Using secondary outcomes and covariates to sharpen inference in randomized experiments with noncompliance[1]

Fabrizia Mealli

Department of Statistics - University of Florence - mealli@ds.unifi.it

Barbara Pacini

Department of Statistics and Mathematics - University of Pisa - barbara.pacini@sp.unipi.it

Please do not quote without permission of the authors

Abstract: We develop new methods for analyzing randomized experiments with noncompliance and, by extension, instrumental variable settings, when the often controversial, but key, exclusion restriction assumption is violated. We show how existing large-sample bounds on intention-to-treat effects for the subpopulations of compliers, never-takers and always-takers can be tightened by exploiting the joint distribution of the outcome of interest with a secondary outcome, for which the exclusion restriction is satisfied, or with a pretreatment variable. The derived bounds can be used to detect violations of the exclusion restriction and the magnitude of these violations in instrumental variables settings. It is shown that the reduced width of the bounds depends on the the strength of the association of the auxiliary variable with the primary outcome and the compliance status. We also show how the setup we consider offers new identifying assumptions of intention-to-treat effects. The role of the auxiliary information is shown in two examples of a real social job training experiment and a simulated medical randomized encouragement study. We also discuss issues of inference in finite samples and show how to conduct Bayesian analysis in our partial and point identified settings.

Keywords: Exclusion restrictions, instrumental variables, ITT effects, large-sample bounds, multiple outcomes, noncompliance.

[1]The authors would like to thank participants in the 7th IZA Conference on Labor Market Policy Evaluation, the Statistical Science Seminar at Duke University, the Causal Inference Seminar at the Harvard School of Public Health, the Statistics Seminar at the Collegio Carlo Alberto, and the Statistics Seminar at the Università Cattolica Milan for discussions, and Guido Imbens for helpful comments.

1 Introduction

The paper develops new methods for analyzing randomized experiments suffering from noncompliance. Recent literature on instrumental variables (IV: Imbens and Angrist, 1994; Angrist et al., 1996) and principal stratification (Frangakis and Rubin, 2002) shows that in these settings one can identify the causal effect of assignment among compliers, subjects who would take the treatment if offered to take it and would not take it if not offered. This result relies on some assumptions, the key one being the exclusion restriction (ER), which essentially says that the assignment to treatment has no *direct* effect on the outcome other than through the effect on the treatment receipt. This result applies also in observational studies with IV, where the instrument plays the parallel role of treatment assignment. More specifically, the ER required for *identification* rules out any effect of assignment on the outcome for those whose treatment status is not affected by assignment, i.e., the noncompliers, usually distinguished in never-takers and always-takers, whereas the ER required for *interpretation* of the effect of assignment for compliers as the causal effect of treatment receipt assumes the effect of assignment for compliers being solely due to the actual treatment receipt. In many empirical studies the ER is a controversial assumption. This is particularly true in open-label experiments, which are the norm in the social sciences (e.g., Jo, 2002, Duflo et al., 2008), in randomized encouragement studies (Hirano et al., 2000), and in observational studies with instrumental variables (e.g., Angrist, 1990; Angrist and Kruger, 1991; Hoogerheide et al., 2007).

We focus on settings where the ER is possibly violated and show how to tighten bounds on the partially identified effects of assignment using an auxiliary variable, that is, a secondary outcome for which the ER holds or a covariate that is not affected by assignment. We take an approach that is closer in spirit to Hirano et al. (2000), in the sense that we focus on partial identification of intention-to-treat (ITT) effects, that is, the effects of the assignment on the subpopulations of compliers, never-takers and always-takers.

Our choice of focussing on ITT effects has the following reasons. First, ITT effects only involve outcomes that can potentially be observed. In IV setting another *local* causal estimand of interest is the LATE (Local Average Treatment Effect), that is, the effect of the treatment receipt

for compliers. When the ER does not hold, the LATE estimand involves an *a priori counterfactual* quantity, namely the outcome for compliers when they are assigned to take the treatment and do not take it (for details see, for example, Flores and Flores-Lagunes, 2012). Because compliers *do* take the treatment if assigned to take it, data contain no information on this outcome. In this regard, focussing on ITT effects makes a clear distinction of what can be learnt from the data regarding potentially observable quantities, and what can be extrapolated on a priori counterfactuals using additional assumptions. Second, ITT effects for noncompliers are interesting per se because they provide information on the extent of the violation of ERs, and inference on these effects can be used to assess the plausibility of ERs themselves.

When the ER is violated, and the analysis is not augmented with additional assumptions, *local* ITT effects can only be partially identified. In the literature, bounds on some of these effects have been derived (Richardson et al., 2011; Huber and Mellace, 2011), borrowing results on large-sample bounds derived for so-called principal strata *direct* effects (Zhang and Rubin, 2003; Imai, 2008; Lee, 2009; Mattei and Mealli, 2011).

Other strands of literature focus on identifying different causal estimands with an invalid instrument. Manski and Pepper (2000) extend results in Manski (1990, 1994) and Balke and Pearl (1997) who derive bounds for the average treatment effect (ATE) under the ER. Manski and Pepper study partial identification of the ATE, when the usual ER is replaced by a weaker monotone instrumental variable assumption. A similar approach is followed by Flores and Flores-Lagunes (2012), who derive bounds on the LATE, without assuming the ER, but investigating different sets of assumptions imposing weak-inequality restrictions on the mean potential outcomes. Other authors develop sensitivity analysis of IV estimates in linear models under local violations of the ER (see Conley et al., 2008, 2012, and, similarly, Small, 2007, Kraay, 2012, and Nevo and Rosen, 2012).

The setup we consider is one with a binary random assignment, a binary treatment, and binary outcomes and covariates. This should not be viewed as a limit of our framework. Our results can in fact be used to point-wise bound the cumulative distribution function of a continuous outcome Y for different levels of the outcome, i.e., to derive bounds on the probabilities of the events $Y \leq y$

for each $y \in \mathcal{Y}$, where $\mathcal{Y}$ is the support of the outcome variable Y.

We derive tighter sharp bounds on ITT effects, and specifically on the distribution of potential outcomes by compliance type and assignment values, by exploiting the additional information provided by the joint distribution of the outcome of interest with secondary outcomes or covariates. Specifically, we exploit restrictions on the joint distribution of the primary outcome with an auxiliary variable implied by the randomization of treatment assignment and by the ER on the secondary outcome, and show that the reduced width of the bounds depends on the the strength of the association of the auxiliary variable with the primary outcome and the compliance status. This is a novel insight that can be useful in empirical work using instrumental variables and principal stratification strategies in general.

In what follows, we first introduce our framework and notation (Section 2). We then review, in Section 3, identification results of ITT effects on a single outcome, with and without exclusion restriction assumptions. In Section 4, sharp tighter bounds on the ITT effects on the primary outcomes are derived using auxiliary variables. In Section 5 two limiting cases are analyzed, under which bounds collapse. Section 6 introduces some additional assumptions, in the form of latent independences, that can be used as identifying conditions for ITT effects. Section 7 discusses issues of inference in finite samples and shows how to conduct Bayesian analysis in our partial and point identified settings. This is applied in two examples of a real social job training experiment and a simulated medical randomized encouragement study. Some concluding remarks are offered in Section 8.

2 Framework and Notation

Let introduce the potential outcome notation. Throughout the paper we will make the stability assumption (SUTVA; Rubin, 1978) that there is neither interference between units nor different versions of the treatment. Under SUTVA, let Z_i be a binary treatment assignment for unit i ($Z_i = 0$ if unit i is assigned to the control group, $Z_i = 1$ if unit i is assigned to the treatment group). We denote by $D_i(z)$ the binary treatment receipt for unit i (1 = treatment, 0 = control) when

assigned treatment z. $D_i(Z_i)$ denotes the actual treatment received. The two potential indicators $D_i(0)$ and $D_i(1)$ describe the compliance status and define four subpopulations: compliers (c), for whom $D_i(z) = z$ for $z \in \{0, 1\}$; never-takers (n), for whom $D_i(z) = 0$ for $z \in \{0, 1\}$; always-takers (a), for whom $D_i(z) = 1$ for $z \in \{0, 1\}$; and defiers (d), for whom $D_i(z) = 1 - z$ for $z \in \{0, 1\}$ (Angrist et al., 1996). Because only one of the two potential indicators of treatment receipt is observed, these four subpopulations are latent, in the sense that in general it is not possible to identify the specific subpopulation a unit i belongs to. We denote as G_i the subpopulation membership, which takes on values in $\{c, n, a, d\}$. We define four potential outcomes for a bivariate binary outcome,$\mathbf{Y}_i(z, d) = [Y_{i1}(z, d), Y_{i2}(z, d)]'$, for all possible combinations of treatment assignment and treatment received, $z \in \{0, 1\}$ and $d \in \{0, 1\}$. However, for every subject i, only two of the four potential outcomes are potentially observed, namely, $\mathbf{Y}_i(z, D_i(z))$, $z \in \{0, 1\}$, the other two potential outcomes being *a priori counterfactuals* (Frangakis and Rubin, 2002). In order to avoid the use of such counterfactuals, we let the binary outcome variables depend only on treatment assignment: $\mathbf{Y}_i(z) = [Y_{i1}(z), Y_{i2}(z)]'$. In our setting, the first outcome will be considered as the outcome of primary interest and the second outcome as an auxiliary variable. We will also consider cases where the auxiliary variable is a binary covariate, X_i.

In what follows we will maintain the following assumptions:

Assumption 1 *Random assignment: Z_i is randomly assigned, implying that*

$$Z_i \perp\!\!\!\perp D_i(1), D_i(0), \mathbf{Y}_i(1), \mathbf{Y}_i(0), X_i, \quad \forall i.$$

Random assignment of Z_i usually holds by design in randomized experiments; it is often plausible in instrumental variable settings, although sometimes assumed to hold conditional on covariates depending on the nature of the instrumental variable.

Assumption 2 *Nonzero effect of Z on D: $E(D_i(1) - D_i(0) \neq 0$.*

This assumption states that the assignment has a non-null effect on treatment receipt and can be verified from the data (Angrist et al., 1996).

Assumption 3 *Monotonicity of compliance:* $D_i(1) \geq D_i(0)$, $\forall i$, *which rules out the presence of defiers.*

Assumption 3 implies that the population is only composed of compliers (c), never-takers (n) and always-takers (a). This is often a reasonable assumption, which cannot be directly tested, but can be falsified by the data (e.g., Angrist et al, 1996). We denote as π_c, π_n, and π_a the proportions of c, n and a in the target population, respectively. Assumptions 2 and 3 imply that $\pi_c \neq 0$.

We introduce the following notation for the joint distribution of potential outcomes, which is summarized in Table 1:

$$P[\mathbf{Y}_i(z) = (y_1, y_2)|G_i = g, Z_i = z] = P_{gz}^{(y_1 y_2)} \tag{1}$$

for $y_1 y_2 = \{00, 01, 10, 11\}$, $z = \{0, 1\}$, $g = \{c, n, a\}$. For the corresponding marginal distributions we define:

$$P[Y_{i1}(z) = y_1|G_i = g, Z_i = z] = P_{gz}^{(y_1 \cdot)}, \tag{2}$$

$$P[Y_{i2}(z) = y_2|G_i = g, Z_i = z] = P_{gz}^{(\cdot y_2)}. \tag{3}$$

For the secondary outcome we will maintain the following stochastic exclusion restriction assumption for always-takers and never-takers:

Assumption 4 *Partial stochastic exclusion restriction:* $P_{n1}^{(\cdot 1)} = P_{n0}^{(\cdot 1)}$ *and* $P_{a1}^{(\cdot 1)} = P_{a0}^{(\cdot 1)}$.

ER assumptions are never satisfied by design and require subject matter knowledge. Typically, ERs appear plausible in blind or double blind placebo-controlled experiments: if subject do not know their initial assignment, it is reasonable to argue that assignment can affect their outcome only through the effect of treatment received. ERs may however be questionable in open-label experiments, in randomized encouragement studies, and in observational studies with instrumental variables. Open-label experiments are the norm in the social sciences, where subjects, as well as experimenters, cannot be blinded the treatment received because they actively participate to the treatment, and in general assignment can affect the outcome through channels other than the treatment. The exclusion restriction of some instruments in observational studies is also usually debated.

We argue that ERs are often more plausible for secondary outcomes (rather than for primary outcomes) for which the study was not specifically designed. For example, in open-label randomized experiments, the ER on secondary outcomes, such as side-effects, is usually plausible, because some side-effects can manifest themselves only if treatment is actually received.

For a covariate, note that, due to random assignment (Assumption 1), $Z_i \perp\!\!\!\perp X_i | D_i(1), D_i(0)$, $\forall i$. This implies that $P[X_i = 1 | Z_i = 0, G_i = g] = P[X_i = 1 | Z_i = 1, G_i = g]$ $\forall g$, $\forall i$, and this equality can be interpreted as a form of stochastic exclusion restriction which holds by design, i.e., by the randomization of the instrument, for covariates within all three latent subpopulations.

We focus on identifying intention-to-treat (ITT) effects on the first outcome, Y_1, for the subgroups of compliers, never-takers and always-takers, which are defined as:

$$E[Y_{i1}(1) - Y_{i1}(0) | G_i = g] = P_{g1}^{(1\cdot)} - P_{g0}^{(1\cdot)} \quad g \in \{c, n, a\}. \tag{4}$$

ITT effects for always-takers and never-takers reflect the effect of the assignment/instrument and can thus highlight possible violations of the exclusion restriction on the primary outcome. Differently, the ITT effect for compliers includes both the effect of assignment and the effect of treatment, and so provides information on their joint magnitude.

The data we can observe are Z_i, X_i, $D_i^{obs} = D_i(Z_i)$ and $\mathbf{Y}_i^{obs} = \mathbf{Y}_i(Z_i)$, so that the distributions that are asymptotically revealed by the sampling process are the following:

$$P[\mathbf{Y}_i^{obs} = y_1 y_2 | Z_i = z, D_i^{obs} = d],$$

$$P[Y_{i1}^{obs} = y_1 | Z_i = z, D_i^{obs} = d],$$

$$P[Y_{i2}^{obs} = y_2 | Z_i = z, D_i^{obs} = d],$$

$$P[X_i = x | Z_i = z, D_i^{obs} = d],$$

$$P[D_i^{obs} = d | Z_i = z]$$

for $y_1 \in \{0, 1\}$, $y_2 \in \{0, 1\}$, $z \in \{0, 1\}$, $d \in \{0, 1\}$, $x \in \{0, 1\}$. Notation for marginal and joint distribution of observed outcomes is summarized in Table 2. We assume these distributions are known or can be consistently estimated. Specific statistical problems related to inference in finite samples are discussed in Section 7 and detailed in Appendix C.

