Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2011/05

Identification of causal effects in the presence of nonignorable missing outcome values

Alessandra Mattei,
Fabrizia Mealli, Barbara Pacini

Università degli Studi di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Identification of causal effects in the presence of nonignorable missing outcome values

By A. Mattei, F. Mealli
Department of Statistics, University of Florence, Viale Morgagni 59, 50134 Florence, Italy
mattei@ds.unifi.it mealli@ds.unifi.it

and B. Pacini
Department of Statistics and Mathematics, University of Pisa, Via Ridolfi 10, 56124 Pisa, Italy
barbara.pacini@sp.unipi.it

Summary

We consider a new approach to identify the causal effects of a binary treatment when the outcome is missing on a subset of units and dependence of nonresponse on the outcome cannot be ruled out even after conditioning on observed covariates. We provide sufficient conditions under which the availability of a binary instrument for nonresponse allows us to partially identify causal effects in some latent subgroups of units, defined by their nonresponse behavior in all possible combinations of treatment and instrument, named Principal Strata. Examples are provided as possible scenarios where our assumptions may be plausible; they are used to discuss the key role of the instrument for nonresponse in identifying average causal effects in presence of nonignorable missing outcomes and provide new guidelines on study designs for causal inference.

Some key words: Bounds; Causal Inference; Missing Outcomes; Principal Stratification; Instrumental Variables.

1. Introduction

In the potential outcome approach to causal inference (Rubin, 1974, 1978), a causal inference problem is viewed as a problem of missing data, where the assignment mechanism is explicitly posed as a process for revealing the observed data. The assumptions on the assignment mechanism are crucial for identifying and deriving methods to estimate causal effects. A commonly invoked identifying assumption is unconfoundedness (Rosenbaum & Rubin D.B., 1983), which usually holds by design in randomized experiments. However, even under such assumption, inference on causal effects may be invalidated due to the presence of post-treatment complications, such as noncompliance (Angrist et al., 1996), censoring *due to death* (Rubin, 1998; Lee, 2009; Zhang et al., 2009) and missing outcome values (Frangakis & Rubin, 1999). Here, we focus on identifying causal effects in the presence of missing outcome values, primarily due to nonresponse. Because nonresponse occurs after treatment assignment, respondents are not comparable by treatment status: the observed and unobserved characteristics of respondents in each treatment group are likely to differ and may be associated with the values of the missing outcome, making the missing mechanism nonignorable (e.g., Little & Rubin, 2002).

Outcome missingness is a pervasive problem in empirical studies, characterizing most of the longitudinal surveys and medical and social experiments with follow-ups[1]. Often analysts use

[1] Just to give a couple of examples, in the Tennessee Student/Teacher achievement ratio study (STAR) of 1985 the percentage of missing in reading and math scores ranged between 30% and 50% (Krueger, 1999); while in the National Job Corps Study

ad hoc procedures to handle missing data, such as dropping cases with missing observations, or sample mean substitution, which lead to valid inferences only under strong ignorability assumptions of the missing mechanism (Little & Rubin, 2002; Rubin, 1976).

A relatively recent approach to deal with post-treatment complications within the potential outcome approach is principal stratification, introduced by Frangakis & Rubin (2002). Principal stratification can be viewed as having its roots in the instrumental variables method: the approach to adjust for noncompliance applied in Angrist et al. (1996) is a special case of principal stratification, where the compliers are a principal stratum with respect to the post-treatment compliance behavior. In recent years, various identification strategies, leading to either partially or point identify the causal estimands of interest, and several applications for the concepts of principal stratification have been developed (e.g., Cheng et al. (2009); Frangakis et al. (2007); Imai (2008); Lee (2009); Zhang & Rubin (2003); Zhang et al. (2009))

In this paper, we apply principal stratification in order to develop a novel approach to deal with nonignorable missing outcome values without imposing any restriction on treatment effect heterogeneity. We rely on the presence of a binary instrument for nonresponse and provide new sufficient conditions for partial identification of causal effects for subsets of units (unions of principal strata) defined by their nonresponse behavior in all possible combinations of treatment and instrument values. The framework allows us to clarify and discuss substantive behavioral assumptions, which may differ from those required by other approaches.

The paper is organized as follows: in Section 2 principal stratification is presented and identification issues with missing outcome values are briefly described. Section 3 introduces a binary instrument for nonresponse; alternative sets of assumptions are proposed, which allow one to either partially or point identify causal estimands for specific subpopulations of units. Some examples are used to characterize these latent groups and related causal estimands. Section 4 provides some discussion and concludes.

2. PRINCIPAL STRATIFICATION AND ITS ROLE FOR CAUSAL INFERENCE

Principal stratification was first introduced by Frangakis & Rubin (2002), in order to address post-treatment complications, i.e., events which cannot be ignored when inferring on causal effects, and require adjusting for them, although conditioning on their observed values (e.g., including them in a regression model) may lead to estimating parameters which are not, in general, causal effects. We first introduce *potential outcomes* for one post-treatment variable, Y, and a binary treatment, T. If unit i in the study ($i = 1, \dots, N$) is assigned to treatment $T_i = t$ ($t = 1$ for treatment and $t = 0$ for no treatment), we denote with $Y_i(T_i = 1) = Y_i(1)$ and $Y_i(T_i = 0) = Y_i(0)$ the two potential outcomes, either of which can be observed depending on the value taken by T. A causal effect of T on Y is defined, on a single unit, as a comparison between $Y_i(1)$ and $Y_i(0)$. The fact that only two potential outcomes for each unit are defined reflects the acceptance of the stable unit treatment value assumption (SUTVA; Rubin, 1980) that there is no interference between units and that both levels of the treatment define a single outcome for each unit. We also denote with $S_i(t)$ the post-treatment potential variable, which represents a response indicator for $Y_i(t)$: the observation of $Y_i(t)$ is missing if $S_i(t) = 0$. To simplify the notation, we will drop the i subscript in the sequel.

Throughout the paper, we will maintain the assumption that treatment assignment is unconfounded given a vector X of observed pre-treatment variables and that in infinite samples treated and controls can be compared for all values of X (overlap):

(Burghardt et al., 2001) missing rate due to attrition on post-treatment occupational status varied between 12% and 30% depending on the time from the initial treatment assignment.

