*}
\]

(10)

where \( I(\cdot) \) is the indicator function that equals 1 if the event inside the parentheses is true and 0 otherwise, \( \expit(x) = e^x / (1 + e^x) \) and \( e \) is Euler's number. To be clear, \( \beta_2 \) and \( \gamma_4 \) are each vectors of length 4 where the elements are independent normal draws. The model in Equation (10) with \( \delta_1 = \delta_2 = 0 \) generates distributions under Figure 2(a), with \( \delta_1 = 1, \delta_2 = 0 \) generates distributions under Figure 2(b), with \( \delta_1 = 0, \delta_2 = 1 \) generates distributions under Figure 2(c), and with \( \delta_1 = \delta_2 = 1 \) generates distributions under Figure 2(d). To generate data under the no defiers assumption, we fixed \( \upsilon_4 = 0 \). For the settings in Figure 1, we instead set \( X = R \), and set \( \delta_2 = 0 \). Then, with \( \delta_1 = 0 \) we have Figure 1(a), and with \( \delta_1 = 1 \) we have Figure 1(b). To investigate the impact of the amount of missingness on the informativeness of the study, we generated distributions with a \( \gamma_4 \) fixed to yield a certain proportion of missingness given the other parameters. To investigate the impact of the proportion of compliers on the bounds, we fixed \( \upsilon_3 \) at a particular value.

Figure 3 shows the distributions of the absolute widths of the different bounds computed when 1000 distributions are generated under each of the scenarios in Figure 2, with the darker gray showing when the bounds do not cover the true intervention effect. This clearly illustrates that, in settings with noncompliance, the best/worst-case bounds often do not contain the intervention effect. It can also be seen that the bounds for the setting of Figure 2(a) sometimes do not cover the intervention effect under the data generation for settings of Figures 2(b)–(d). Though not shown in Figure 3, when generating distributions under the no defiers assumption, we find that the no defiers bounds for setting of Figure 2(a) are narrower than the bounds allowing defiers 31.2% of the time (out of 1000 distributions). Additional simulation results for the settings of Figures 1 and 2 are available in the supplementary materials.

Figure 4 shows the average width of the bounds as functions of the proportion observed (left panel) and the proportion of compliers (right panel). Even with relatively small amounts of missing data \( < 5\% \), we can see that the bounds quickly become very wide, particularly in the settings of Figure 2. The width of the bounds also appears to be approximately linearly increasing in the proportion missing. The bounds tend to get narrower as the proportion of compliers increases, but not as dramatically as when the proportion observed increases.

We also investigated the relationships between the size of the effect of \( X \) on \( Y \) on the widths of the bounds, and the size of the

![Figure 3](image-url)  
**Figure 3.** Comparison of the width of the true bounds for distributions that are generated under Figure 2. The y-axis shows the absolute width of the bounds for each setting and the best/worst-case bounds (denoted bw in the figure). The light gray dots and boxes indicate cases where the true value is within the bounds, and the dark gray bounds that where that is not the case.
effect of $R$ on $X$ on the widths of the bounds. These parameters did not have a noticeable relationship with the widths of the bounds except in extreme scenarios, as those parameters do not impact the constraints on observed proportions in the same way that the proportion missing or proportion compliers do. That is, even when the true effects are large, the observed proportions do not rule out the possibility that the observed associations are due to the unmeasured confounding. These additional simulations and several others are reported in the supplemental materials.