Due to Assumption 3, the strata proportions π_c, π_a, and π_n can be point identified[2] as

$$\pi_a = P[D_i^{obs} = 1|Z_i = 0] \tag{5}$$

$$\pi_n = P[D_i^{obs} = 0|Z_i = 1]$$

$$\pi_c = 1 - \pi_a - \pi_n.$$

3 Existing identification results for a single binary outcome

We first present identification results for the primary outcome Y_1 without imposing the ER. The bounds in the following Corollary 1 coincide with the bounds on principal *direct* effects in Zhang and Rubin (2003), Imai (2008), Lee (2009) (see also Richardson et al., 2011), and are only reformulated using our notation. In Appendix A, the proof of the following proposition is sketched, because it is instrumental in deriving subsequent results.

Proposition 1 *Under Assumptions 1, 2 and 3,* $P_{n1}^{(1\cdot)}$ *and* $P_{a0}^{(1\cdot)}$ *are point-identified as* $P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]$ *and* $P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 1]$ *respectively, while* $P_{c0}^{(1\cdot)}$ *and* $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$ *and* $P_{a1}^{(1\cdot)}$ *can be bounded. Detailed expressions are reported in the Appendix. For example, bounds for* $P_{c0}^{(1\cdot)}$ *are:*

$$L_{P_{c0}^{(1\cdot)}} = \max\left\{\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c}, 0\right\} \leq P_{c0}^{(1\cdot)} \leq \min\left\{\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, 1\right\} = U_{P_{c0}^{(1\cdot)}} \tag{6}$$

Corollary 1 *Under Assumptions 1, 2 and 3, ITT effects can be bounded as:*

$$L_{P_{c1}^{(1\cdot)}} - U_{P_{c0}^{(1\cdot)}} \leq P_{c1}^{(1\cdot)} - P_{c0}^{(1\cdot)} \leq U_{P_{c1}^{(1\cdot)}} - L_{P_{c0}^{(1\cdot)}}, \tag{7}$$

$$P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] - U_{P_{n0}^{(1\cdot)}} \leq P_{n1}^{(1\cdot)} - P_{n0}^{(1\cdot)} \leq P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] - L_{P_{n0}^{(1\cdot)}}, \tag{8}$$

$$L_{P_{a1}^{(1\cdot)}} - P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 1] \leq P_{a1}^{(1\cdot)} - P_{a0}^{(1\cdot)} \leq U_{P_{a1}^{(1\cdot)}} - P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 1]. \tag{9}$$

Let now focus on the secondary outcome, Y_2, and report the identification results in the presence of stochastic exclusion restrictions for never-takers and always-takers (Assumption 4), previously derived in Angrist et al. (1996). The proof of the following proposition is sketched in Appendix A.

[2]We say that an estimand of interest is point identified when explicit formulas for it can be provided in terms of the distribution of observed data. As is well known, such explicit formulas imply identification in the sense of Hurwicz (1950).

Proposition 2 *Under Assumptions 1, 2, 3 and 4,* $P_{c0}^{(\cdot 1)}$, $P_{c1}^{(\cdot 1)}$, $P_{n0}^{(\cdot 1)}$ *and* $P_{a1}^{(\cdot 1)}$ *can be identified as:*

$$P_{n1}^{(\cdot 1)} = P_{n0}^{(\cdot 1)} = P[Y_{i2}^{obs} = 1 | Z_i = 1, D_i^{obs} = 0], \tag{10}$$

$$P_{a0}^{(\cdot 1)} = P_{a1}^{(\cdot 1)} = P[Y_{i2}^{obs} = 1 | Z_i = 0, D_i^{obs} = 1]. \tag{11}$$

$$P_{c0}^{(\cdot 1)} = \frac{P[Y_{i2}^{obs} = 1 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 1 | Z_i = 1, D_i^{obs} = 0]}{\pi_c}, \tag{12}$$

$$P_{c1}^{(\cdot 1)} = \frac{P[Y_{i2}^{obs} = 1 | Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a) - \pi_a P[Y_{i2}^{obs} = 1 | Z_i = 0, D_i^{obs} = 1]}{\pi_c}. \tag{13}$$

The same framework can be used to derive identification results for the distribution of a binary covariate, X, within subpopulations. In this case the stochastic exclusion restriction holds by design, i.e., by the randomization of the instrument, within all three latent subpopulations, so that the distribution of X within subpopulations can be identified using analogous results:

$$P[X_i = 1 | Z_i = 1, G_i = a] = P[X_i = 1 | Z_i = 0, G_i = a] = P[X_i = 1 | Z_i = 0, D_i^{obs} = 1] \tag{14}$$

$$P[X_i = 1 | Z_i = 0, G_i = n] = P[X_i = 1 | Z_i = 1, G_i = n] = P[X_i = 1 | Z_i = 1, D_i^{obs} = 0] \tag{15}$$

$$P[X_i = 1 | Z_i = 1, G_i = c] = P[X_i = 1 | Z_i = 0, G_i = c] = \frac{P[X_i = 1 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[X_i = 1 | Z_i = 1, D_i^{obs} = 0]}{\pi_c}, \tag{16}$$

where $P[X_i = 1 | Z_i = 1, G_i = c]$ is also equal to $\frac{P[X_i=1|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a) - \pi_a P[X_i=1|Z_i=0,D_i^{obs}=1]}{\pi_c}$.

4 A bivariate binary outcome and partial exclusion restrictions

Consider now the bivariate case, with two binary outcomes. Quantities of interest related to the primary outcome can be written as follows:

$$P_{c0}^{(1\cdot)} = P_{c0}^{(11)} + P_{c0}^{(10)},$$

$$P_{c1}^{(1\cdot)} = P_{c1}^{(11)} + P_{c1}^{(10)},$$

$$P_{n0}^{(1\cdot)} = P_{n0}^{(11)} + P_{n0}^{(10)},$$

$$P_{a1}^{(1\cdot)} = P_{a1}^{(11)} + P_{a1}^{(10)}.$$

As a consequence, bounds of these quantities can be obtained by first bounding the joint probabilities and then summing up the bounds. It can be easily shown that, without imposing the exclusion restriction on any of the two outcomes, the same bounds in Proposition 1 and Corollary 1 are obtained. The secondary outcome does not help sharpening the bounds if no exclusion restriction is

imposed on it. This is a different result from the parametric case, where the joint modelling of two outcomes usually improves inference (both from a frequentist and a Bayesian perspective) in terms of increased precision and reduced bias, even if no exclusion restriction on the second is imposed (Frumento et al., 2011a; Mattei et al., 2012; Mercatanti et al., 2012; see also Jo and Muthen, 2001).

Assume now that the partial stochastic exclusion restriction (Assumption 4) holds. Assumption 4 can be also expressed as follows:

$$P_{a0}^{(\cdot 1)} = P_{a0}^{(11)} + P_{a0}^{(01)} = P_{a1}^{(\cdot 1)} = P_{a1}^{(11)} + P_{a1}^{(01)}, \tag{17}$$

$$P_{n0}^{(\cdot 1)} = P_{n0}^{(11)} + P_{n0}^{(01)} = P_{n1}^{(\cdot 1)} = P_{n1}^{(11)} + P_{n1}^{(01)}.$$

Using the following constraints characterizing joint and marginal probabilities:

$$0 \leq P_{n0}^{(11)} \leq P_{n0}^{(\cdot 1)}, 0 \leq P_{a1}^{(11)} \leq P_{a1}^{(\cdot 1)}, 0 \leq P_{c0}^{(11)} \leq P_{c0}^{(\cdot 1)}, 0 \leq P_{c1}^{(11)} \leq P_{c1}^{(\cdot 1)}, \tag{18}$$

$$0 \leq P_{n0}^{(10)} \leq P_{n0}^{(\cdot 0)}, 0 \leq P_{a1}^{(10)} \leq P_{a1}^{(\cdot 0)}, 0 \leq P_{c0}^{(10)} \leq P_{c0}^{(\cdot 0)}, 0 \leq P_{c1}^{(10)} \leq P_{c1}^{(\cdot 0)}, \tag{19}$$

together with (17), leads to tighter bounds. The intuition is that the joint probabilities are bounded above by marginal probabilities that can be identified due to the partial ER on the secondary outcome. This intuition is also reflected in the bounds, which have the structure of Fréchet bounds (Fréchet, 1951) used to bound joint distributions given marginals. This is formally shown in the proof of the following proposition (see Appendix A).

Proposition 3 *Under Assumptions 1, 2, 3 and 4, $P_{c0}^{(11)}$, $P_{c0}^{(10)}$, $P_{c1}^{(11)}$, $P_{c1}^{(10)}$, $P_{n0}^{(11)}$, $P_{n0}^{(10)}$, $P_{a1}^{(11)}$ and $P_{a1}^{(10)}$ can be bounded. Detailed expressions are reported in the Appendix. For example, bounds for $P_{c0}^{(11)}$ and $P_{c0}^{(10)}$ are:*

$$P_{c0}^{(11)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs} = 11 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 1 | Z_i = 1, D_i^{obs} = 0], 0 \right) = L_{P_{c0}^{(11)}} \tag{20}$$

$$P_{c0}^{(11)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs} = 11 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, \frac{P[Y_{i2}^{obs} = 1 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 1 | Z_i = 1, D_i^{obs} = 0]}{\pi_c} \right) = U_{P_{c0}^{(11)}}$$

$$P_{c0}^{(10)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs} = 10 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 0 | Z_i = 1, D_i^{obs} = 0], 0 \right) = L_{P_{c0}^{(10)}} \tag{21}$$

$$P_{c0}^{(10)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs} = 10 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, \frac{P[Y_{i2}^{obs} = 0 | Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 0 | Z_i = 1, D_i^{obs} = 0]}{\pi_c} \right) = U_{P_{c0}^{(10)}}$$

Corollary 2 *Under Assumptions 1, 2, 3 and 4, sharp bounds for $P_{c0}^{(1\cdot)}$, $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$ and $P_{a1}^{(1\cdot)}$, can be obtained as follows:*

$$L_{P_{c0}^{(11)}} + L_{P_{c0}^{(10)}} \leq P_{c0}^{(1\cdot)} \leq U_{P_{c0}^{(11)}} + U_{P_{c0}^{(10)}}, \tag{22}$$

$$L_{P_{c1}^{(11)}} + L_{P_{c1}^{(10)}} \leq P_{c1}^{(1\cdot)} \leq U_{P_{c1}^{(11)}} + U_{P_{c1}^{(10)}}.$$
$$L_{P_{n0}^{(11)}} + L_{P_{n0}^{(10)}} \leq P_{n0}^{(1\cdot)} \leq U_{P_{n0}^{(11)}} + U_{P_{n0}^{(10)}},$$
$$L_{P_{a1}^{(11)}} + L_{P_{a1}^{(10)}} \leq P_{a1}^{(1\cdot)} \leq U_{P_{a1}^{(11)}} + U_{P_{a1}^{(10)}}.$$

Let us take a closer look at the bounds; take the sum $L_{P_{c0}^{(11)}} + L_{P_{c0}^{(10)}}$ as an example. This sum would correspond to the lower bound obtained in Section 3 if both $L_{P_{c0}^{(11)}}$ and $L_{P_{c0}^{(10)}}$ were greater than zero. The lower bound in (22) becomes strictly greater than the lower bound in (6) if at least one of the two terms is equal to zero. For example, suppose that

$$\frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] \tag{23}$$

in (20) is > 0 and

$$\frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 0|Z_i = 1, D_i^{obs} = 0] \tag{24}$$

in (21) is < 0; in this case the *new* lower bound is equal to $\frac{P[\mathbf{Y}_i^{obs}=11|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]$ in (20) and it is greater than $\frac{P[Y_{i1}^{obs}=1|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c}$ in (6), which is implicitly obtained by adding a negative quantity, (24), to (23). The same is true for the lower bounds of $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$ and $P_{a1}^{(1\cdot)}$. As for the upper bounds, if $U_{P_{c0}^{(11)}} = \frac{P[\mathbf{Y}_i^{obs}=11|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_c}$ and $U_{P_{c0}^{(10)}} = \frac{P[\mathbf{Y}_i^{obs}=10|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_c}$, then their sum in (22) would be exactly equal to the upper bound in (6). On the contrary, if either $U_{P_{c0}^{(11)}}$ or $U_{P_{c0}^{(10)}}$ is different from the above quantities, then a strictly smaller upper bound for $P_{c0}^{(1\cdot)}$ is obtained. A similar argument holds for the upper bounds of $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$ and $P_{a1}^{(1\cdot)}$.

The bounds in Corollary 2 are therefore sharp because they improve on existing sharp bounds (Imai, 2008), by excluding the set of values of $P_{gz}^{(1\cdot)}$, $g \in \{c, a, n\}$, $z \in \{0, 1\}$, that are not consistent with the ER assumption on the secondary outcome.

Corollary 3 *Under Assumptions 1, 2, 3 and 4, sharp bounds for ITT effects are:*

$$L^*_{P_{c1}^{(1\cdot)}} - U^*_{P_{c0}^{(1\cdot)}} \leq P_{c1}^{(1\cdot)} - P_{c0}^{(1\cdot)} \leq U^*_{P_{c1}^{(1\cdot)}} - L^*_{P_{c0}^{(1\cdot)}}, \tag{25}$$

$$P[Y_{1i}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] - U^*_{P_{n0}^{(1\cdot)}} \leq P_{n1}^{(1\cdot)} - P_{n0}^{(1\cdot)} \leq P[Y_{1i}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] - L^*_{P_{n0}^{(1\cdot)}},$$

$$L^*_{P_{a1}^{(1\cdot)}} - P[Y_{1i}^{obs} = 1|Z_i = 0, D_i^{obs} = 1] \leq P_{a1}^{(1\cdot)} - P_{a0}^{(1\cdot)} \leq U^*_{P_{a1}^{(1\cdot)}} - P[Y_{1i}^{obs} = 1|Z_i = 0, D_i^{obs} = 1],$$

where

$$L^{*}_{P^{(1\cdot)}_{c0}} = L_{P^{(11)}_{c0}} + L_{P^{(10)}_{c0}}, \quad U^{*}_{P^{(1\cdot)}_{c0}} = U_{P^{(11)}_{c0}} + U_{P^{(10)}_{c0}},$$
$$L^{*}_{P^{(1\cdot)}_{c1}} = L_{P^{(11)}_{c1}} + L_{P^{(10)}_{c1}}, \quad U^{*}_{P^{(1\cdot)}_{c1}} = U_{P^{(11)}_{c1}} + U_{P^{(10)}_{c1}},$$
$$L^{*}_{P^{(1\cdot)}_{n0}} = L_{P^{(11)}_{n0}} + L_{P^{(10)}_{n0}}, \quad U^{*}_{P^{(1\cdot)}_{n0}} = U_{P^{(11)}_{n0}} + U_{P^{(10)}_{n0}},$$
$$L^{*}_{P^{(1\cdot)}_{a1}} = L_{P^{(11)}_{a1}} + L_{P^{(10)}_{a1}}, \quad U^{*}_{P^{(1\cdot)}_{a1}} = U_{P^{(11)}_{a1}} + U_{P^{(10)}_{a1}}.$$

As said in the introduction, bounds on ITT for noncompliers can inform on the extent of the violation ER on the primary outcome. Because these bounds are tighter than existing ones, they can sometimes identify the sign of the violation as well as its size. These bounds could be used to formally check if data falsify exclusion restriction assumptions on the primary outcome, similarly to what has been done by Huber and Mellace (2011). Note that, because our bounds are tighter than the ones derived from moment inequalities used to prove Proposition 1 (see Appendix A) as in Huber and Mellace (2011), the implied testing procedure will have higher power. Huber and Mellace (2011), as well as other authors (e.g., Zhang and Rubin, 2003), impose additional restrictions related to the primary outcome distribution of different subpopulations, such as stochastic dominance, to tighten bounds and thus increase testing power. We instead obtained tighter bounds without imposing any additional assumption on the primary outcome, but using restrictions following from randomization and ER on an auxiliary variable. Our use of a secondary outcome differs from common practice. Usually, in the presence of multiple outcomes, analysis is conducted separately for one outcome at a time, and the joint analysis of two (or more) outcomes is not pursued, unless analyzing their association is the goal. Joint analysis of multiple outcomes is sometimes used but only to address issues of adjustments for multiple comparisons (e.g., Hsu, 1996).