ASSUMPTION 1. *(Strong ignorability - Rosenbaum & Rubin, 1983)*

$$T \perp\!\!\!\perp S(0), S(1), Y(0), Y(1) \mid X \qquad \text{and} \qquad 0 < P(T = 1 \mid X) < 1.$$

Strong ignorability amounts to assuming that, within cells defined by the values of pre-treatment variables X, the treatment is randomly assigned or, at least, is assigned independently of the relevant post-treatment variables. We define

$$S^{obs} = TS(1) + (1 - T)S(0) \quad \text{and} \quad \mathrm{Y^{obs}} = \begin{cases} TY(1) + (1 - T)Y(0) & (S^{obs} = 1), \\ \text{missing} & (S^{obs} = 0). \end{cases}$$

Strong ignorability guarantees that the comparison of treated and control units with the same value of X leads to valid inference on causal effects. In general, though, we cannot compare treated with controls conditional on the observed value of S, S^{obs} (i.e., respondent treated with respondent control units), because these two groups are obtained by conditioning on different variables for units under treatment and under control, $S(1)$ and $S(0)$, respectively.

Consider now the potential response indicators $S(0)$ and $S(1)$. Within each cell defined by values of the covariates, units under study can be stratified into four latent groups, named Principal Strata, according to the joint values $(S(0), S(1))$: stratum $11 : S(1) = S(0) = 1$ comprises those who would respond under treatment and under control; stratum $10 : S(1) = 1, S(0) = 0$ comprises those who would respond under treatment but not under control; stratum $01 : S(1) = 0, S(0) = 1$ comprises those who would not respond under treatment but would respond under control; and stratum $00 : S(1) = S(0) = 0$ comprises those who would never respond regardless of treatment assignment.

This stratification of units corresponds to the *basic* principal stratification, as defined in Frangakis & Rubin (2002). More generally, a principal stratification with respect to the post-treatment variable S is a partition of the units, whose sets are unions of sets in the basic principal stratification. The principal stratum membership, $G = \{11, 10, 01, 00\}$, is not affected by treatment assignment by definition, so it only reflects characteristics of subjects, and can be regarded as a covariate, which is only partially observed in the sample (Angrist et al., 1996).

Assumption 1 implies the following: (a) $S(0), S(1) \perp\!\!\!\perp T \mid X$, so that G is guaranteed to have the same distribution in both treatment arms, within cells defined by pre-treatment variables; (b) $Y(0), Y(1) \perp\!\!\!\perp T \mid S(0), S(1), X$, so that potential outcomes are independent of the treatment given the principal strata. While it is in general improper to condition on S^{obs}, treated and control units can instead be compared conditional on a principal stratum, $(S(0), S(1))$; (c) $Y(0), Y(1)$ $\perp\!\!\!\perp T, S^{obs} \mid S(0), S(1), X$, so that, conditional on a principal stratum, comparison of respondent treated and respondent controls leads to valid inference on causal effects.

Note that, although causal effects of the treatment are well defined for the whole population, and thus for all latent groups, only in stratum 11 we can observe $Y(1)$ for some respondent units under treatment and $Y(0)$ for some other respondent units under control. On the contrary, in the other three strata we can observe the outcome only for respondents in at most one of the two treatment arms. What makes stratum 11 interesting is the fact that only in this stratum can we hope to learn something about the causal effect, even if it may not be an interesting stratum *per se*. This is a different case from noncompliance with treatment assignment or censoring due to death, such as, for instance, censoring of quality of life due to death, which can also be regarded as special cases of principal stratification. In this setting, causal effects are well defined only for specific subgroups of units which may be relevant subpopulations. For example, the group of the always survivors is the only group for which the treatment effect on quality of life is well defined, quality of life being well defined only for units who are still alive (Rubin, 1998). Although conceptually a different problem, the identification issues in estimating the effect for the stratum of the always respondents are analogous to those related to the identification of the treatment effect on the always survivors. Zhang et al. (2009) and Mealli & Pacini (2008) point identify the effect on the always survivors and the always respondent respectively, but using additional distributional assumptions. Zhang & Rubin (2003), Lee (2009), and Imai (2008) derive large sample bounds for the effect on the always survivors, but without using instruments.

Some of the assumptions that may be invoked to deal with nonresponse essentially assume that nonresponse is ignorable. The missing completely at random (MCAR) assumption states that the response probability is constant across units, thus allowing one to ignore nonresponse and use only the sample of respondents. This assumption can be tested and is usually rejected by the data. A weaker assumption is missing at random (MAR), that is, independence between the missing data mechanism and the outcome of interest, after conditioning on observed variables. MAR allows the probability of nonresponse to depend on observed but not on unobserved values, thus assuming that missing values carry no information about the probability of missingness. Unlike MCAR, MAR is not testable without auxiliary information. Both assumptions describe ignorable missing data mechanisms (Little & Rubin, 2002)[2], which are convenient because they allow us to avoid an explicit probability model for nonresponse. If the response probability depends on both observable and unobserved characteristics, then nonresponse is nonignorable.

In the econometric literature alternative ways to deal with nonresponse include instrumental variable assumptions (e.g., Manski (2003)). Plausible instrumental variables for nonresponse can be found relatively easily (unlike finding instruments for other intermediate variables): data collection characteristics, for example, are likely to affect the response probability but not the outcome values. Characteristics of the interviewer (e.g., gender), interview mode, length and design of the questionnaire can be convincing instruments for nonresponse (see, for example, Fumagalli et al. (2010); Lepkowski & Couper (2002); Nicoletti & Peracchi (2005); Nicoletti (2010); Schräpler (2004)).

We use a binary instrument for nonresponse in a causal inference framework; in this context complications arise because we have to deal simultaneously with the nonresponse behavior under treatment and under control. Recently, Chen et al. (2009) have addressed the problem of identifying causal effects in randomized experiments with noncompliance and completely nonignorable missing data, using principal stratification only to represent noncompliance, and proposing alternative nonignorable missing data models, under which the complier average causal effect is identifiable. Although similar, the setting and the framework we consider are different. First, we focus on the problem of missing outcomes, by proposing a new approach to handle missingness, which can be applied both in randomized experiment where compliance is perfect and in observational studies, both suffering from missing outcome data. Second, we account for nonignorability of the missing data mechanism by defining principal strata according to the joint values of the potential response indicators in all possible combinations of treatment and instrument for nonresponse.