Up to this point the bounds have only been discussed based on true probabilities. However, all proposed bounds are functions of probabilities that can be estimated by their sample proportions to produce estimated bounds. To account for the statistical uncertainty in the estimates due to sampling we suggest the nonparametric bootstrap (Efron 1979). We investigated the properties of the bootstrap extensively, Table 1 displays one such investigation where we use quantile-based 95% confidence limits for the lower and upper bounds. Coverage of the 95% bootstrap confidence intervals for the estimated bounds are shown in Table 1 for trial sizes of 200 and 2000 for a missingness probability of 25%. We considered several values of $\beta_2 \in \{0, -\log(2), -\log(5)\}$, over 1000 simulated replicates. We fixed the values of the parameters and generated trials of size $n = 200$ or 2000 from those distributions, and calculated the empirical proportions needed to compute the bounds. The parameter values are fixed as follows: $(v_1, v_2, v_3, v_4) = (1, .3, .5, 1)$, $\beta_u = (-0.66, 2.30, -0.15, -1.19)$, $\beta_1 = -0.1$, $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (1.5, -1.56, 1.1, -0.98)$, and $\gamma_u = (-0.17, 0.01, -0.45, .25)$. We observe that a few of the confidence intervals have somewhat too small or too large coverage probability, but most have nearly 95% coverage, as expected. Using the upper confidence limit of the upper bound and the lower confidence limit of the lower bound, we observe 100% coverage of the true risk difference in these scenarios.

### Table 1. Coverage of 95% bootstrap CI for the true upper and lower bounds.

| True $\theta$ | Trial size | Causal diagrams |
|---------------|------------|----------------|
|               |            | 1a (Lower,Upper) | 1b (Lower,Upper) | 2b–d (Lower,Upper) |
| $-0.2$        | 200        | (0.92, 0.95)     | (0.93, 0.94)     | (0.94, 0.93)     |
| $-0.2$        | 2000       | (0.95, 0.95)     | (0.94, 0.92)     | (0.95, 0.94)     |
| $-0.1$        | 200        | (0.93, 0.93)     | (0.95, 0.96)     | (0.96, 0.96)     |
| $-0.1$        | 2000       | (0.96, 0.94)     | (0.97, 0.96)     | (0.96, 0.95)     |
| $0.0$         | 200        | (0.94, 0.96)     | (0.94, 0.94)     | (0.94, 0.92)     |
| $0.0$         | 2000       | (0.95, 0.92)     | (0.94, 0.94)     | (0.96, 0.94)     |

5. Real Data Application

Du Toit et al. (2015) presented the findings from a randomized controlled trial designed to estimate the causal effect of peanut consumption on the development of allergy to peanuts in infants. Total 640 participants between 4 and 11 months of age were randomized to either consume peanuts or avoid peanuts until the age of 60 months. Compliance with the assigned intervention was assessed weekly by using a food frequency questionnaire, and by manual inspection of the infants’ cribs for peanut crumbs in a randomly selected subset of participants. We defined the intervention $X$ having consumed any peanut products, which is slightly different from the intervention as used in Du Toit et al. (2015). At 60 months, the primary outcome of peanut allergy was assessed using an oral food challenge. Outcome data were missing in some participants primarily due to loss to follow up, as other protocols were in place to account for failure of the oral food challenge procedures. We therefore defined our observation variable according to loss to follow up, in which case neither the compliance nor the outcome were observable. The publicly available trial data were downloaded from the Immune Tolerance Network TrialShare website on 2020-06-15 (https://www.itntrialshare.org/, study identifier: ITN032AD).

This study clearly falls into one of the settings of Figure 2, as both compliance and missing data were issues in the study. The
primary results in the article were reported as the proportion with food allergy at 60 months in the assigned intervention groups. The per-protocol analysis and the worst-case imputation analysis were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses. The assignment with food allergy at 60 months in the assigned intervention were reported as sensitivity analyses.
in the intervention or outcome in some settings, which may lead to tighter bounds. Investigation of such additional monotonicity assumptions is a current area of research for the authors. Another avenue for future work is to consider settings where the missingness in \( X \) and \( Y \) are caused by different mechanisms. Missingness graphs (\( m \)-graphs) may be a useful tool to graphically communicate assumptions in these settings (Mohan and Pearl 2021).

**Funding**

EEG is partially supported by Swedish Research Council grant 2017-01898, AS by Swedish Research Council grant 2016-01267 and MCS by Swedish Research Council grant 2019-00227.