4.1 Using a covariate as an auxiliary variable

When using the joint distribution of the primary outcome and a covariate, $[Y_{i1}(z), X_i)]'$, under Assumptions 1, 2 and 3 only, bounds for the quantities in Proposition 2, Corollary 1 and Corollary 2 are obtained simply substituting Y^{obs}_{i2} with X_i in all expressions. When using a covariate, we essentially exploit the independence of the covariate and the assignment within subpopulations, which plays the same role of the exclusion restriction on the secondary outcome. In Appendix

B, we show that our bounds correspond to those that would be obtained by averaging bounds on conditional ITTs, $E[Y_{i1}(1) - Y_{i1}(0)|G_i = g, X_i = x]$ for $g \in \{c, n, a\}$ and $x \in \{0, 1\}$, over the distribution of $(X_i|G_i = g)$.

This result deserves some special remarks.

In observational studies, covariates are usually used to make identifying assumptions more plausible if stated conditional on them (Manski, 1990; Abadie, 2003; Frolich, 2006; Hong and Nekipelov, 2010): under these assumptions, bounds on conditional quantities are derived and then averaged over the distribution of covariates (e.g., Lechner and Melly, 2010). Alternatively, assuming that conditional effects are constant, bounds are tightened by intersecting bounds on conditional quantities (Manski, 1990). Lee (2009) seems to be one of the few papers where the contribution of the covariates in tightening bounds is shown explicitly in a context where conditioning on covariates is not required by the assumptions. Therefore, our result highlights the usefulness of using covariates not only when this is required by the assumptions, but in general as a tool to reduce the identified set of partially identified estimands.

In addition, in randomized experiments, covariates are usually conditioned on in order to improve the precision of causal estimates, by improving the prediction of the compliance status and the missing potential outcomes (Hirano et al., 2000). We show, however, that using the covariates can not only increase precision, but it can also tighten the bounds of partially identified estimands.

It would be possible to proceed with a conditional analysis also with a secondary outcome, but the conditional analysis would not be straightforward, as it is with a covariate. In fact, it would not simply involve stratifying on Y_{i2}^{obs}, but it would involve conditioning on $Y_{i2}(z)$ separately by treatment arm, and then combining results in a non-standard fashion (see Appendix B).

5 The role of the association of the auxiliary variable with the primary outcome and the compliance status

We have shown that assuming the exclusion restriction for a and n for the secondary outcome or using a covariate helps tightening the bounds. Now, we investigate how the width of the bounds

depends on the strength of the association of the auxiliary variable with the primary outcome and the compliance status. The intuition is that, on one hand, the auxiliary variable should help tightening the bounds the stronger its association is with the compliance status. On the other hand, we expect to sharpen inference also the stronger its association is with the primary outcome. To support these intuitions, we now consider two limiting cases. The first one is when Y_2 is perfectly associated with the compliance behavior. Specifically, suppose that $Y_{i2}^{obs} = \mathcal{I}(G_i = c)$, where $\mathcal{I}$ represents the indicator function. This implies the following equalities:

$$P_{n1}^{(\cdot 1)} = P_{n0}^{(\cdot 1)} = 0, \quad P_{a1}^{(\cdot 1)} = P_{a0}^{(\cdot 1)} = 0, \quad P_{c1}^{(\cdot 1)} = P_{c0}^{(\cdot 1)} = 1, \tag{26}$$

$$P_{n1}^{(11)} = P_{n0}^{(11)} = 0, \quad P_{a1}^{(11)} = P_{a0}^{(11)} = 0, \quad P_{c1}^{(10)} = P_{c0}^{(10)} = 0. \tag{27}$$

These equalities have implications for the observable distributions. For example, $P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]$ will be exactly 0, and $P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 0]$ will be equal to $\frac{\pi_c}{\pi_c+\pi_n}$. It is thus trivial to prove the following Corollary 4, showing that bounds in Proposition 3 would collapse if $Y_{i2}^{obs} = \mathcal{I}(G_i = c)$.

Corollary 4 *Under Assumptions 1, 2, 3, 4 and if $Y_{i2}^{obs} = \mathcal{I}(G_i = c)$, the upper and lower bounds for $P_{c0}^{(1\cdot)}$, $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$ and $P_{a1}^{(1\cdot)}$ in Proposition 3 (and so bounds for ITT effects) would be identical, and thus bounds would collapse as follows:*

$$P_{c0}^{(1\cdot)} = P_{c0}^{(11)} = \frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, \tag{28}$$

$$P_{c1}^{(1\cdot)} = P_{c1}^{(11)} = \frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c}, \tag{29}$$

$$P_{n0}^{(1\cdot)} = P_{n0}^{(10)} = \frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_n}, \tag{30}$$

$$P_{a1}^{(1\cdot)} = P_{a1}^{(10)} = \frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_a}. \tag{31}$$

Bounds collapse if the secondary outcome predicts with no uncertainty the compliance status; this is also true if we use a covariate such that $X_i = \mathcal{I}(G_i = c)$.

The second limiting case is when the secondary outcome is perfectly dependent on the primary outcome conditional on the compliance status and the treatment assignment. Specifically, suppose that, within the two observed groups where compliers are mixed with either never-takers or always-takers, $(Z_i = 0, D_i^{obs} = 0)$ and $(Z_i = 1, D_i^{obs} = 1)$, we have:

$$P_{n0}^{(11)} = P_{n0}^{(\cdot 1)} = P_{n0}^{(1\cdot)}, \quad P_{c0}^{(11)} = P_{c0}^{(\cdot 1)} = P_{c0}^{(1\cdot)}, \tag{32}$$

and

$$P^{(11)}_{a1} = P^{(\cdot 1)}_{a1} = P^{(1\cdot)}_{a1}, \quad P^{(11)}_{c1} = P^{(\cdot 1)}_{c1} = P^{(1\cdot)}_{c1}. \tag{33}$$

As before, these equalities have implications for the observable distributions. For example, $P[\mathbf{Y}^{obs}_i = 11|Z_i = 0, D^{obs}_i = 0]$ will be exactly equal to $P[Y^{obs}_{i2} = 1|Z_i = 0, D^{obs}_i = 0]$. It is thus trivial to prove the following Corollary 5, showing that bounds in Proposition 3 collapse.

Corollary 5 *Under Assumptions 1, 2, 3, 4 and if (32) and (33) hold, the lower and upper bounds for $P^{(1\cdot)}_{c0}$, $P^{(1\cdot)}_{c1}$, $P^{(1\cdot)}_{n0}$ and $P^{(1\cdot)}_{a1}$ in Proposition 3 (and so bounds for ITT effects) would be identical, so that bounds would collapse as follows:*

$$P^{(1\cdot)}_{c0} = P^{(11)}_{c0} = \frac{P[\mathbf{Y}^{obs}_i = 11|Z_i = 0, D^{obs}_i = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y^{obs}_{i2} = 1|Z_i = 1, D^{obs}_i = 0], \tag{34}$$

$$P^{(1\cdot)}_{c1} = P^{(11)}_{c1} = \frac{P[\mathbf{Y}^{obs}_i = 11|Z_i = 1, D^{obs}_i = 1](\pi_c + \pi_a)}{\pi_c} - \frac{\pi_a}{\pi_c} P[Y^{obs}_{i2} = 1|Z_i = 0, D^{obs}_i = 1], \tag{35}$$

$$P^{(1\cdot)}_{n0} = P^{(11)}_{n0} = P[Y^{obs}_{i2} = 1|Z_i = 1, D^{obs}_i = 0], \tag{36}$$

$$P^{(1\cdot)}_{a1} = P^{(11)}_{a1} = P[Y^{obs}_{i2} = 1|Z_i = 0, D^{obs}_i = 1]. \tag{37}$$

Note that perfect dependence in the sense of (32) and (33) does not imply the exclusion restriction to hold also for the primary outcome, because $P^{(1\cdot)}_{n0}$ may differ from $P^{(1\cdot)}_{n1}$, and $P^{(1\cdot)}_{a0}$ from $P^{(1\cdot)}_{a1}$. Note that perfect dependence can only occur in the data when the two outcomes have the same marginal distributions, that is, when the frequencies in one of the two diagonals are zero[3].

6 Latent independence as an identifying assumption

We have just shown that one may need to find auxiliary variables that are strongly associated with the primary outcome in order to reduce the width of the bounds.

The values of $P^{(11)}_{gz}$, $g \in \{c, a, n\}$, $z \in \{0, 1\}$ regulate the strength of the association between the two outcomes conditional on compliance status, and imply different levels of associations within observed groups. We can thus think of using restriction on these parameters, that is, restrictions on the (latent) association between the two outcomes as identifying assumption of ITT effects. As

[3]This is also the only case where the correlation coefficient (that coincides with the phi-coefficient for contingency tables) may reach its maximum absolute value of 1.

usual, identifying restrictions consider extreme cases: ER for example assumes some ITT effects to be exactly zero. Here, we investigate the role of independence, as a restriction describing an extreme form of association. Specifically, we can show that ITT effects can be point-identified if we further assume that the two outcomes are independent conditional on the compliance status. This identifying assumption is a form of *latent independence*, in the sense that independence holds only conditional on a latent variable, the compliance status (Frangakis and Rubin, 1999). This is formalized as follows:

$$P_{gz}^{(11)} = P_{gz}^{(1\cdot)} P_{gz}^{(\cdot 1)}, \quad P_{gz}^{(10)} = P_{gz}^{(1\cdot)} P_{gz}^{(\cdot 0)}, \tag{38}$$

$$P_{gz}^{(01)} = P_{gz}^{(0\cdot)} P_{gz}^{(\cdot 1)}, \quad P_{gz}^{(00)} = P_{gz}^{(0\cdot)} P_{gz}^{(\cdot 0)}, \tag{39}$$

for $g \in \{c, n, a\}$ and $z \in \{0, 1\}$.

The following proposition is proved in Appendix A.

Proposition 4 *Under Assumptions 1, 2, 3, 4, and (38) and (39), the quantities* $P_{c0}^{(1\cdot)}$, $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$ *and* $P_{a1}^{(1\cdot)}$ *can be point-identified as follows:*

$$P_{c0}^{(1\cdot)} = \frac{\pi_n + \pi_c}{\pi_c} \cdot \left\{ \frac{P[\mathbf{Y}_i^{obs} = 10 | Z_i = 0, D_i^{obs} = 0)] P_{n0}^{(\cdot 1)} - P[\mathbf{Y}_i^{obs} = 11 | Z_i = 0, D_i^{obs} = 0](1 - P_{n0}^{(\cdot 1)})}{P_{n0}^{(\cdot 1)} - P_{c0}^{(\cdot 1)}} \right\}, \tag{40}$$

$$P_{c1}^{(1\cdot)} = \frac{\pi_a + \pi_c}{\pi_c} \cdot \left\{ \frac{P[\mathbf{Y}_i^{obs} = 10 | Z_i = 1, D_i^{obs} = 1)] P_{a1}^{(\cdot 1)} - P[\mathbf{Y}_i^{obs} = 11 | Z_i = 1, D_i^{obs} = 1](1 - P_{a1}^{(\cdot 1)})}{P_{a1}^{(\cdot 1)} - P_{c1}^{(\cdot 1)}} \right\}, \tag{41}$$

$$P_{n0}^{(1\cdot)} = \frac{\pi_n + \pi_c}{\pi_n} \cdot \left\{ \frac{P[\mathbf{Y}_i^{obs} = 11 | Z_i = 0, D_i^{obs} = 0](1 - P_{c0}^{(\cdot 1)}) - P[\mathbf{Y}_i^{obs} = 10 | Z_i = 0, D_i^{obs} = 0)] P_{c0}^{(\cdot 1)}}{P_{n0}^{(\cdot 1)} - P_{c0}^{(\cdot 1)}} \right\}, \tag{42}$$

$$P_{a1}^{(1\cdot)} = \frac{\pi_a + \pi_c}{\pi_a} \cdot \left\{ \frac{P[\mathbf{Y}_i^{obs} = 11 | Z_i = 1, D_i^{obs} = 1](1 - P_{c1}^{(\cdot 1)}) - P[\mathbf{Y}_i^{obs} = 10 | Z_i = 1, D_i^{obs} = 1)] P_{c1}^{(\cdot 1)}}{P_{a1}^{(\cdot 1)} - P_{c1}^{(\cdot 1)}} \right\}. \tag{43}$$

Proposition 4 essentially shows that, if an outcome or a covariate is found that can be plausibly assumed latent independent of the primary outcome, it essentially serves as an additional instrument, and this restriction, when imposed, allows point-identification of ITT effects and so also identification of the effect of the instrument for never-takers and always-takers. As the ER, latent independence cannot be directly tested, but does have testable implications and can thus be falsified by the data. For example, if it holds, sample analogs of the quantities in (40)-(43) should lie in $[0, 1]$. Note that, if the two outcomes are indeed latent independent but latent independence is not imposed, i.e., not used as a restriction, bounds in Proposition 3 do not *automatically* collapse, unlike in the case of perfect dependence discussed in previous Section. Indeed, consistently with

what said in Section 5, bounds remain rather wide *just because* the association between the two outcomes is absent.