[2] Ignorability requires that, in addition to MAR, the parameters of a MAR missing data process be distinct from those of the data distribution.

3. Identifying causal effect with nonignorable nonresponse on the outcome and an instrumental variable

3·1. *Principal stratification with a binary treatment and a binary instrument for nonresponse*

We assume that the distributions that are asymptotically revealed by the sampling process are known, or can be consistently estimated, thereby not taking account of specific statistical inference problems related to estimation in finite samples.

In addition to treatment T, whose causal effect on Y is still our primary interest, suppose that units are exposed to an additional treatment Z which is related to nonresponse S but unrelated to the outcome Y. Consider the following three simplified examples, as potential empirical scenarios which allow us to informally introduce and later discuss some of the identifying assumptions.

Example 1. Consider a randomized trial to assess the effects of a campaign for AIDS prevention. Let T be a binary treatment which represents the offer of free condoms. T is randomly assigned to a group of individuals at high risk of HIV infection. The post-assignment HIV infection status Y may be missing due to refusal of some patients to participate in the HIV-test; presumably non-participants are more likely to be HIV-positive than individuals who take the test[3]. The identity of nurses, Z, can be reasonably used as an instrument for nonresponse if (a) the propensity to take the HIV test varies with the nurses; (b) nurses, whose identity cannot affect the result of the test (HIV infection status), are randomly assigned to patients.

Example 2. Consider a social experiment to assess the effects of a training program on employment. Let T be a binary treatment which represents the offer to enroll in the program. T is randomly assigned to a group of target disadvantaged individuals. The post-assignment employment status Y may be missing due to refusal to respond to the follow-up interview. Presumably nonresponse is related with the employment status, e.g., unemployed individuals, especially if trained, may be less likely to declare their occupational status. Due to budget constraints, some questionnaires are administered by a phone interview, while some others with a direct interview. The mode of interview, Z, can be reasonably used as an instrument for nonresponse if (a) the propensity to respond on the employment status varies with the mode; (b) the mode of interview, which cannot affect the employment status, is randomly assigned to participants.

Example 3. Consider an observational study concerning the evaluation of firms' subsidies. Let T be a binary treatment which represents public financial assistance to firms. T is assumed strongly ignorable given a vector of pre-treatment covariate X. The outcome variable of interest is sales, Y; post-treatment questionnaires are administered via phone interviews. Not all interviewed firms respond to the question on sales, which can be potentially nonignorably missing. The interviewer's gender, Z, can be reasonably used as an instrument for nonresponse if (a) the propensity to provide information on sales varies with the interviewer's gender; (b) the interviewer's gender, which cannot affect sales, is randomly assigned to participants.

In all these examples the variable Z can be regarded as a treatment, because an intervention on it can be contemplated. The assignment of two binary treatments, T and Z, implies that four potential outcomes can be defined for each post-treatment variable, the primary outcome, Y, and the response indicator, S, in our case: $S(t, z)$, $Y(t, z)$ for $t = 0, 1$ and $z = 0, 1$. Principal strata are defined according to the joint values of $S(0, 0)$, $S(0, 1)$, $S(1, 0)$, and $S(1, 1)$. Because the response indicator is binary, the stratum membership, G, takes on 16 values (see Table 1). Unlike the case discussed in the previous Section with no instrument, there is more than one stratum from which we can hope to learn something about the causal effect of T on Y, i.e.,

[3] This example was suggested by the study in Janssens et al. (2008).

Table 1. *Principal strata with a binary treatment and a binary instrument for nonresponse*

G	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$S(0,0)$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
$S(0,1)$	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
$S(1,0)$	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
$S(1,1)$	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

all the strata where some units respond under treatment and some units respond under control ($G = 6, 7, 8, 10, 11, 12, 14, 15, 16$). In this setting, estimands of interest are causal effects for some (union) of these strata. Note that these strata include subjects who are more responsive to the instrument, i.e., are more inclined to respond if properly "encouraged".

3·2. *Basic Assumptions*

Due to the presence of two treatments, assumptions are required on the compound assignment mechanism. Both treatments are assumed randomized conditional on a set of pre-treatment covariates, so that:

ASSUMPTION 2.

$$T, Z \perp\!\!\!\perp S(0,0), S(0,1), S(1,0), S(1,1), Y(0,0), Y(0,1), Y(1,0), Y(1,1) \mid X$$

and

$$0 < Pr(T = t, Z = z \mid X) < 1 \quad (t = 0, 1; z = 0, 1);$$

where the second condition is an overlap assumption which guarantees that in large samples we can find treated and control units, as well as units with the different values of the instrument, for all values of X. For the sake of notational simplicity we will omit an explicit indication of conditioning on X in the sequel.

In order to characterize Z as an instrument, we propose the following exclusion-restriction assumption:

ASSUMPTION 3. $Y(0,0) = Y(0,1)$ and $Y(1,0) = Y(1,1)$,

which says that the value of the instrument is unrelated to the outcome. We further require that the instrument Z has some effect on S, both under treatment and under control:

ASSUMPTION 4. $E(S(0,1) - S(0,0)) \neq 0$, and $E(S(1,1) - S(1,0)) \neq 0$.

Assumption 4 warrants that there is at least one principal stratum where the response behaviour is different depending on the value of the instrument.

3·3. *Main Identification Results*

Let us now analyze how the presence of an instrument can be exploited to achieve identification of some causal estimands. Some identification assumptions can be stated as forms of monotonicity[4] of S:

[4] An alternative stronger assumption, trivially leading to identification of population average treatment effect, (and other estimands defined as comparisons of features of the marginal distributions of $Y(T = 1)$ and $Y(T = 0)$) is the following *perfect* instrument assumption: $S(0, 1) = 1$ and $S(1, 1) = 1$. This assumption can be easily falsified by the data. In our examples, having such an instrument would amount to finding persons with the same job task, all of whom provided information on the outcome variable, or finding a nurse with whom all the patients took the HIV test. In Appendix A we show how the population average treatment effect can be identified under the perfect instrument assumption.