**ORCID**

Erin E. Gabriel @ http://orcid.org/0000-0002-0504-8404
Arvid Sjölander @ http://orcid.org/0000-0001-5226-6685
Michael C. Sachs @ http://orcid.org/0000-0002-1279-8676

**References**

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 91, 444–455. [687]

Balke, A. (1995), *Probabilistic Counterfactuals: Semantics, Computation, and Applications*, PhD thesis, UCLA Cognitive Systems Laboratory. [688]

Balke, A., and Pearl, J. (1994), “Counterfactual Probabilities: Computational Methods, Bounds and Applications,” in *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., pp. 46–54. [684,687]

——— (1997), “Bounds on Treatment Effects From Studies With Imperfect Compliance,” *Journal of the American Statistical Association*, 92, 1171–1176. [684,687,688,689]

Cheng, J., and Small, D. S. (2006), “Bounds on Causal Effects in Three-Arm Trials With Non-Compliance,” *Journal of the Royal Statistical Society*, Series B, 68, 815–836. [684]

Efron, B. (1979), “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics*, 7, 1–26. [690]

European Medicines Agency: CPMP/EWP/1776/99. Guideline on missing data in confirmatory clinical trials, rev. 1. pages 1–12, 2010. Available at https://www.ema.europa.eu/en/missing-data-confirmatory-clinical-trials. [684]

Fleming, T. R. (2011), “Addressing Missing Data in Clinical Trials,” *Annals of Internal Medicine*, 154, 113–117. [691]

Gabriel, E. E., Sachs, M. C., and Sjölander, A. (2020), “Causal Bounds for Outcome-Dependent Sampling in Observational Studies,” *Journal of the American Statistical Association*. [685,687,688]

Horowitz, J. L., and Manski, C. F. (2000), “Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data,” *Journal of the American Statistical Association*, 95, 77–84. [684,687]

Kuroki, M., Cai, Z., and Geng, Z. (2010), “Sharp Bounds on Causal Effects in Case-Control and Cohort Studies,” *Biometrika*, 97, 123–132. [685,687]

Lee, D. S. (2009), “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *The Review of Economic Studies*, 76, 1071–1102. [684]

Lee, W., Sjölander, A., Larsson, A., and Pawitan, Y. (2018), “Likelihood-Based Inference for Bounds of Causal Parameters,” *Statistics in Medicine*, 37, 4695–4706. [691]

Louis, T. A., and Zeger, S. L. (2009), “Effective Communication of Standard Errors and Confidence Intervals,” *Biostatistics*, 10, 1–2. [691]

Marden, J. R., Wang, L., Tchetgen, E. J. T., Walter, S., Glymour, M. M., and Wirth, K. E. (2018), “Implementation of Instrumental Variable Bounds for Data Missing Not at Random,” *Epidemiology* (Cambridge, Mass.), 29, 364. [684]

Mohan, K., and Pearl, J. (2021), “Graphical Models for Processing Missing Data,” *Journal of the American Statistical Association*, 1023–1037. [692]

Pearl, J. (2009), *Causality: Models, Reasoning, and Inference* (2nd ed.), New York: Cambridge University Press. [685,686]

Robins, J. (1989), “The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies,” in *Health Service Research Methodology: A Focus on AIDS*, eds. L. Sechrest, H. Freeman, and A. Mulley. US Public Health Service, National Center for Health Services Research, pp. 113–159. [687,688]

Rubin, D. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology*, 66, 688–701. [685]

Spirtes, P., and Glymour, C. (1991). “An Algorithm for Fast Recovery of Sparse Causal Graphs,” *Social Science Computer Review*, 9, 62–72. [691]

Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M., and Richardson, T. S. (2018), “Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes,” *Journal of the American Statistical Association*, 113, 933–947. [686]

Toit, G. D., Roberts, G., Sayre, P. H., Bahson, H. T., Radulovic, S., Santos, H., Brough, A., Phippard, D., Basting, M., Feeney, M. (2015), “Randomized Trial of Peanut Consumption in Infants at Risk for Peanut Allergy,” *The New England Journal of Medicine*, 372, 803–813. [685,690,691]