Results in Proposition 4 have results in Ding et al. (2011) as a special case. In Ding et al. (2011) a covariate, independent of the primary outcome, is used to identify the effect on the subpopulation of the *always-survivors*, which is analogous to the subpopulation of the always-takers. We instead use the auxiliary variable to identify effects in all the three subgroups, and extend the result to a secondary outcome.

Given this identification results, inference on ITT effects can proceed in various ways. In Section 7 we provide an empirical example, where we propose a Bayesian analysis to conduct inference under latent independence.

7 Inference

So far we have been concerned with identification as a question separate from statistical inference, essentially supposing to have access to an arbitrarily large sample size. In empirical settings, one needs to account for statistical uncertainty arising from finite sample sizes, using either a Bayesian or a frequentist perspective. The Bayesian perspective appears to have appealing properties, as advocated by, among others, Liao and Jiang (2010) and Gustafson (2010), and it is the approach followed in this paper.

In a frequentist perspective, if the bounds are explicit functionals of the observed data distributions, they can in general be consistently estimated substituting the observable distributions with their sample counterparts. The asymptotic approximations to the sampling distributions of these estimators can then be used to construct confidence regions that cover the identified set or the true parameter with at least a nominal probability (Imbens and Manski, 2004). Closed-form characterization of the asymptotic distribution of these estimators, usually involving min and max operators, are very difficult to derive, especially in multiparameter settings, and standard resampling techniques, such as the bootstrap or subsampling, typically fail, that is, applying them to the endpoints directly yield to confidence sets that are not necessarily uniformly valid (Andrews, 2000; Andrews

and Guggenberger, 2009; Romano, 1989; Romano and Shaikh, 2008, 2010). This has generated a growing literature on inference methods for partially identified parameters (see Tamer, 2010 for a review). Some authors propose to use subsampling, or similar re-sampling techniques, to derive critical values and thereby constructing confidence sets by inverting tests (e.g., Chernozhukov et al., 2007; Romano and Shaikh, 2008, 2010). An alternative approach is followed by Chernozhukov et al. (2011).

In multiparameter settings as ours, applying these methods and characterizing the confidence set is not straightforward, and the finite sample properties of commonly proposed procedures have not been thoroughly investigated and may, in general, be rather poor (e.g., Andrews and Guggenberger, 2009). Also, deriving confidence intervals for one (function of) the parameter vectors, by projecting the confidence region for the full parameter vector, in general produces very conservative confidence intervals.

Regarding these and other issues discussed in the following Section, the Bayesian perspective appears to have some advantages, at least in the setting analyzed in the paper.

7.1 Bayesian credible sets

Here we present the Bayesian approach for conducting inference for partially identified parameters, and discuss the characteristics of the posterior distribution of these parameters that allow us to construct exact credible intervals for the partially identified ITT effects, without relying on asymptotic approximations.

Identification and inference are more integrated under a Bayesian analysis: from a Bayesian perspective, there is no conceptual difference between fully and partially identified parameters: starting with a proper prior distribution over all parameters, one can interpret a posterior credible set as being likely to contain the true value of the estimands, given the combination of observed data and prior beliefs. Works considering Bayesian interval estimators in partially identified models include Moon and Schorfheide (2012), Liao and Jiang (2010), Gustafson and Greenland (2009). As the sample size increases, the marginal posterior distribution of partially identified parameters converges to a non-degenerate distribution with support equal to the identification region. Thus,

a Bayesian obtains a distribution over the identification region that conveys varying plausibility of values across the region; the shape of the posterior distribution may help to distinguish some values in the region as more plausible than others, in light of the data. We view this, as do Liao and Jiang (2010) and Gustafson (2010), as a possible strength of the Bayesian approach.

It is known that, unlike the point-identified case, Bayesian probability statements about partially identified parameters do not coincide, not even asymptotically, with the frequentist confidence statements. The Bayesian interval estimates are shorter than the frequentist ones, and they asymptotically lie strictly inside of the frequentist confidence intervals. However, the usual calibration property of Bayesian procedures is unaffected by the lack of identification. That is, the average frequentist coverage of the Bayesian credible set, taken with respect to the prior distribution over the parameter space, equals the nominal coverage (Gustafson and Greenland, 2009). In general, from a Bayesian perspective, frequentist confidence statements suffer from an extreme conservatism that lacks posterior probability justification.

We propose taking account of finite sample uncertainty using a Bayesian approach. We specify the likelihood as a function of partially and point identified parameters and use it to update the (non informative) prior distribution over these parameters. Alternative parametrizations are sometimes used (Richardson et al., 2010), but we view our choice as more natural for the setting considered in the paper. Details on the prior specifications and posterior inference based on MCMC methods are presented in Appendix C.

7.2 Two illustrative empirical examples

For illustrative purposes, a randomized study with noncompliance, where the ER of the random assignment has been questioned, is analyzed. The study is the National Job Corps (JC) Study, a randomized experiment performed in the mid-1990s to evaluate the effects of participation in JC (D), a large job training program for economically disadvantaged youths aged 16 to 24 years. A random sample of eligible applicants ($N = 13987$) was randomly assigned into treatment and control groups (Z), with the second group being denied access to JC for three years. Both groups were tracked at baseline, soon and at 12, 30 and 48 months after randomization. Previous works

have concentrated on global ITT effects, i.e., effects of being assigned to enroll in Job Corps (e.g., Lee, 2009; Zhang et al., 2009). However, noncompliance was present, as only 68% of those assigned to the treatment group actually enrolled in JC within 6 months from assignment. When estimating the effect on compliers, the ER for never-takers was always maintained (e.g., Frumento et al., 2011b). However being denied enrollment in JC, as opposed to deciding not to accept the offer to enroll, may, in principle, affect the labor market behavior of never-takers, especially in the short-term. For example, the denial may encourage applicants to temporarily look for alternative forms of training, possibly reducing their job search intensity. However, because they are people who are not willing to be trained when offered the opportunity, the overall amount of training that never-takers are expected to get is plausibly the same irrespective of initial assignment. As a consequence, assignment should not have any effect on long-term employment, so that the ER is more plausible for long-term labor market outcomes.

To limit exposition, here we concentrate only on short-term (52 weeks after randomization) effects on employment (Y_1) and use the long-term employment indicator (130 weeks after randomization) as a secondary outcome (Y_2); we use only observations where both outcomes and the treatment indicator are not missing (N=13193). Descriptive statistics are reported in Table 3; the two observed outcomes are strongly associated when $Z = 0$, so that we expect our bounds to be a lot tighter than the ones derived without using the secondary outcome. In Table 4 we report summaries of posterior inference for the partially identified ITT effects for compliers and never-takers, derived with and without using the secondary outcome. We also report estimates of the bounds, based on plug-in estimators, that is, using sample counterparts of observable disributions, in order to check if the posterior distribution concentrates much probability in the interior of the set, as suggested by Moon and Schorfheide (2012). As expected, due to the large sample size, 99.9% credible intervals go only a little beyond the estimated identified set. We found no evidence of the violation of the ER for never-takers: the 99% credible intervals for ITT_n are narrower when using the secondary outcome (with interval length less than half) but still cover 0. On the other hand, 99% credible intervals for ITT_c are narrower and point to a negative effect on employment for compliers of at least 0.5% points, confirming lock-in effects of those participating in the program

(van Ours, 2004; Lechner and Wunsch, 2009; Frumento et al., 2011b).

As a second illustrative example, we generate artificial data, reproducing a realistic setting, where the assumptions of latent independence could be plausible. We suppose to have data on an open-label encouragement randomized study, aimed at assessing the effects of a new drug on cholesterol levels on subjects with total cholesterol level between 200 and 239 mg/dL (borderline high) and no other risk factor for cardiovascular diseases; we assume these types of subjects were not included in previous drug experimentation studies, which typically include subjects at higher risk. The new drug is known to have some frequent side effects, such as headaches and muscle aches. These events are assumed to be very frequent even in the absence of the drug.

We suppose that half of a sample of patients was randomly encouraged to take the new drug via a prescription, while the other half was only warned of the risks related to high cholesterol levels. To simplify the example, we hypothesize that this group of subjects cannot have access to the new drug, because it would require a prescription.

We define the primary outcome as an indicator of total cholesterol above 200 mg/dL, while the secondary outcome is an indicator of headache or muscle ache events. We further assume that the two outcomes are independent conditional on compliance status and treatment assignment: the development of side effects is due to individual specific reaction to the drug, which is plausibly unrelated to the level of cholesterol. We assume that only 60% of the patients receiving a prescription actually take the drug.

While for the secondary outcome it is plausible to assume that the encouragement to take the new drug has no effect for never-takers, those who decide not to take the treatment (i.e., headache and muscle ache do not intensify just because of the encouragement, if the drug is not taken), the encouragement may have an effect on cholesterol level for never-takers: their explicit decision to avoid taking the drug may induce them to undertake other means to reduce their cholesterol level, for example changing diet or increasing physical activity.

Table 5 (first panel on the left) describes the scenario chosen as the data generating process. The scenario is characterized by a negative effect (reduction of cholesterol) of the encouragement to take the drug on cholesterol for compliers (-0.3) and for never-takers (-0.1). There is a positive

effect of the encouragement to take the drug on headaches and muscle aches for compliers (0.3), while the same effect for never-takers is 0 (ER). Two samples of size 10000 and 5000 were drawn and some descriptive statistics for the two samples are reported in Table 5. The two samples were used to draw inference on ITT effects, imposing and not imposing latent independence (LI).

Results, reported in Tables 6 and 7, show that the estimated bounds using the secondary outcome, but without imposing LI, remain rather wide. This is due to the fact that the two outcomes are independent (or very weakly associated in the two finite samples). This is also confirmed by the Bayesian analysis: without imposing LI the credible intervals derived from the univariate and bivariate analysis almost coincide.

However, when LI is imposed as an identifying restriction, inference is much sharper: the moment-based estimates, computed by substituting the observable distributions in (40)-(43) with their sample counterparts, are very close to the true parameter values; credible intervals are shorter, the posterior distributions are better behaved, the posterior mean is closer to the true parameter values and the posterior standard deviation is smaller. This behavior of the posterior distribution reflects the fact that under LI all the parameters are point identified; the shape of the posterior distribution, however, shows that in finite samples it would be dangerous to rely on standard normal asymptotic approximations to the sampling distribution of the moment-based estimators, as the shape is far from normal even in a relatively large samples.

8 Concluding remarks

We used restrictions on the joint distribution of a primary outcome and an auxiliary variable (a secondary outcome or a covariate) to derive large-sample bounds for ITT effects on the primary outcome on the subpopulations defined by compliance behavior, without requiring the ER on the primary outcome. We also provide alternative point-identifying assumptions, in the form of latent independences.

Issues of inference in finite samples were discussed, and a Bayesian solution proposed together with details on how to conduct posterior analysis in our partial and point-identified settings. This

was illustrated in two empirical examples, which highlighted the benefits of using auxiliary variables. The size of exact 99% credible intervals on ITT effects was more than halved when using a secondary outcome for which the ER is plausible.

ITT effects for noncompliers provide information on the extent of the violation of ERs: the sign of the violation is sometimes identified, and separately for never-takers and always-takers. ITT effects for compliers provide information on the possible extent of the effect of the treatment, particularly when compared with ITT effects for noncompliers.

The novel results we provide can be useful in empirical work using instrumental variables and also in other settings of broken randomized experiments, other than settings with noncompliance, where typically some local (principal strata) effects of interest can only be partially identified. Our results not only suggest new analysis with existing data, but also reveal the importance of design issues, by showing that it may be valuable to collect data on additional outcomes that may satisfy the ER, even if these secondary outcomes are not of substantive interest. We also provide guidelines on which auxiliary variables should be collected and jointly analyzed. Specifically, the stronger the association of an auxiliary variable with the compliance status and/or the primary outcome the narrower the bounds. In this regard, secondary outcomes can be particularly useful because they are expected to be highly associated with the primary outcome and compliance status.

References

[1] Abadie, A. (2003) Semiparametric instrumental variable estimation of treatment response models. Journal of Econometrics, 113, 231-263.

[2] Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica, 68, 399–405.

[3] Andrews, D. W. K., Guggenberger, P. (2009). Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. Econometric Theory, 25, 669–709.

[4] Angrist, J.D. (1990) Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records. American Economic Review, 80, 313-336.

[5] Angrist, J.D., Krueger, A.B. (1991) Does Compulsory School Attendance Affect Schooling and Earnings? The Quarterly Journal of Economics, 106, 979-1014.

[6] Angrist, J.D., Imbens, G. W., Rubin, D.B. (1996) Identification of causal effects using instrumental variables (with discussion). Journal of the American Statistical Association, 91, 444-472.

[7] Balke, A., Pearl, J. (1997) Bounds on treatment effects from studies with imperfect compliance. Journal of the American Statistical Association, 92, 1171-1176.

[8] Chernozhukov V., Hong H., Tamer E. (2007) Estimation and confidence regions for parameter sets in econometric models, Econometrica, 75, 1243-84.

[9] Chernozhukov V., Lee S., Rosen A. (2011) Intersection bounds: Estimation and Inference, mimeo, MIT, Department of Economics.

[10] Conley T., Hansen C., McCulloch R.E., Rossi P.E. (2008) A non-parametric Bayesian approach to the instrumental variable problem. Journal of Econometrics, 144, 276-305.

[11] Conley, T., Hansen, C., Rossi, P. (2012) Plausibly Exogenous. Review of Economics and Statistics, 94, 260-272.

[12] Ding P., Geng Z., Yan W., Zhou X.H. (2011) Identifiability and Estimation of Causal Effects by Principal Stratification with Outcomes Truncated by Death. Journal of the American Statistical Association, 106, 1578-1591.

[13] Duflo, E., Glennerster, R., Kremer, M. (2008) Using Randomization in Development Economics Research: a Toolkit. In T.P. Schultz and J. Strauss (Eds.) Handbook of Development Economics, Vol. 4, (3895-3962), Elsevier Science North Holland.

[14] Flores C.A., Flores-Lagunes A. (2012) Partial Identification of Local Average Treatment Effects with an Invalid Instrument. Mimeo, Department of Economics, University of Miami.

[15] Frangakis, C.E., Rubin, D.B. (1999) Addressing complications of intention-to-treat analysis in the presence of combined all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika, 86, 365-379.

[16] Frangakis, C.E., Rubin, D.B. (2002) Principal stratification in causal inference. Biometrics, 58, 191–199.

[17] Fréchet, M. (1951) Sur les tableaux de corrélation dont les marges sont donnes. Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie, 9, 53-77.