Table 2. *Principal strata with a binary treatment and binary instrument for nonresponse under Assumptions 5 and 6*

G	1	2	4	6	8	16
S(0,0)	0	0	0	0	0	1
S(0,1)	0	0	0	1	1	1
S(1,0)	0	0	1	0	1	1
S(1,1)	0	1	1	1	1	1

ASSUMPTION 5. $S(t, 0) \leq S(t, 1) \quad \forall\, t,$

and

ASSUMPTION 6. $S(0, z) \leq S(1, z) \quad \forall\, z.$

Assumption 5 relates to the response behavior with respect to the instrument: for a fixed treatment level, units responding when $Z = 0$ would respond also when $Z = 1$. Assumption 6 relates to the response behavior with respect to the treatment: for a fixed value of the instrument, units responding under control would respond also when treated. These assumptions may often be plausible. Taking Example 2, participants in a training program usually have a higher response rate than nonparticipants. In addition, if an individual is willing to provide information on his/her occupational status in a phone interview, he/she would reasonably do so also in a direct interview. In Example 3, we may reasonably assume that exposure to the treatment, i.e., the receipt of public incentives, makes the interviewed person more responsive to administrative requests, and also that the interviewee may be more willing to provide information on sales if the interviewer is a female.

Assumptions 5 and 6 reduce the number of strata to 6 (Table 2)[5], allowing us to point identify the proportion of units who belong to the first and the last principal stratum (see Table 2), and derive large sample bounds for the other principal stratum proportions and the causal estimands of interest. Note that, under the monotonicity assumptions, the strata containing information on causal effects are strata 6, 8 and 16, so that the goal is to isolate these three strata from the remaining ones. For this purpose, identification (at least partial) of strata proportions is crucial for disentangling the mixtures and to bound the marginal distributions of $Y(T = 0)$ and $Y(T = 1)$ within strata.

We first obtain large sample bounds for the proportions in each principal stratum. Let $P_{s|t,z} = \mathrm{pr}(S^{obs} = s \mid T = t, Z = z)$, $s = 0, 1$, $t = 0, 1$ and $z = 0, 1$, be the conditional distribution of the observed response indicator given the treatment and instrument values, and define $\pi_j = \mathrm{pr}(G = j)$, $j = 1, 2, 4, 6, 8, 16$.

Note that, due to Assumption 2, $\pi_1 = \mathrm{pr}(G = 1) = \mathrm{pr}(G = 1 \mid T = 1, Z = 1)$. Also, $\mathrm{pr}(G = 1 \mid T = 1, Z = 1) = \mathrm{pr}(S(1, 1) = 0 | T = 1, Z = 1)$ (see fourth row of Table 2), and $\mathrm{pr}(S(1, 1) = 0 \mid T = 1, Z = 1) = \mathrm{pr}(S^{obs} = 0 \mid T = 1, Z = 1)$, so in large sample we have:

$$\pi_1 = 1 - P_{1|1,1}. \tag{1}$$

Analogously, in large sample the proportion of stratum $G = 16$ is the proportion of respondents within the observed group where $T = 0$ and $Z = 0$ (see first row of Table 2):

$$\pi_{16} = P_{1|0,0}. \tag{2}$$

[5] Assumption 5 together with 4 is a form of *strict* monotonicity, because inequality in 5 must hold for at least one unit; Assumption 6 is not *strict* so that the strata implied by these assumptions are at most those reported in Table 2.

Equations (1) and (2) imply that

$$\pi_2 + \pi_4 = P_{1|1,1} - P_{1|0,1} \quad (3)$$
$$\pi_2 + \pi_6 = P_{1|1,1} - P_{1|1,0} \quad (4)$$
$$\pi_4 + \pi_8 = P_{1|1,0} - P_{1|0,0} \quad (5)$$
$$\pi_6 + \pi_8 = P_{1|0,1} - P_{1|0,0} \quad (6)$$

In order for Equations (3) - (6) to hold, the differences on their right must be non negative. Note that $P_{1|1,1} - P_{1|0,1}$ and $P_{1|1,0} - P_{1|0,0}$ are the average causal effects of the treatment on the response indicator among units randomly assigned to $Z = 1$ and $Z = 0$, respectively; and $P_{1|1,1} - P_{1|1,0}$ and $P_{1|0,1} - P_{1|0,0}$ are the average causal effects of the instrument on the response indicator among units randomly assigned to the standard and active treatment, respectively. Therefore, Assumptions 5 and 6 are not falsified by the data if in large sample these causal effects are non negative.

Using Equations (1), (2), (3) and (4), and taking into account that the principal strata proportions need to add up to one ($1 = \pi_1 + \pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16}$), we have

$$\pi_4 = P_{1|1,1} - P_{1|0,1} - \pi_2 \quad (7)$$
$$\pi_6 = P_{1|1,1} - P_{1|1,0} - \pi_2 \quad (8)$$
$$\pi_8 = \pi_2 + (P_{1|1,0} - P_{1|0,0}) - (P_{1|1,1} - P_{1|0,1}) \quad (9)$$

Equations (7), (8) and (9) hold for any π_2 such that

$$\max\left\{0; (P_{1|1,1} - P_{1|0,1}) - (P_{1|1,0} - P_{1|0,0})\right\} \leq \pi_2 \leq \min\left\{P_{1|1,1} - P_{1|0,1}; P_{1|1,1} - P_{1|1,0}\right\}. \quad (10)$$

Under the previously stated assumptions (Assumptions 2 through 6), we can now derive large sample bounds on the causal effect of T for the union of strata 6, 8, and 16, which include units *reacting to the instrument* under control and/or under treatment (strata 6 and 8) and always respondents (stratum 16). For the sake of simplicity, henceforth we focus on average treatment effects. Note that the same identification strategies could be used to identify the entire outcome distribution under both values of the treatment for particular strata.