[18] Frolich, M. (2006) Nonparametric IV estimation of local average treatment effects with covariates. Journal of Econometrics, 139, 35-75.

[19] Frumento, M., Mealli, F., Pacini, B. (2011a) Causal inference with multivariate outcome. A simulation study. In Ingrassia, S., Rocci, R. and Vichi, M. (Eds.) New Perspectives in Statistical Modeling and Data Analysis, Springer.

[20] Frumento, M., Mealli, F., Pacini, B., Rubin, D.B. (2011b) Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. Journal of the American Statistical Association,107, 450-466.

[21] Gelfand, A. E., Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, 85, 398-409.

[22] Gelman, A., Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences (with discussion). Statistical Science, 7, 457-511.

[23] Gustafson, P. (2010) Bayesian inference for partially identified models, International Journal of Biostatistics, 6, issue 2, article 17.

[24] Gustafson, P. and Greenland, S. (2009) Interval estimation for messy observational data. Statistical Science 24, 328-342.

[25] Hirano, K., Imbens, G.W., Rubin, D.B., Zhou, X-H. (2000) Assessing the effect of an influenza vaccine in an encouragement design. Biostatistics, 1, 69–88.

[26] Hoogerheide, L., Kleibergen, F., van Dijk, H. (2007) Natural conjugate priors for the instrumental variables regression model applied to the Angrist-Krueger data. Journal of Econometrics, 138, 63–103.

[27] Hong, H., Nekipelov, D. (2010) Semiparametric Efficiency in Nonlinear LATE Models. Quantitative Economics, 1, 279–304.

[28] Hsu, J. C. (1996) Multiple Comparisons: Theory and Methods. London: Chapman and Hall.

[29] Huber, M., Mellace, G. (2011) Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints, mimeo.

[30] Hurwicz, L. (1950) Generalization of the Concept of Identification. In Statistical Inference in Dynamic Ecnomic Models, Cowles Comission Monograph, Vol. 10. New York: Wiley.

[31] Imai, K. (2008) Sharp bounds on causal effects in randomized experiments with truncation-by-death. Statistics and Probability Letters, 78, 144–149.

[32] Imbens, G.W., Angrist, J. (1994). Identification and estimation of local average treatment effects. Econometrica, 62, 467-476.

[33] Imbens, G.W., Manski C. (2004). Confidence Intervals for Partially Identified . Parameters. Econometrica, 74, 1845-1857.

[34] Imbens, G. W., Rubin, D. B. (1997) Bayesian inference for causal effects in randomized experiments with noncompliance. Annals of Statistics, 25, 305–327.

[35] Jo, B. (2002) Estimation of Intervention Effects with Noncompliance: Alternative Model SPecifications (with comments and rejoinder). Journal of Educational and Behavioral Statistics, 27, 385–415.

[36] Jo, B., Muthen, B. (2001) Modeling of intervention effects with noncompliance: a latent variable approach for randomized trials, in Marcoulides, G.A., Schumacker, R.E. (eds.) New developments and techniques in structrual equation modeling, 57-87, Lawrence Erlbaum Associates, Publishers. Mahwah, New Jersey.

[37] Kraay, A. (2012) Instrumental variables regressions with uncertain exclusion restrictions : a bayesian approach. Journal of Applied Econometrics, 27, 108-128.

[38] Lechner, M., Melly, B. (2010). Partial Identification of Wage Effects of Training Programs?, Brown University, Working Paper 2010-8.

[39] Lechner, M., Wunsch, C. (2007). Are training programs more effective when unemployment is high? Journal of Labor Economics, 27, 653-692.

[40] Lee, D. (2009) Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. Review of Economic Studies, 76, 1071-102 .

[41] Liao Y., Jiang W. (2010) Bayesian Analysis in Moment Inequality Models. The Annals of Statistics, 38, 275-316.

[42] Manski, C. (1990) Nonparametric bounds on treatment effects. American Economic Review: Paper and Proceedings, 80, 319-323.

[43] Manski, C. (1994) The selection problem. In C. Sims *et al.* (Eds.) Advances in Econometrics: Sixth World Congress. 143-170.

[44] Manski, C. , Pepper, J. (2000) Monotone Instrumental Variables: With an Application to the Returns to Schooling. Econometrica, 68, 997-1010.

[45] Mattei, A., Mealli, F. (2012) Augmented designs to assess principal strata direct effects. Journal of the Royal Statistical Society - Series B, 73, 729-752.

[46] Mattei A., Mealli, F., Pacini, B. (2011), Exploiting Multivariate outcomes in Bayesian inference for causal effects with noncompliance, in Studies in Theoretical and Applied Statistics, Springer, to appear.

[47] Mercatanti, A., Li, F., Mealli, F. (2012) Improving inference of normal mixtures using auxiliary variables. Mimeo.

[48] Moon, H.R., Schorfeide, F. (2012) Bayesian and frequentist inference in partially identified modelds. Econometrica, 80, 755-782.

[49] Nevo, A., Rosen, A. (2012) Identification with Imperfect Instrumental Variables. The Review of Economics and Statistics, 94, 659-671.

[50] Richardson T.S., Evans R.J., Robins J.M. (2011) Transparent Parametrizations of Models for Potential Outcomes, in Bernardo J.M., Bayarri M.J., Berger J.O., Dawid A. P., Heckerman D., Smith A.F.M. and West M. (Eds.) Bayesian Statistics, Oxford University Press.

[51] Romano, J. P. (1989). Do bootstrap confidence procedures behave well uniformly in P? The Canadian Journal of Statistics, 17, 75-80.

[52] Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. Journal of Statistical Planning and Inference, Special Issue in Honor of Ted Anderson, 138, 2786-2807.

[53] Romano, J. P. and Shaikh, A. M. (2010). Inference for the identified set in partially identified econometric models. Econometrica, 78, 169-211.

[54] Rubin, D.B. (1978) Bayesian Inference for Causal Effects: The Role of Randomization. The Annals of Statistics, 6, 34-58.

[55] Small, D. (2007) Sensitivity analysis for instrumental variables regression with overidentifying restrictions. Journal of the American Statistical Association, 102, 1049-1058.

[56] Tanner, M. , Wong, W. (1987) The calculation of posterior distributions by data augmentation (with discussion).Journal of the American Statistical Association, 82, 528-550.

[57] Tamer, E. (2010) Partial Identification in Econometrics. Annual Review of Economics, 2, 167-195.

[58] van Ours, J. (2004). The locking-in effect of subsidized jobs. Journal of Comparative Economics, 32, 37-52.

[59] Zhang, J.L., Rubin, D.B. (2003) Estimation of causal effects via principal stratification when some outcomes are truncated by death. Journal of Educational and Behavioral Statististics, 28, 353-368.

[60] Zhang, J.L., Rubin, D.B., Mealli, F. (2009) Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification. Journal of the American Statistical Association, 104, 166-176.

Appendix A

Proof of Proposition 1 Under Assumptions 1, 2 and 3, the four observable distributions are equal to:

$$
\begin{aligned}
P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 1] &= P_{a0}^{(1\cdot)}, \\
P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] &= P_{n1}^{(1\cdot)}, \\
P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0] &= \frac{\pi_c P_{c0}^{(1\cdot)} + \pi_n P_{n0}^{(1\cdot)}}{\pi_c + \pi_n}, \\
P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 1] &= \frac{\pi_c P_{c1}^{(1\cdot)} + \pi_a P_{a1}^{(1\cdot)}}{\pi_c + \pi_a}.
\end{aligned} \tag{44}
$$

Given that $0 \leq P_{c0}^{(1\cdot)}, P_{c1}^{(1\cdot)}, P_{n0}^{(1\cdot)}, P_{a1}^{(1\cdot)} \leq 1$, worst case bounds are derived. For example, the lower (upper) bound for $P_{c0}^{(1\cdot)}$ is obtained as the maximum (minimum) of 0 (1) and the value derived from (44) when $P_{n0}^{(1\cdot)} = 1$ ($P_{n0}^{(1\cdot)} = 0$):

$$
L_{P_{c0}^{(1\cdot)}} = \max\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c}, 0\right) \leq P_{c0}^{(1\cdot)} \leq \min\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, 1\right) = U_{P_{c0}^{(1\cdot)}}.
$$

Analogously, the following bounds are derived:

$$
L_{P_{c1}^{(1\cdot)}} = \max\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c} - \frac{\pi_a}{\pi_c}, 0\right) \leq P_{c1}^{(1\cdot)} \leq \min\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c}, 1\right) = U_{P_{c1}^{(1\cdot)}},
$$

$$
L_{P_{n0}^{(1\cdot)}} = \max\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_n} - \frac{\pi_c}{\pi_n}, 0\right) \leq P_{n0}^{(1\cdot)} \leq \min\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_n}, 1\right) = U_{P_{n0}^{(1\cdot)}},
$$

$$
L_{P_{a1}^{(1\cdot)}} = \max\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_a} - \frac{\pi_c}{\pi_a}, 0\right) \leq P_{a1}^{(1\cdot)} \leq \min\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_a}, 1\right) = U_{P_{a1}^{(1\cdot)}}.
$$

To simplify notation, in the expressions of the bounds we used the point identified strata proportions π_c, π_a, π_n in place of the observable distributions $P[D_i^{obs} = 1|Z_i = 1] - P[D_i^{obs} = 1|Z_i = 0]$, $P[D_i^{obs} = 1|Z_i = 0]$, $P[D_i^{obs} = 0|Z_i = 1]$, respectively.

Proof of Proposition 2 Under Assumptions 1, 2, 3 and 4, the four observable distributions are equal to:

$$
\begin{aligned}
P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 1] &= P_{a0}^{(\cdot 1)}, \\
P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0] &= P_{n1}^{(\cdot 1)}, \\
P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 0] &= \frac{\pi_c P_{c0}^{(\cdot 1)} + \pi_n P_{n0}^{(\cdot 1)}}{\pi_c + \pi_n} = \frac{\pi_c P_{c0}^{(\cdot 1)} + \pi_n P_{n1}^{(\cdot 1)}}{\pi_c + \pi_n}, \qquad (45) \\
P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 1] &= \frac{\pi_c P_{c1}^{(\cdot 1)} + \pi_a P_{a1}^{(\cdot 1)}}{\pi_c + \pi_a} = \frac{\pi_c P_{c1}^{(\cdot 1)} + \pi_a P_{a0}^{(\cdot 1)}}{\pi_c + \pi_a}, \qquad (46)
\end{aligned}
$$

where the second equalities in (45) and in (46) are due to the exclusion restrictions, so that the system can be univocally solved also for $P_{c0}^{(\cdot 1)}$ and $P_{c1}^{(\cdot 1)}$ as

$$
\begin{aligned}
P_{c0}^{(\cdot 1)} &= \frac{P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]}{\pi_c}, \\
P_{c1}^{(\cdot 1)} &= \frac{P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a) - \pi_a P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 1]}{\pi_c}.
\end{aligned}
$$

Proof of Proposition 3 In order to bound $P_{c0}^{(11)}$, $P_{c0}^{(10)}$, $P_{c1}^{(11)}$, $P_{c1}^{(10)}$, $P_{n0}^{(11)}$, $P_{n0}^{(10)}$, $P_{a1}^{(11)}$ and $P_{a1}^{(10)}$, the relevant observable joint distributions are equal to the following:

$$P[\mathbf{Y}_i^{obs} = 11|Z_i = 0, D_i^{obs} = 0] = \frac{\pi_c P_{c0}^{(11)} + \pi_n P_{n0}^{(11)}}{\pi_c + \pi_n}, \tag{47}$$

$$P[\mathbf{Y}_i^{obs} = 11|Z_i = 1, D_i^{obs} = 1] = \frac{\pi_c P_{c1}^{(11)} + \pi_a P_{a1}^{(11)}}{\pi_c + \pi_a},$$

$$P[\mathbf{Y}_i^{obs} = 10|Z_i = 0, D_i^{obs} = 0] = \frac{\pi_c P_{c0}^{(10)} + \pi_n P_{n0}^{(10)}}{\pi_c + \pi_n},$$

$$P[\mathbf{Y}_i^{obs} = 10|Z_i = 1, D_i^{obs} = 1] = \frac{\pi_c P_{c1}^{(10)} + \pi_a P_{a1}^{(10)}}{\pi_c + \pi_a}.$$

Also, the following inequalities follow from the relationship between joint and marginal distributions:

$$0 \leq P_{n0}^{(11)} \leq P_{n0}^{(\cdot 1)} = P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0], \tag{48}$$

$$0 \leq P_{c0}^{(11)} \leq P_{c0}^{(\cdot 1)} =$$

$$= \frac{P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]}{\pi_c},$$

$$0 \leq P_{a1}^{(11)} \leq P_{a1}^{(\cdot 1)} = P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 1],$$

$$0 \leq P_{c1}^{(11)} \leq P_{c1}^{(\cdot 1)} =$$

$$\frac{P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a) - \pi_a P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 1]}{\pi_c},$$

where the equalities follow from results in Proposition 2. Under these restrictions, bounds are obtained by using the equalities in (47) and substituting the maximum and minimum values of relevant quantities in (48):

$$P_{c0}^{(11)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0], 0\right) = L_{P_{c0}^{(11)}},$$

$$P_{c0}^{(11)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, \frac{P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]}{\pi_c}\right) = U_{P_{c0}^{(11)}},$$

$$P_{c0}^{(10)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[Y_{i2}^{obs} = 0|Z_i = 1, D_i^{obs} = 0], 0\right) = L_{P_{c0}^{(10)}},$$

$$P_{c0}^{(10)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_c}, \frac{P[Y_{i2}^{obs} = 0|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 0|Z_i = 1, D_i^{obs} = 0]}{\pi_c}\right) = U_{P_{c0}^{(10)}},$$

$$P_{c1}^{(11)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c} - \frac{\pi_a}{\pi_c} P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 1], 0\right) = L_{P_{c1}^{(11)}},$$

$$P_{c1}^{(11)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs} = 11|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c}, \frac{P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a) - \pi_a P[Y_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 1]}{\pi_c}\right) = U_{P_{c1}^{(11)}},$$

$$P_{c1}^{(10)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c} - \frac{\pi_a}{\pi_c} P[Y_{i2}^{obs} = 0|Z_i = 0, D_i^{obs} = 1], 0\right) = L_{P_{c1}^{(10)}},$$

$$P_{c1}^{(10)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs} = 10|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a)}{\pi_c}, \frac{P[Y_{i2}^{obs} = 0|Z_i = 1, D_i^{obs} = 1](\pi_c + \pi_a) - \pi_a P[Y_{i2}^{obs} = 0|Z_i = 0, D_i^{obs} = 1]}{\pi_c}\right) = U_{P_{c1}^{(10)}},$$

$$P_{n0}^{(11)} \geq \max\left(\frac{P[Y_i^{obs} = 11|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n)}{\pi_n} - \frac{\pi_c}{\pi_n} \frac{P[\mathbf{Y}_{i2}^{obs} = 1|Z_i = 0, D_i^{obs} = 0](\pi_c + \pi_n) - \pi_n P[Y_{i2}^{obs} = 1|Z_i = 1, D_i^{obs} = 0]}{\pi_c}, 0\right) = L_{P_{n0}^{(11)}},$$

$$P_{n0}^{(11)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs}=11|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_n}, P[Y_{i2}^{obs}=1|Z_i=1,D_i^{obs}=0]\right) = U_{P_{n0}^{(11)}},$$

$$P_{n0}^{(10)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs}=10|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_n} - \frac{\pi_c}{\pi_n}\frac{P[Y_{i2}^{obs}=0|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)-\pi_n P[Y_{i2}^{obs}=0|Z_i=1,D_i^{obs}=0]}{\pi_c}, 0\right) = L_{P_{n0}^{(10)}},$$

$$P_{n0}^{(10)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs}=10|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n)}{\pi_n}, P[Y_{i2}^{obs}=0|Z_i=1,D_i^{obs}=0]\right) = U_{P_{n0}^{(10)}},$$

$$P_{a1}^{(11)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs}=11|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a)}{\pi_a} - \frac{\pi_c}{\pi_a}\frac{P[Y_{i2}^{obs}=1|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a)-\pi_a P[Y_{i2}^{obs}=1|Z_i=0,D_i^{obs}=1]}{\pi_c}, 0\right) = L_{P_{a1}^{(11)}},$$

$$P_{a1}^{(11)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs}=11|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a)}{\pi_a}, P[Y_{i2}^{obs}=1|Z_i=0,D_i^{obs}=1]\right) = U_{P_{a1}^{(11)}},$$

$$P_{a1}^{(10)} \geq \max\left(\frac{P[\mathbf{Y}_i^{obs}=10|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a)}{\pi_a} - \frac{\pi_c}{\pi_a}\frac{P[Y_{i2}^{obs}=0|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a)-\pi_a P[Y_{i2}^{obs}=0|Z_i=0,D_i^{obs}=1]}{\pi_c}, 0\right) = L_{P_{a1}^{(10)}},$$

$$P_{a1}^{(10)} \leq \min\left(\frac{P[\mathbf{Y}_i^{obs}=10|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a)}{\pi_a}, P[Y_{i2}^{obs}=0|Z_i=0,D_i^{obs}=1]\right) = U_{P_{a1}^{(10)}}.$$

Proof of Proposition 4 Substituting (38) and (39) in the equalities in (47), we have the following system of four equations:

$$P[\mathbf{Y}_i^{obs}=11|Z_i=0,D_i^{obs}=0](\pi_c+\pi_n) = \pi_c P_{c0}^{(1\cdot)}P_{c0}^{(\cdot 1)} + \pi_n P_{n0}^{(1\cdot)}P_{n0}^{(\cdot 1)}, \tag{49}$$

$$P[\mathbf{Y}_i^{obs}=11|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a) = \pi_c P_{c1}^{(1\cdot)}P_{c1}^{(\cdot 1)} + \pi_a P_{a1}^{(1\cdot)}P_{a1}^{(\cdot 1)},$$

$$P[\mathbf{Y}_i^{obs}=10|Z_i=0,D_i^{obs}=0)](\pi_n+\pi_c) = \pi_c P_{c0}^{(1\cdot)}(1-P_{c0}^{(\cdot 1)}) + \pi_n P_{n0}^{(1\cdot)}(1-P_{n0}^{(\cdot 1)}),$$

$$P[\mathbf{Y}_i^{obs}=10|Z_i=1,D_i^{obs}=1](\pi_c+\pi_a) = \pi_c P_{c1}^{(1\cdot)}(1-P_{c1}^{(\cdot 1)}) + \pi_a P_{a1}^{(1\cdot)}(1-P_{a1}^{(\cdot 1)}).$$

Now, $P_{c0}^{(\cdot 1)}$, $P_{n0}^{(\cdot 1)}$, $P_{c1}^{(\cdot 1)}$ and $P_{a1}^{(\cdot 1)}$ are identified (see Proposition 2) so that the linear system (49) has only four unknowns $P_{c0}^{(1\cdot)}$, $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$, and $P_{a1}^{(1\cdot)}$, and can be solved, giving results in (40)-(43).

Appendix B (Web Appendix)

Deriving bounds by conditioning on X. Under Assumptions 1, 2 and 3, we show that, when the auxiliary variable is a binary covariate X, bounds in Corollary 2 can be obtained also with a conditional analysis. To this end, write for example $P_{c0}^{(1\cdot)}$ as

$$P_{c0}^{(1\cdot)} = P[Y_{i1}(0)|G_i = c, X_i = 0] \cdot P[X_i = 0|G_i = c] + P[Y_{i1}(0)|G_i = c, X_i = 1] \cdot P[X_i = 1|G_i = c] \tag{50}$$

$$= P_{c0}^{(1|0)} \cdot P[X_i = 0|G_i = c] + P_{c0}^{(1|1)} \cdot P[X_i = 1|G_i = c].$$

Let introduce the following additional notation: $P[G_i = g|X_i = x] = \pi_{g|x}, \quad g = c, n, a; \quad x = 0, 1$. These conditional strata proportions are point identified as

$$\pi_{a|1} = P[D_i^{obs} = 1|Z_i = 0, X_i = 1],$$

$$\pi_{n|1} = P[D_i^{obs} = 0|Z_i = 1, X_i = 1],$$

$$\pi_{c|1} = 1 - \pi_{a|1} - \pi_{n|1}.$$

This identification result follows from Assumption 1, because $P[G_i = a|Z_i = 0, X_i = 1] = P[G_i = a|Z_i = 1, X_i = 1] = P[G_i = a|X_i = 1]$. Bounds for the conditional quantities, $P_{c0}^{(1|1)}$ and $P_{c0}^{(0|1)}$ in (50), can be obtained applying results in Proposition 1:

$$P_{c0}^{(1|1)} \geq \max\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 1](\pi_{c|1} + \pi_{n|1})}{\pi_{c|1}} - \frac{\pi_{n|1}}{\pi_{c|1}}, 0 \right) = L_{P_{c0}^{(1|1)}},$$

$$P_{c0}^{(1|1)} \leq \min\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 1](\pi_{c|1} + \pi_{n|1})}{\pi_{c|1}}, 1 \right) = U_{P_{c0}^{(1|1)}},$$

$$P_{c0}^{(1|0)} \geq \max\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 0](\pi_{c|0} + \pi_{n|0})}{\pi_{c|0}} - \frac{\pi_{n|0}}{\pi_{c|0}}, 0 \right) = L_{P_{c0}^{(1|0)}},$$

$$P_{c0}^{(1|0)} \leq \min\left(\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 0](\pi_{c|0} + \pi_{n|0})}{\pi_{c|0}}, 1 \right) = U_{P_{c0}^{(1|0)}}.$$

In order to obtain lower and upper bounds for $P_{c0}^{(1\cdot)}$, $L_{P_{c0}^{(1|1)}}$ and $L_{P_{c0}^{(1|0)}}$, as well as $U_{P_{c0}^{(1|1)}}$ and $U_{P_{c0}^{(1|0)}}$, must be weighted by $P[X_i = 1|G_i = c]$ and $P[X_i = 0|G_i = c]$, respectively, and summed. $P[X_i = 1|G_i = c]$ and $P[X_i = 0|G_i = c]$ are identified as in (16). As an example, denote $\pi_{1|c} = P[X_i = 1|G_i = c]$ and $\pi_1 = P[X_i = 1]$; by weighting $\frac{P[Y_{i1}^{obs}=1|Z_i=0,D_i^{obs}=0,X_i=1](\pi_{c|1}+\pi_{n|1})}{\pi_{c|1}} - \frac{\pi_{n|1}}{\pi_{c|1}}$ in $L_{P_{c0}^{(1|1)}}$, we obtain the following

$$\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 1](\pi_{c|1} + \pi_{n|1})}{\pi_{c|1}} \cdot \pi_{1|c} - \frac{\pi_{n|1}}{\pi_{c|1}} \pi_{1|c} =$$

$$\frac{P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 1](\pi_{c|1} + \pi_{n|1})}{\pi_{c|1}} \cdot \left(\frac{\pi_1 \pi_{c|1}}{\pi_c} \right) - \frac{\pi_{n|1}}{\pi_{c|1}} \left(\frac{\pi_1 \pi_{c|1}}{\pi_c} \right) =$$

$$P[Y_{i1}^{obs} = 1|Z_i = 0, D_i^{obs} = 0, X_i = 1](\pi_{c|1} + \pi_{n|1}) \cdot \frac{\pi_1}{\pi_c} - \pi_{n|1} \frac{\pi_1}{\pi_c} =$$

$$\frac{P[Y_{i1}^{obs} = 1, D_i^{obs} = 0, X_i = 1|Z_i = 0]}{\pi_c} - \frac{\pi_n}{\pi_c} P[X_i = 1|Z_i = 1, D_i^{obs} = 0],$$

where the last equality follows from

$$(\pi_{c|1} + \pi_{n|1}) \cdot \pi_1 = P[G_i = c, X_i = 1] + P[G_i = n, X_i = 1] = P[D_i^{obs} = 0, X_i = 1|Z_i = 0]$$

and

$$\pi_{n|1} \cdot \pi_1 = \pi_n \cdot \pi_{1|n} = \pi_n P[X_i = 1|Z_i = 1, D_i^{obs} = 0].$$

Analogous result can be obtained by weighting $\frac{P[Y_{i1}^{obs}=1|Z_i=0,D_i^{obs}=0,X_i=0](\pi_{c|0}+\pi_{n|0})}{\pi_{c|0}} - \frac{\pi_{n|0}}{\pi_{c|0}}$ in $L_{P_{c0}^{(1|0)}}$. Consider now the weighted terms:

$$a^* = \frac{P[Y_{i1}^{obs}=1, D_i^{obs}=0, X_i=1|Z_i=0]}{\pi_c} - \frac{\pi_n}{\pi_c} P[X_i=1|Z_i=1, D_i^{obs}=0],$$

$$b^* = \frac{P[Y_{i1}^{obs}=1, D_i^{obs}=0, X_i=0|Z_i=0]}{\pi_c} - \frac{\pi_n}{\pi_c} P[X_i=0|Z_i=1, D_i^{obs}=0],$$

and define $a = \max(0, a^*)$ and $b = \max(0, b^*)$. The lower bound for $P_{c0}^{(1\cdot)}$, $L_{P_{c0}^{(1\cdot)}}$, is obtained as $(a+b)$. If $a = a^*$ and $b = b^*$, then $L_{P_{c0}^{(1\cdot)}}$ is the same lower bound in (6), derived using information only on the primary outcome. If $a^* < 0$ or $b^* < 0$ then $L_{P_{c0}^{(1\cdot)}}$ is equal to the lower bound obtained in Corollary 2. For example, if $a^* < 0$ and $b^* > 0$, then $L_{P_{c0}^{(1\cdot)}} = b^*$, and can be rewritten as

$$\frac{P[Y_{i1}^{obs}=1, X_i=0|Z_i=0, D_i^{obs}=0]P[D_i^{obs}=0|Z_i=0]}{\pi_c} - \frac{\pi_n}{\pi_c} P[X_i=0|Z_i=0, D_i^{obs}=0] =$$

$$\frac{P[Y_{i1}^{obs}=1, X_i=0|Z_i=0, D_i^{obs}=0](\pi_c+\pi_n)}{\pi_c} - \frac{\pi_n}{\pi_c} P[X_i=0|Z_i=0, D_i^{obs}=0].$$

Analogous equivalence results can be derived for $U_{P_{c0}^{(1\cdot)}}$, as well as for upper and lower bounds for $P_{c1}^{(1\cdot)}$, $P_{n0}^{(1\cdot)}$, and $P_{a1}^{(1\cdot)}$.

Deriving bounds by conditioning on Y_2. Given the equivalence result obtained above, one may argue that a conditional strategy could be used also when the auxiliary variable is a secondary outcome. Here we show why a simple stratification on the observed value of Y_2 does not yield valid results. To see this, write $P_{c0}^{(1\cdot)}$ as

$$P_{c0}^{(1\cdot)} = P[Y_{i1}(0)|G_i=c, Y_{i2}(0)=0] \cdot P[Y_{i2}(0)=0|G_i=c] + P[Y_{i1}(0)|G_i=c, Y_{i2}(0)=1] \cdot P[Y_{i2}(0)=1|G_i=c]. \tag{51}$$

In order to bound $P[Y_{i1}(0)|G_i=c, Y_{i2}(0)=0]$ and $P[Y_{i1}(0)|G_i=c, Y_{i2}(0)=1]$ we need the strata proportions conditional on $Y_{i2}(0)$. These cannot be simply identified (as it is instead the case with a covariate) from Assumption 1. We can in fact identify the following conditional strata proportions:

$$P[G_i=a|Y_{i2}(0)=1] = P[D_i^{obs}=1|Z_i=0, Y_{i2}^{obs}=1],$$

and

$$P[G_i=n|Y_{i2}(1)=1] = P[D_i^{obs}=0|Z_i=1, Y_{i2}^{obs}=1],$$

which are in general different from $P[G_i=a|Y_{i2}(1)=1]$ and $P[G_i=n|Y_{i2}(0)=1]$, respectively. The proportion of compliers also differs depending on whether it is condition on $Y_{i2}(0)=1$ or $Y_{i2}(1)=1$; the conditional proportions can be identified by exploiting the additional exclusion restriction assumption. For example, $P[G_i=c|Y_{i2}(0)=1]$ is identified as:

$$P[G_i=c|Y_{i2}(0)=1] = P[G_i=c|Y_{i2}(0)=1, Z_i=0] = \frac{P[G_i=c|Z_i=0]P(Y_{i2}(0)=1|G_i=c, Z_i=0)}{P(Y_{i2}(0)=1|Z_i=0)},$$

where $P(Y_{i2}(0)=1|G_i=c, Z_i=0)$ is identified thank to the exclusion restriction (see Proposition 2).

This result shows that it would be feasible to derive bounds on quantities that are conditional on the values of the secondary outcome. However, this would not be as straightforward as with a covariate; we cannot simply stratify on the observed values of Y_{i2}^{obs}, but analysis must be conducted separately by treatment arm, and so conditional also on Z_i. Also, working with joint probabilities for Y_{i1} and Y_{i2} (rather than conditional probabilities) makes it clear how the partial exclusion restriction helps tightening the bounds on ITT effects for the primary outcome: the ER imposes inequality constraints (as in (18) and (19)) on these joint probabilities that allow to remove values of the joint probabilities that imply marginal probabilities of the primary outcome are not admissible (not consistent with the ER on the secondary outcome).