Define $E_{tz1}(Y^{obs}) = E(Y^{obs} \mid T = t, Z = z, S^{obs} = 1)$ and let $E^{\leq\alpha}_{tz1}(Y^{obs})$ and $E^{\geq\alpha}_{tz1}(Y^{obs})$ be the conditional expectations of Y^{obs} in the α ($0 < \alpha < 1$) fraction of the observed respondents ($S^{obs} = 1$) assigned to $T = t$ and $Z = z$ with the smallest and largest values of the outcome variable, Y, respectively. The following proposition is proved in Appendix A.

PROPOSITION 1. *If Assumptions 2–6 hold, then the following bounds on the average treatment effect for the union of strata 6, 8 and 16 can be derived:*

$$E^{\leq\pi_{6,8,16|111}}_{111}(Y^{obs}) - E_{011}(Y^{obs}) \leq$$
$$E(Y(T=1) - Y(T=0) \mid G \in \{6,8,16\}) \leq E^{\geq\pi_{6,8,16|111}}_{111}(Y^{obs}) - E_{011}(Y^{obs}) \quad (11)$$

where $\pi_{6,8,16|111} = \mathrm{pr}(G \in \{6,8,16\} \mid T = 1, Z = 1, S^{obs} = 1) = \frac{P_{1|0,1}}{P_{1|1,1}}$.

Under Assumptions 2 through 6, we can also derive large sample bounds on the causal effect of T separately for each of the three principal strata 6, 8, and 16, as well as bounds on the causal effects of T for the pair unions of strata 6 and 8, 6 and 16, and 8 and 16. Bounds on these six alternative causal estimands are provided in Appendix B. The choice of focusing on the causal effect for a specific subgroup of units among those belonging to principal strata either 6, or 8, or 16, rather than the causal effect for the union of the strata 6, 8, and 16, is a subject matter, and depends on the specific evaluation problem at hand. Comparing bounds for the alternative estimands may be of interest also for assessing the heterogeneity of the treatment effect across subpopulations characterized by different response behaviors. If the treatment effect is highly heterogeneous with respect to the response behavior, the average treatment causal effects within sub-unions of principal strata can provide useful information, which could be difficult to detect by focusing only on the aggregate average causal effect of the treatment for the union of principal strata 6, 8, and 16.

The sampling process allows us to identify the conditional distributions, $P_{s|t,z}$, the conditional expectations $E_{tz1}(Y_i^{\text{obs}})$, and the conditional lower and upper trimmed means $E_{tz1}^{\leq\alpha}(Y_i^{\text{obs}})$ and $E_{tz1}^{\geq\alpha}(Y_i^{\text{obs}})$, $0 < \alpha < 1$. Therefore finding estimators for the bounds defined in Proposition 1 is relatively straightforward. For instance, a moment-based estimator can be derived by replacing the means of Y and the strata proportions by their sample counterparts:

$$\widehat{P}_{s|t,z} = \frac{\sum_{i=1}^{n} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)\mathbb{1}(S_i^{obs} = s)}{\sum_i \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)}, \qquad (s = 0, 1)$$

$$\widehat{E}_{tz1}(Y_i^{obs}) = \frac{\sum_{i=1}^{n} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)S_i^{obs}Y_i^{obs}}{\sum_{i=1}^{n} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)S_i^{obs}} \equiv \overline{Y}_{tz1}$$

$$\widehat{E}_{tz1}^{\leq\alpha}(Y^{obs}) = \frac{\sum_{i=1}^{[n\alpha]} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)S_i^{obs}Y_{(i)}^{obs}}{\sum_{i=1}^{[n\alpha]} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)S_i^{obs}} \equiv \overline{Y}_{tz1}^{\leq\alpha}$$

$$\widehat{E}_{tz1}^{\geq\alpha}(Y^{obs}) = \frac{\sum_{i=n-[n\alpha]+1}^{n} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)S_i^{obs}Y_{(i)}^{obs}}{\sum_{i=n-[n\alpha]+1}^{n} \mathbb{1}(T_i = t)\mathbb{1}(Z_i = z)S_i^{obs}} \equiv \overline{Y}_{tz1}^{\leq\alpha}$$

$t, z = 0, 1$, where $\mathbb{1}(.)$ is the indicator function, $[n\alpha]$ is the largest integer not greater than $n\alpha$, and $Y_{(i)}^{obs}$, $i = 1, \ldots, n$, are the ordered statistics. In small samples, bounds can be wrapped in confidence bands to account for sampling variability in various ways (e.g., Imbens & Manski (2004)).

The benefit of using an instrument for nonresponse is due to the fact that more information can be extracted from the data about the causal effects of the treatment. Specifically, in the presence of an instrument for nonresponse, strata containing information on the causal effects are strata 6, 8 and 16, which in general include a larger proportion of units than the group of the always respondents without instrument (stratum 11, see Section 2). The bound on the average treatment effect for the always respondents, $E(Y(1) - Y(0) \mid S(1) = S(0) = 1)$, depends on the proportion of the always respondents (see, for instance, Manski (2003) and Zhang & Rubin (2003)), as well as the bound on $E(Y(1) - Y(0) \mid G \in \{6, 8, 18\})$ depends on the proportion of strata 6, 8, and 16; therefore, when the instrument is not available or is ignored, we have a loss of information. In other words, the presence of an instrument for nonresponse provides information on the causal effect also for subjects who, without the instrument, would not respond under either the standard treatment or the active treatment (i.e., principal strata 10 and 01), but would respond regardless treatment assignment when assigned to $Z = 1$. Each principal stratum defined by $(S(0), S(1))$ is split into more principal strata when an instrument for nonresponse is introduced, digging out information on a larger proportion of units. If causal effects are homogeneous, this implies using more information to estimate the same causal estimands (leading also to a better precision if the instrument is used in a parametric estimation approach). If causal effects are heterogeneous, this implies estimating an average effect for a larger proportion of units, which has higher chances to mimic the behavior of the target overall population. Therefore, when an instrument for nonresponse is available, using it might help identification and estimation of causal effects. Our discussion suggests that an instrument for nonresponse should be included as a design variable in the planning phase of the study design.

3·4. *Latent Ignorability of Nonresponse and Restrictions on the Number of Principal strata*

The bounds in Proposition 1 can be tightened if additional assumptions are introduced. Here we investigate different sets of assumptions, whose plausibility should be judged in specific empirical case studies.