Appendix C (Web Appendix)

Bayesian inference with noncompliance The structure of Bayesian inference for causal effects in the presence of noncompliance was first developed in Imbens and Rubin (1997). In what follows, we show how to conduct posterior analysis under the set of conditions of the two applications in the paper, i.e., assuming to have a random sample of size N where there are no always-takers and the auxiliary variable is a binary secondary outcome.

We parameterize the likelihood function of the observed data as a function of point and partially identified parameters. Specifically, we write the joint probabilities $P_{gz}^{(y_1y_2)}$ as $P_{gz}^{(.y_2)}P_{gz}^{(y_1|y_2)}$, where $P_{gz}^{(y_1|y_2)} = P[Y_{i1}(z) = y_1|G_i = g, Y_{i2}(z) = y_2]$, $g \in \{c, n\}$ and $z \in \{0, 1\}$. This parametrization allows us to easily impose the stochastic exclusion restriction on the secondary outcome, by setting $P_{n1}^{(.y_2)} = P_{n0}^{(.y_2)} = P_{n}^{(.y_2)}$. Our vector of parameters is thus $\theta = (P_{c0}^{(.1)}, P_{c0}^{(1|1)}, P_{c0}^{(1|0)}, P_{c1}^{(.1)}, P_{c1}^{(1|1)}, P_{c1}^{(1|0)}, P_{n}^{(.1)}, P_{n0}^{(1|1)}, P_{n0}^{(1|0)}, P_{n1}^{(1|1)}, P_{n1}^{(1|0)}, \pi_n)$. The parameters of primary interest are functions of θ: $ITT_g = P_{g1}^{(1.)} - P_{g0}^{(1.)} = (P_{g1}^{(1|1)}P_{g1}^{(.1)} + P_{g1}^{(1|0)}(1 - P_{g1}^{(.1)})) - (P_{g0}^{(1|1)}P_{g0}^{(.1)} + P_{g0}^{(1|0)}(1 - P_{g0}^{(.1)}))$, $g \in c, n$.

Denote with $\pi_c = 1 - \pi_n$. Let $\pi(\theta)$ be the assumed prior distribution for the parameters θ; the posterior distribution of θ is:

$$\pi(\theta|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{Z}) \propto \pi(\theta)\times \tag{52}$$

$$\prod_i \left(\pi_c P_{c1}^{(\cdot1)} P_{c1}^{(1|1)}\right)^{Z_i D_i^{obs} Y_{i1}^{obs} Y_{i2}^{obs}} \prod_i \left(\pi_c (1 - P_{c1}^{(\cdot1)}) P_{c1}^{(1|0)}\right)^{Z_i D_i^{obs} Y_{i1}^{obs} (1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_c P_{c1}^{(\cdot1)} (1 - P_{c1}^{(1|1)})\right)^{Z_i D_i^{obs} (1-Y_{i1}^{obs}) Y_{i2}^{obs}} \prod_i \left(\pi_c (1 - P_{c1}^{(\cdot1)})(1 - P_{c1}^{(1|0)})\right)^{Z_i D_i^{obs} (1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_n P_{n}^{(\cdot1)} P_{n1}^{(1|1)}\right)^{Z_i (1-D_i^{obs}) Y_{i1}^{obs} Y_{i2}^{obs}} \prod_i \left(\pi_n (1 - P_{n}^{(\cdot1)}) P_{n1}^{(1|0)}\right)^{Z_i (1-D_i^{obs}) Y_{i1}^{obs} (1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_n P_{n}^{(\cdot1)} (1 - P_{n1}^{(1|1)})\right)^{Z_i (1-D_i^{obs}) (1-Y_{i1}^{obs}) Y_{i2}^{obs}} \prod_i \left(\pi_n (1 - P_{n}^{(\cdot1)})(1 - P_{n1}^{(1|0)})\right)^{Z_i (1-D_i^{obs}) (1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_c P_{c0}^{(\cdot1)} P_{c0}^{(1|1)} + \pi_n P_{n}^{(\cdot1)} P_{n0}^{(1|1)}\right)^{(1-Z_i)(1-D_i^{obs}) Y_{i1}^{obs} Y_{i2}^{obs}}$$

$$\prod_i \left(\pi_c (1 - P_{c0}^{(\cdot1)}) P_{c0}^{(1|0)} + \pi_n (1 - P_{n}^{(\cdot1)}) P_{n0}^{(1|0)}\right)^{(1-Z_i)(1-D_i^{obs}) Y_{i1}^{obs} (1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_c P_{c0}^{(\cdot1)} (1 - P_{c0}^{(1|1)}) + \pi_n P_{n}^{(\cdot1)} (1 - P_{n0}^{(1|1)})\right)^{(1-Z_i)(1-D_i^{obs}) (1-Y_{i1}^{obs}) Y_{i2}^{obs}}$$

$$\prod_i \left(\pi_c (1 - P_{c0}^{(\cdot1)})(1 - P_{c0}^{(1|0)}) + \pi_n (1 - P_{n}^{(\cdot1)})(1 - P_{n0}^{(1|0)})\right)^{(1-Z_i)(1-D_i^{obs}) (1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}$$

where $\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{Z}$ are the N-row matrix and vectors with elements $\mathbf{Y}_i^{obs}, D_i^{obs}, Z_i$, respectively. The sum in the likelihood is because the units with $(Z_i = 0, D_i^{obs} = 0)$ are a mixture of never-takers and compliers. Direct posterior inference of θ from (52) is made easier using data augmentation to impute the missing $D_i^{mis} = D_i(1 - Z_i)$ when $Z_i = 0$. When all D_i^{mis} are imputed we essentially *observe* the compliance status G_i for every unit. Specifically, we can first obtain the joint posterior distribution of $(\theta, \mathbf{D^{mis}})$ from a Gibbs sampler by iteratively sampling from $P[\theta|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{D}^{mis}, \mathbf{Z}]$ and $P[\mathbf{D}^{mis}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{Z}, \theta]$, which in turn provides the marginal posterior distribution $\pi(\theta|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{Z})$. The key to the posterior computation is the evaluation of the complete compliance status-data posterior distribution $P[\theta|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{D}^{mis}, \mathbf{Z}]$, which has the following simple form:

$$\pi(\theta|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{D}^{mis}, \mathbf{Z}) \propto \pi(\theta)\times \tag{53}$$

$$\prod_i \left(\pi_c P_{c1}^{(\cdot1)} P_{c1}^{(1|1)}\right)^{Z_i D_i^{obs} Y_{i1}^{obs} Y_{i2}^{obs}} \prod_i \left(\pi_c (1 - P_{c1}^{(\cdot1)}) P_{c1}^{(1|0)}\right)^{Z_i D_i^{obs} Y_{i1}^{obs} (1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_c P_{c1}^{(\cdot 1)}(1-P_{c1}^{(1|1)})\right)^{Z_i D_i^{obs}(1-Y_{i1}^{obs})Y_{i2}^{obs}} \prod_i \left(\pi_c(1-P_{c1}^{(\cdot 1)})(1-P_{c1}^{(1|0)})\right)^{Z_i D_i^{obs}(1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_n P_n^{(\cdot 1)} P_{n1}^{(1|1)}\right)^{Z_i(1-D_i^{obs})Y_{i1}^{obs}Y_{i2}^{obs}} \prod_i \left(\pi_n(1-P_n^{(\cdot 1)})P_{n1}^{(1|0)}\right)^{Z_i(1-D_i^{obs})Y_{i1}^{obs}(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_n P_n^{(\cdot 1)}(1-P_{n1}^{(1|1)})\right)^{Z_i(1-D_i^{obs})(1-Y_{i1}^{obs})Y_{i2}^{obs}} \prod_i \left(\pi_n(1-P_n^{(\cdot 1)})(1-P_{n1}^{(1|0)})\right)^{Z_i(1-D_i^{obs})(1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_c P_{c0}^{(\cdot 1)} P_{c0}^{(1|1)}\right)^{(1-Z_i)(1-D_i^{obs})D_i^{mis}Y_{i1}^{obs}Y_{i2}^{obs}} \prod_i \left(\pi_n P_n^{(\cdot 1)} P_{n0}^{(1|1)}\right)^{(1-Z_i)(1-D_i^{obs})(1-D_i^{mis})Y_{i1}^{obs}Y_{i2}^{obs}}$$

$$\prod_i \left(\pi_c(1-P_{c0}^{(\cdot 1)})P_{c0}^{(1|0)}\right)^{(1-Z_i)(1-D_i^{obs})D_i^{mis}Y_{i1}^{obs}(1-Y_{i2}^{obs})} \prod_i \left(\pi_n(1-P_n^{(\cdot 1)})P_{n0}^{(1|0)}\right)^{(1-Z_i)(1-D_i^{obs})(1-D_i^{mis})Y_{i1}^{obs}(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_c P_{c0}^{(\cdot 1)}(1-P_{c0}^{(1|1)})\right)^{(1-Z_i)(1-D_i^{obs})D_i^{mis}(1-Y_{i1}^{obs})Y_{i2}^{obs}} \prod_i \left(\pi_n P_n^{(\cdot 1)}(1-P_{n0}^{(1|1)})\right)^{(1-Z_i)(1-D_i^{obs})(1-D_i^{mis})(1-Y_{i1}^{obs})Y_{i2}^{obs}}$$

$$\prod_i \left(\pi_c(1-P_{c0}^{(\cdot 1)})(1-P_{c0}^{(1|0)})\right)^{(1-Z_i)(1-D_i^{obs})D_i^{mis}(1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}$$

$$\prod_i \left(\pi_n(1-P_n^{(\cdot 1)})(1-P_{n0}^{(1|0)})\right)^{(1-Z_i)(1-D_i^{obs})(1-D_i^{mis})(1-Y_{i1}^{obs})(1-Y_{i2}^{obs})}.$$

All the parameters in θ are parameters of simple Bernoulli distributions, so that conjugate priors exist. The prior distribution of θ can be specified as the product of independent Beta(α_0, β_0).

Prior to Posterior Computation The posterior distributions of the parameters are obtained from Markov Chain Monte Carlo (MCMC) methods. The MCMC algorithm that we adopt uses Gibbs sampling (Gelfand and Smith, 1990) with data augmentation (Tanner and Wong, 1987) to impute, at each step, the missing compliance indicators D_i^{mis}, and exploits the complete compliance data posterior distribution to update the distribution of parameters.

Let $(G^{(t)}, \theta^{(t)})$ denote the state of the chain at time t. The state of the chain at time $t+1$ follows from applying the following steps.

1. The compliance status $G_i^{(t+1)}$ is drawn according to:

   for observations with $Z_i = 1$

   $$P\left[G_i^{(t+1)} = c | Z_i = 1, D_i^{obs} = 1, \mathbf{Y}_i^{obs}, \theta^{(t)}\right] = 1;$$
   $$P\left[G_i^{(t+1)} = n | Z_i = 1, D_i^{obs} = 0, \mathbf{Y}_i^{obs}, \theta^{(t)}\right] = 1.$$

   and for observations with $Z_i = 0, D_i^{obs} = 0$

   $$P\left[G_i^{(t+1)} = c | Z_i = 0, D_i^{obs} = 0, \mathbf{Y}_i^{obs}, \theta^{(t)}\right] = \frac{\pi_c^{(t)} \cdot P_{c0}^{(\cdot Y_{i2}^{obs})(t)} P_{c0}^{(Y_{i1}^{obs}|Y_{i2}^{obs})(t)}}{\pi_c^{(t)} \cdot P_{c0}^{(\cdot Y_{i2}^{obs})(t)} P_{c0}^{(Y_{i1}^{obs}|Y_{i2}^{obs})(t)} + \pi_n^{(t)} \cdot P_{n0}^{(\cdot Y_{i2}^{obs})(t)} P_{n0}^{(Y_{i1}^{obs}|Y_{i2}^{obs})(t)}}$$

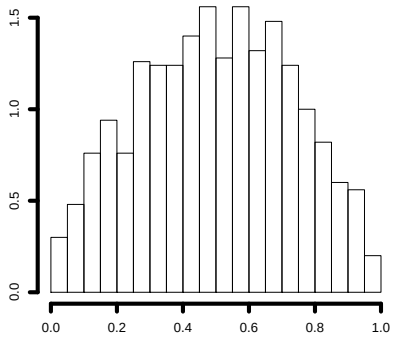
2. All the other parameters are drawn from their full conditional distribution, that is, from updated beta distributions. For example $P_{c0}^{(\cdot 1)(t+1)}$ is drawn from $Beta(\alpha, \beta)$ with $\alpha = \alpha_0 + \sum_{i=1}^{N} \mathbf{1}_{\{G_i^{(t+1)}=c\}}(1-Z_i)Y_{i2}^{obs}$ and $\beta = \beta_0 + \sum_{i=1}^{N} \mathbf{1}_{\{G_i^{(t+1)}=c\}}(1-Z_i)(1-Y_{i2}^{obs})$.

3. The compliance probability $\pi_c^{(t+1)}$ is drawn from $Beta(\alpha, \beta)$ with $\alpha = \alpha_0 + \sum_{i=1}^{N} \mathbf{1}_{\{G_i^{(t+1)}=c\}}$ and $\beta = \beta_0 + \sum_{i=1}^{N} \mathbf{1}_{\{G_i^{(t+1)}=n\}}$.

Bayesian inference in the univariate case Bayesian inference for the univariate model can be conducted for comparison purposes, assuming that the observed data are only $\mathbf{Y}_1^{obs}, \mathbf{D}^{obs}, \mathbf{Z}$. The set of parameters is reduced to $\theta = (P_{c0}^{(1.)}, P_{c1}^{(1.)}, P_{n0}^{(1.)}, P_{n1}^{(1.)}, \pi_n)$. The posterior analysis in this univariate setting was analyzed in Imbens and Rubin (1997) and Hirano et al. (2000) and the MCMC steps provided there and can be viewed as a special case of the bivariate analysis presented above. Prior distributions should be chosen consistently to those specified for the bivariate case.

Bayesian inference under latent independence Bayesian inference under latent independence can be viewed as a special case of the posterior inference we developed for the bivariate case, now the set of parameters being $\theta = (P_{c0}^{(.1)}, P_{c0}^{(1.)}, P_{c1}^{(.1)}, P_{c1}^{(1.)}, P_{n}^{(.1)}, P_{n0}^{(1.)}, P_{n1}^{(1.)}, \pi_n)$. Both the observed-data likelihood in (52) and the complete-data likelihood in (53) can be easily modified to account for latent independence.