A first set of additional assumptions we focus on includes *latent* ignorability (Frangakis & Rubin, 1999) of nonresponse. Latent ignorability defines a nonignorable missing data process, assuming that potential outcomes are independent of missingness conditional on variables that are only partially observed (e.g., union of principal strata). The following assumption requires latent ignorability for a union of strata, having the same response behavior when $Z = 1$:

ASSUMPTION 7. $Y(1,0) \perp\!\!\!\perp S(0,0), S(1,0) \mid S(0,1) = 1, S(1,1) = 1.$

Assumption 7 amounts to stating that the distribution of $Y(1,0)$ is the same within strata 6, 8 and 16 (see Table 2); heuristically, for units with a similar response behavior pattern, differences in response behavior can be considered random and not related to Y.

PROPOSITION 2. *If Assumptions 2–6 and 7 hold and* $\pi_2 < P_{1|1,1} - P_{1|1,0}$, *then the following bounds on the average treatment effect for the union of strata 6, 8 and 16,* $E(Y(T = 1) - Y(T = 0) \mid G \in \{6,8,16\})$, *can be derived:*

$$\max\left\{E_{111}^{\leq \pi_{6,8,16|111}}(Y^{obs}); \min_{\pi_2}\left\{\frac{\Delta - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2}\right\}\right\} - E_{011}(Y^{obs})$$
$$\leq E(Y(T = 1) - Y(T = 0) \mid G \in \{6,8,16\}) \leq \qquad (12)$$
$$\min\left\{E_{111}^{\geq \pi_{6,8,16|111}}(Y^{obs}); \max_{\pi_2}\left\{\frac{\Delta - \pi_2 E_{111}^{\leq \pi_{2|111}}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2}\right\}\right\} - E_{011}(Y^{obs})$$

where $\Delta = E_{111}(Y^{obs})(\pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16}) - E_{101}(Y^{obs})(\pi_4 + \pi_8 + \pi_{16})$, $\pi_{6,8,16|111} = \mathrm{pr}(G \in \{6,8,16\} \mid T = 1, Z = 1, S^{obs} = 1) = \frac{P_{1|0,1}}{P_{1|1,1}}$, $\pi_{2|111} = \mathrm{pr}(G = 2 \mid T = 1, Z = 1, S^{obs} = 1) = \frac{\pi_2}{P_{1|1,1}}$, $\sum_{j\in\{2,4,6,8,16\}} \pi_j = P_{1|1,1}$, $\sum_{j\in\{4,8,16\}} \pi_j = P_{1|1,0}$, *and bounds on* π_2 *are given in Equation (10).*

The proof of Proposition 2 is given in Appendix A. Note that, if π_2 were equal to $P_{1|1,1} - P_{1|1,0}$, then stratum 6 would not exist ($\pi_6 = 0$), and the average treatment effect would be for the union of strata 8 and 16. As we could expect, the bounds in Equation 12, derived under the latent ignorability Assumption 7, are not larger than those in Equation 11. However, they can be the same in some situations. Therefore, we might be interested in understanding when the latent ignorability Assumption 7 leads to strictly more informative bounds on the average treatment effect for the union of strata 6, 8, and 16, $E(Y(T = 1) - Y(T = 0) \mid G \in \{6,8,16\})$. A sufficient condition under which the bounds in Equation (12) are strictly tighter than those in Equation (11) is given in Appendix A.

It is worth to note that, if $\pi_2 = 0$, that is, if stratum 2 does not exist, the principal strata proportions π_g, $g = 1,4,6,8,16$ can be point identified from the four observed response proportions $P_{1|t,z}$, $t = 0,1$ and $z = 0,1$, and the large sample bounds in Equation (12) degenerate, leading to point identify the average causal effect of T for the union of principal strata 6, 8 and 16. Formally, the following proposition holds (see Appendix A).

PROPOSITION 3. *If Assumptions 2–6 and 7 hold and $\pi_2 = 0$, then the average treatment effect for the union of strata 6, 8 and 16, $E()Y(T = 1) - Y(T = 0)|G \in \{6, 8, 16\})$, is*

$$E(Y(T = 1) - Y(T = 0)|G \in \{6, 8, 16\}) = \tag{13}$$
$$\frac{E_{111}(Y^{obs})(\pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16}) - E_{101}(Y^{obs})(\pi_4 + \pi_8 + \pi_{16})}{\pi_6} - E_{011}(Y^{obs})$$

where $(\pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16}) = P_{1|1,1}$, $(\pi_4 + \pi_8 + \pi_{16}) = P_{1|10}$, *and* $\pi_6 = P_{1|1,1} - P_{1|1,0}$.

Nonexistence of stratum 2 is a restriction on the response behavior, and amounts to assuming that there is no unit who reacts to the instrument only if treated: if one reacts under treatment, the same must be true under control. For example, this assumption appears plausible in double blind experiments, where both doctors and patients do not know the treatment the patient is assigned to. If the doctor is able to convince the patient to take a diagnostic test, the same should plausibly be true under control.

The following proposition, proved in Appendix A, shows that point identification of the causal estimand $E(Y(T = 1) - Y(T = 0)|G \in \{6, 8, 16\})$ can be also reached under an alternative set of assumptions, which does not impose latent ignorability of nonresponse, but requires that both stratum 2 and stratum 4 do not exist: $\pi_2 = 0$ and $\pi_4 = 0$.

PROPOSITION 4. *If Assumptions 2–6 hold and $\pi_2 = 0$ and $\pi_4 = 0$, then the average treatment effect for the union of strata 6, 8 and 16, $E(Y(T = 1) - Y(T = 0)|G \in \{6, 8, 16\})$, is*

$$E(Y(T = 1) - Y(T = 0) \mid G \in \{6, 8, 16\}) = E_{111}(Y^{obs}) - E_{011}(Y^{obs}). \tag{14}$$

The assumption that both principal strata 2 and 4 do not exist ($\pi_2 = 0$ and $\pi_4 = 0$) can be easily falsified by the data. For example, if $\pi_2 = 0$ and $\pi_4 = 0$, then under Assumptions 5 and 6, one would expect to observe the same proportion of respondents among treated and controls with $Z = 1$, and a larger proportion of respondents among treated with $Z = 0$ than among controls with $Z = 0$. Note that the set of assumptions 2 through 6, along with the assumptions of nonexistence of principal strata 2 and 4, is very similar to the perfect instrument assumption because, with the exception of the stratum of the never respondents, all the other units respond when the instrument is equal to one.

4. CONCLUDING REMARKS

In this paper, we tackled the problem of identifying treatment effects when some outcome values are missing. Identification results were obtained relying on a binary instrument for nonresponse. We proposed sets of sufficient assumptions allowing identification of causal estimands for some subpopulations of units (union of principal strata) defined by the nonresponse behavior under all possible combinations of treatment and instrument values. Results were derived within the principal stratification framework, where the latent strata are generated by the primitive potential outcomes. Because our main concern was on nonparametric identification, when the assumptions are stated conditional on the covariates, we assumed to be within cells defined by them and provided moment-based estimators of identifiable causal estimands. Our result suggest that an instrument for nonresponse should be included as a design variable in the planning phase of the study design, and it should be considered in drawing causal inference in the presence of missing outcome data, whenever it is available.

The framework we considered was sufficiently rich for discussing and addressing the identification problems. However, in finite samples it would be infeasible to work within cells defined by the covariates, in particular if they are continuous: methods to accommodate covariates should be developed, which may include either flexible parametric specifications (e.g., Hirano et al. (2000)) or be semiparametric as in and Frölich (2007).

Using principal stratification, the result of inference is usually a *local* causal effect. An issue that often arises regarding the principal stratification approach is that we cannot univocally identify the group the causal effect refers to. Note, however, that the fact that proper causal effects can only be identified for latent subgroups of units is a limitation created by the missing mechanism, rather than a drawback of the framework of principal stratification. In this paper, the focus on these subgroups was primarily driven by our goal of providing valid causal effect estimates in the presence of nonignorable missing data under sets of credible assumptions. These subgroups may not be *ex ante* the most interesting ones, but the data is in general not informative about effects for other subgroups without extrapolation.

APPENDIX A

Proofs

Proof of Proposition 1. Assumption 2 guarantees that the distribution of G is the same under treatment and under control; therefore, under Assumptions 5 and 6, the proportion of units who belong to the union of principal strata 6, 8 and 16 can be identified by the observed response proportion $P_{1|01}$: $\pi_{6,8,16} = P_{1|01}$. Under control, the observable $E_{011}(Y^{obs})$ is equal to $E(Y(T=0, Z=1) \mid G \in \{6,8,16\})$ and also to $E(Y(T=0, Z=0) \mid G \in \{6,8,16\})$, by Assumption 3. Therefore, $E_{011}(Y^{obs})$ identifies $E(Y(T = 0) \mid G \in \{6,8,16\})$.
The observed group of units with $T=1$, $Z=1$ and $S^{obs}=1$, is the $\pi_{2|111}$, $\pi_{4|111}$, $\pi_{6|111}$, $\pi_{8|111}$, and $\pi_{16|111}$ mixture of the principal strata 2, 4, 6, 8, and 16. Under Assumption 2 and Assumptions 5 and 6, the conditional probability that a unit belongs to either stratum 6, or stratum 8, or stratum 16 given $T=1$, $Z=1$ and $S^{obs}=1$ is $\pi_{6,8,16|111} = P_{1|01}/P_{1|11}$. In addition, Assumption 3 implies that $E(Y(T=1) \mid G \in \{6,8,16\}) = E(Y(T=1, Z=1) \mid G \in \{6,8,16\})$. Thus, in large samples, the maximum value of $E(Y(T=1) \mid G \in \{6,8,16\})$ is the expected value of Y for the $\pi_{6,8,16|111}$ fraction of largest values of Y for units in the observed group with $T=1$, $Z=1$ and $S^{obs}=1$. Analogously, the minimum value of $E(Y(T=1) \mid G \in \{6,8,16\})$ is the expected value of Y for the $\pi_{6,8,16|111}$ fraction of smallest values of Y for units in the same observed group. Formally, $E_{111}^{\leq \pi_{6,8,16|111}}(Y^{obs}) \leq E(Y(T = 1) \mid G \in \{6,8,16\}) \leq E_{111}^{\geq \pi_{6,8,16|111}}(Y^{obs})$. Therefore, the average treatment effect for the union of strata 6, 8 and 16, $E(Y(T=1) - Y(T=0) \mid G \in \{6,8,16\})$, is at least $E_{111}^{\leq \pi_{6,8,16|111}}(Y^{obs}) - E_{011}(Y^{obs})$, and at most $E_{111}^{\geq \pi_{6,8,16|111}}(Y^{obs}) - E_{011}(Y^{obs})$. □

Proof of Proposition 2. As in proof of Proposition 1, Assumptions 2–6 imply that under control, the expected value $E(Y(T=0) \mid G \in \{6,8,16\})$ can be identified by $E_{011}(Y^{obs})$. In addition, bounds on the estimand $E(Y(T=1) - Y(T=0) \mid G \in \{6,8,16\})$ in Equation (11) derived in Proposition 1 still hold.
On one other hand, under treatment, the observable mean $E_{101}(Y^{obs})$ is equal to

$$\frac{E(Y(T=1,Z=0) \mid G=4)\pi_4 + E(Y(T=1,Z=0) \mid G=8)\pi_8 + E(Y(T=1,Z=0) \mid G=16)\pi_{16}}{\pi_4 + \pi_8 + \pi_{16}}$$

and the observable mean $E_{111}(Y^{obs})$ is equal to

$$\frac{\sum_{j=2,4,6,8,16} E(Y(T=1,Z=1) \mid G=j)\pi_j}{\pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16}}.$$

By Assumption 3, $E(Y(T=1, Z=0) \mid G=j) = E(Y(T=1, Z=1) \mid G=j)$, $j = 2,4,6,8,16$. In addition, Assumption 7 implies that $E(Y(T=1, Z=0) \mid G=6) = E(Y(T=1, Z=0) \mid G=8) =$