Details of the applications In the Job Corps study, we specified the prior distribution of $\theta = (P_{c0}^{(.1)}, P_{c0}^{(1|1)}, P_{c0}^{(1|0)}, P_{c1}^{(.1)}, P_{c1}^{(1|1)}, P_{c1}^{(1|0)}, P_{n}^{(.1)}, P_{n0}^{(1|1)}, P_{n0}^{(1|0)}, P_{n1}^{(1|1)}, P_{n1}^{(1|0)}, \pi_n)$ as the product of independent non-informative Beta distributions, with $\alpha_0 = \beta_0 = 1$, equivalent to uniform distributions in $[0, 1]$. These prior distributions on θ imply that the prior distributions on $P_{gz}^{(1.)}$ are as shown in the following figure, which can be well approximated by a Beta distribution with $\alpha_0 = \beta_0 = 1.8$.



Prior distribution on $P_{gz}^{(1.)}$ implied by the non-informative prior distribution on θ

Checking-diagnostics: JC Study data

Parameter	Univariate		Bivariate	
	Mean	$\hat{R}$	Mean	$\hat{R}$
$P_{c0}^{1.}$	0.384	1.091	0.463	1.069
$P_{c1}^{1.}$	0.354	1.000	0.354	1.000
$P_{n0}^{1.}$	0.563	1.091	0.394	1.070
$P_{n1}^{1.}$	0.429	1.000	0.429	1.000
π_c	0.681	1.000	0.682	1.000

In all applications, the posterior distributions were simulated running five chains from different starting values. Each chain was run for $10,000$ iterations after a burn-in stage of $2,000$ iterations. The potential scale-reduction statistic (Gelman and Rubin, 1992) suggested good mixing of the chains for each parameter, providing no evidence against convergence (see reported table for the case of the JC study application). Inference is based on the remaining $40,000$ iterations, combining the five chains. To initiate each chain, we set the parameters at values equal to proportions in observed groups, then adding small perturbing values. Using the initialized parameters, any unknown compliance status was initialized using the data-augmentation step.

For comparison, we also conducted Bayesian inference for the univariate model. To make the prior distributions consistent to those specified in the bivariate case, priors for $P_{gz}^{(1.)}$ were chosen as Beta distributions with $\alpha_0 = \beta_0 = 1.8$, while the prior distribution for π_n was specified as a Beta distribution with $\alpha_0 = \beta_0 = 1$.

In the two examples based on the two data-sets, simulated under latent independence, the prior distribution of $\theta = (P_{c0}^{(.1)}, P_{c0}^{(1.)}, P_{c1}^{(.1)}, P_{c1}^{(1.)}, P_{n}^{(.1)}, P_{n0}^{(1.)}, P_{n1}^{(1.)}, \pi_n)$ was specified as the product of independent Beta distributions with $\alpha_0 = \beta_0 = 1$.

Tables and Figures

Table 1: Notation for the joint and marginal distributions of the potential outcomes, for $G_i = g$ and $Z_i = z$ ($g \in n, c, a$; $z \in 0, 1$)

	Y_2		
Y_1	1	0	
1	$P_{gz}^{(11)}$	$P_{gz}^{(10)}$	$P_{gz}^{(1\cdot)}$
0	$P_{gz}^{(01)}$	$P_{gz}^{(00)}$	$P_{gz}^{(0\cdot)}$
	$P_{gz}^{(\cdot 1)}$	$P_{gz}^{(\cdot 0)}$	1

Table 2: Notation for the joint and marginal distributions (revealed by the sampling process) of the observed outcomes, for $Z_i = z$ and $D_i^{obs} = d$ ($g \in n, c, a$; $z \in 0, 1$)

	Y_2		
Y_1	1	0	
1	$P[\mathbf{Y}_i^{obs} = 11 \mid Z_i = z, D_i^{obs} = d]$	$P[\mathbf{Y}_i^{obs} = 10 \mid Z_i = z, D_i^{obs} = d]$	$P[Y_{i1}^{obs} = 1 \mid Z_i = z, D_i^{obs} = d]$
0	$P[\mathbf{Y}_i^{obs} = 01 \mid Z_i = z, D_i^{obs} = d]$	$P[\mathbf{Y}_i^{obs} = 00 \mid Z_i = z, D_i^{obs} = d]$	$P[Y_{i1}^{obs} = 0 \mid Z_i = z, D_i^{obs} = d]$
	$P[Y_{i2}^{obs} = 1 \mid Z_i = z, D_i^{obs} = d]$	$P[Y_{i2}^{obs} = 0 \mid Z_i = z, D_i^{obs} = d]$	1

Table 3: Job Corps study: descriptive statistics

Observed marginal distributions

Z	D	Y_1	Y_2
0	0	0.44	0.49
1	0	0.43	0.48
1	1	0.36	0.53
π_c	0.68		
π_n	0.32		

Y_1= employment at week 52, Y_2= employment at week 130

Observed joint distribution under $Z = 0$, $\hat{\rho}_{Y_1 Y_2} = 0.74$

	Y_2		
Y_1	1	0	
1	0.40	0.04	0.44
0	0.09	0.47	0.56
	0.49	0.51	1

Table 4: Job Corps study: inference

	Estimated bounds			
	Univariate		Bivariate	
	Lower	Upper	Lower	Upper
$ITTc$	-0.293	0.174	-0.198	-0.006
$ITTn$	-0.571	0.429	-0.185	0.227

	Posterior inference - Univariate								
	Mean	SD	Median	Min	0.25%	0.5%	99.5%	99.75%	Max
$ITTc$	-0.029	0.100	-0.021	-0.324	-0.279	-0.271	0.166	0.172	0.200
$ITTn$	-0.134	0.214	-0.154	-0.579	-0.550	-0.543	0.374	0.390	0.447
	Posterior inference - Bivariate								
	Mean	SD	Median	Min	0.25%	0.5%	99.5%	99.75%	Max
$ITTc$	-0.109	0.046	-0.111	-0.236	-0.210	-0.205	-0.005	0.001	0.020
$ITTn$	0.035	0.095	0.040	-0.217	-0.181	-0.173	0.230	0.238	0.279

Table 5: Latent independence: data generating process and observed sample distributions

Data generating process:

true underlying parameter values

	Compliers		Never-Takers	
	Y_1	Y_2	Y_1	Y_2
π	0.6		0.4	
$Z = 0$	0.7	0.4	0.5	0.5
$Z = 1$	0.4	0.7	0.4	0.5
True ITT	**-0.3**	0.3	**-0.1**	0

Observed distributions

one simulated sample - N=10000

Z	D	Y_1	Y_2
0	0	0.62	0.46
1	0	0.40	0.51
1	1	0.41	0.69

Observed distributions

one simulated sample - N=5000

Z	D	Y_1	Y_2
0	0	0.63	0.47
1	0	0.39	0.57
1	1	0.39	0.70

Data generating process:

true underlying joint distributions under LI, $Z = 0$

Compliers ($\rho = 0$)

	Y_2		
Y_1	1	0	
1	0.28	0.42	0.7
0	0.12	0.18	0.3
	0.4	0.6	1

Never-Takers ($\rho = 0$)

	Y_2		
Y_1	1	0	
1	0.25	0.25	0.5
0	0.25	0.25	0.5
	0.5	0.5	1

Joint observed distribution

one simulated sample, N=10000, $Z = 0$

	Y_2		
Y_1	1	0	
1	0.28	0.34	0.62
0	0.18	0.20	0.38
	0.46	0.54	1

Joint observed distribution

one simulated sample, N=5000, $Z = 0$

	Y_2		
Y_1	1	0	
1	0.29	0.34	0.63
0	0.18	0.19	0.37
	0.47	0.53	1

Table 6: Latent independence: inference, one sample N=10000

Bounds and moment-based estimates

	Univariate bounds without LI		Bivariate bounds without LI		Moment-based estimates under LI
	Lower	Upper	Lower	Upper	
$ITTc$	−0.593	0.040	−0.593	0.040	−0.300
$ITTn$	−0.598	0.324	−0.598	0.323	−0.102

Posterior inference - Univariate

	Mean	SD	Median	Min	0.25%	0.5%	99.5%	99.75%	Max
$ITTc$	−0.278	0.117	−0.268	−0.591	−0.556	−0.548	0.003	0.014	0.052
$ITTn$	−0.134	0.171	−0.148	−0.598	−0.556	−0.541	0.262	0.272	0.319

Posterior inference - Bivariate without LI

	Mean	SD	Median	Min	0.25%	0.5%	99.5%	99.75%	Max
$ITTc$	−0.274	0.122	−0.268	−0.587	−0.563	−0.556	0.003	0.016	0.062
$ITTn$	−0.139	0.177	−0.148	−0.600	−0.556	−0.537	0.268	0.281	0.347

Posterior inference - Bivariate under LI

	Mean	SD	Median	Min	0.25%	0.5%	99.5%	99.75%	Max
$ITTc$	−0.316	0.075	−0.310	−0.559	−0.525	−0.514	−0.126	−0.108	−0.063
$ITTn$	−0.079	0.109	−0.087	−0.445	−0.378	−0.357	0.214	0.227	0.237

Table 7: Latent independence: inference, one sample N=5000

Bounds and moment-based estimates

	Univariate bounds without LI		Bivariate bounds without LI		Moment-based estimates under LI
	Lower	Upper	Lower	Upper	
$ITTc$	−0.605	0.0234	−0.605	0.0234	−0.324
$ITTn$	−0.607	0.302	−0.607	0.302	−0.105

Posterior inference - Univariate

	Mean	SD	Median	Min	2.5%	5%	95%	97.5%	Max
$ITTc$	−0.275	0.133	−0.266	−0.612	−0.523	−0.497	−0.065	−0.040	0.066
$ITTn$	−0.173	0.192	−0.188	−0.614	−0.510	−0.476	0.148	0.186	0.334

Posterior inference - Bivariate without LI

	Mean	SD	Median	Min	2.5%	5%	95%	97.5%	Max
$ITTc$	−0.247	0.128	−0.239	−0.608	−0.508	−0.480	−0.047	−0.022	0.090
$ITTn$	−0.213	0.185	−0.225	−0.621	−0.537	−0.503	0.124	0.166	0.318

Posterior inference - Bivariate under LI

	Mean	SD	Median	Min	2.5%	5%	95%	97.5%	Max
$ITTc$	−0.336	0.115	−0.329	−0.637	−0.583	−0.553	−0.158	−0.103	0.072
$ITTn$	−0.086	0.168	−0.097	−0.631	−0.422	−0.343	0.230	0.281	0.203

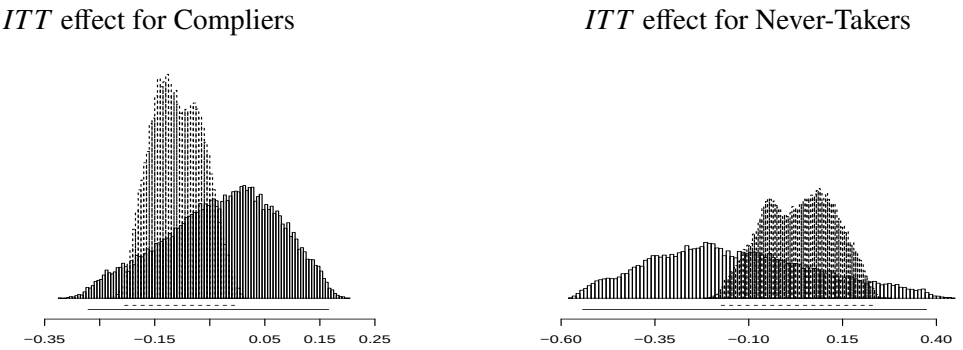


Figure 1: Univariate (continuous line) versus Bivariate (dashed line) Analysis: Histograms and 99% Posterior Intervals of ITT Effects for Compliers and Never-Takers in the JC study

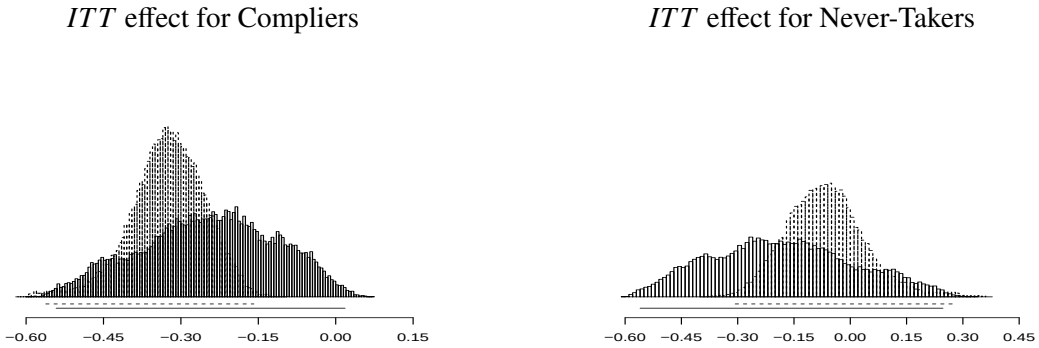


Figure 2: Bivariate Analysis w/o LI (continuous line) and under LI (dashed line): Histograms and 90% Posterior Intervals of ITT Effects for Compliers and Never-Takers, one sample N=10000

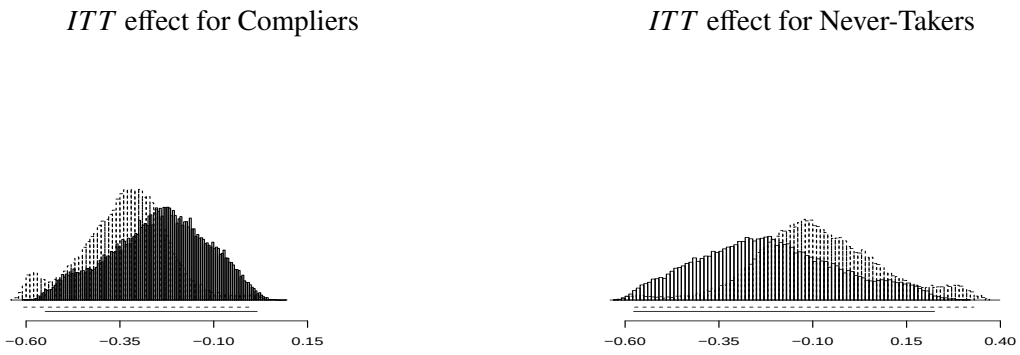


Figure 3: Bivariate Analysis w/o LI (continuous line) and under LI (dashed line): Histograms and 90% Posterior Intervals of ITT Effects for Compliers and Never-Takers, one sample N=5000