$E(Y(T=1,Z=0) \mid G=16)$. Therefore,

$$\frac{\sum_{j=2,4,6,8,16} E(Y(T=1,Z=1) \mid G=j)\pi_j}{\pi_2+\pi_4+\pi_6+\pi_8+\pi_{16}} = \frac{\sum_{j=2,4} E(Y(T=1,Z=1) \mid G=j)\pi_j + E(Y(T=1,Z=1) \mid G=8)(\pi_6+\pi_8+\pi_{16})}{\pi_2+\pi_4+\pi_6+\pi_8+\pi_{16}}.$$

Given a fixed value of π_2, $E(Y(T=1,Z=1) \mid G=8)$ is equal to

$$\frac{E_{111}(Y^{obs})(\pi_2+\pi_4+\pi_6+\pi_8+\pi_{16}) - E_{101}(Y^{obs})(\pi_4+\pi_8+\pi_{16}) - E(Y(T=1,Z=1) \mid G=2)\pi_2}{\pi_6},$$

where $E(Y(T=1,Z=1) \mid G=2) = E(Y(T=1) \mid G=2)$ by Assumption 3, and $\pi_6 = P_{1|1,1} - P_{1|1,0} - \pi_2$, by Assumptions 2, 5 and 6. Using a reasoning as that in proof of Proposition 1, the following bounds on the expected value of the outcome Y under treatment in stratum 2, $E(Y(T=1) \mid G=2)$, can be derived: $E_{111}^{\leq \pi_{2|111}}(Y^{obs}) \leq E(Y(T=1) \mid G=2) \leq E_{111}^{\geq \pi_{2|111}}(Y^{obs})$. Define $\Delta = E_{111}(Y^{obs})(\pi_2 + \pi_4+\pi_6+\pi_8+\pi_{16}) - E_{101}(Y^{obs})(\pi_4+\pi_8+\pi_{16})$, then

$$\frac{\Delta - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{P_{1|1,1}-P_{1|1,0}-\pi_2} \leq E(Y(T=1) \mid G \in \{6,8,16\}) \leq \frac{\Delta - \pi_2 E_{111}^{\leq \pi_{2|111}}(Y^{obs})}{P_{1|1,1}-P_{1|1,0}-\pi_2}.$$

Therefore, given a fixed value of π_2, the average causal effect of T for the union of strata 6, 8 and 16 is at least $\frac{\Delta - E_{111}^{\geq \pi_{2|111}}(Y^{obs})\pi_2}{P_{1|1,1}-P_{1|1,0}-\pi_2} - E_{011}(Y^{obs})$ and at most $\frac{\Delta - E_{111}^{\leq \pi_{2|111}}(Y^{obs})\pi_2}{P_{1|1,1}-P_{1|1,0}-\pi_2} - E_{011}(Y^{obs})$. Minimizing (maximizing), the lower (upper) bound over the possible range of π_2 gives the following bounds on $E(Y(T=1)-Y(T=0) \mid G \in \{6,8,16\})$:

$$\min_{\pi_2}\left\{\frac{\Delta - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{P_{1|1,1}-P_{1|1,0}-\pi_2}\right\} - E_{011}(Y^{obs})$$
$$\leq E(Y(T=1)-Y(T=0) \mid G \in \{6,8,16\}) \leq$$
$$\max_{\pi_2}\left\{\frac{\Delta - \pi_2 E_{111}^{\leq \pi_{2|111}}(Y^{obs})}{P_{1|1,1}-P_{1|1,0}-\pi_2}\right\} - E_{011}(Y^{obs}).$$

Combining these bounds with those in Equation (11) derived in Proposition 1, we have the desired bounds on $E(Y(T=1)-Y(T=0) \mid G \in \{6,8,16\})$ in Equation (12). □

Proof of Proposition 3. Due to Assumptions 5 and 6 and the assumption of nonexistence of stratum 2, at most 5 strata ($G = 1,4,6,8,16$) exist (see Table 2). There are thus 4 free of variation strata proportions to be identified (π_4, π_6, π_8, π_{16}). Under Assumption 2 (which guarantees that the distribution of G is the same under treatment and under control), these proportions can be identified as linear combination of the 4 observable response proportions: $\pi_{16} = P_{1|0,0}$, $\pi_6+\pi_8+\pi_{16} = P_{1|0,1}$, $\pi_4+\pi_8+\pi_{16} = P_{1|1,0}$, $\pi_4+\pi_6+\pi_8+\pi_{16} = P_{1|1,1}$.
As in proof of Proposition 1, Assumptions 2–6 imply that under control, the expected value $E(Y(T=0) \mid G \in \{6,8,16\})$ can be identified by $E_{011}(Y^{obs})$.
Under treatment, the observable mean $E_{101}(Y^{obs})$ is equal to

$$\frac{E(Y(T=1,Z=0) \mid G=4)\pi_4 + E(Y(T=1,Z=0) \mid G=8)\pi_8 + E(Y(T=1,Z=0) \mid G=16)\pi_{16}}{\pi_4+\pi_8+\pi_{16}}$$

and the observable mean $E_{111}(Y^{obs})$ is equal to

$$\frac{\sum_{j=4,6,8,16} E(Y(T=1,Z=1) \mid G=j)\pi_j}{\pi_4+\pi_6+\pi_8+\pi_{16}}.$$

By Assumption 3, $E(Y(T=1,Z=1) \mid G=j)$ is equal to $E(Y(T=1,Z=0) \mid G=j)$, $j = 4,6,8,16$. In addition by Assumption 7, $E(Y(T=1,Z=0) \mid G=6) = E(Y(T=1,Z=0) \mid G=$

$8) = E(Y(T=1, Z=0) \mid G=16)$, so that

$$\frac{\sum_{j=4,6,8,16} E(Y(T=1, Z=1) \mid G=j)\pi_j}{\pi_4+\pi_6+\pi_8+\pi_{16}}$$

is equal to

$$\frac{E(Y(T=1, Z=1) \mid G=4)\pi_4 + E(Y(T=1, Z=1) \mid G=8)(\pi_6+\pi_8+\pi_{16}}{\pi_4+\pi_6+\pi_8+\pi_{16}}.$$

By difference, we can now identify $E(Y(T=1, Z=1) \mid G=8)$ as

$$\frac{E_{111}(Y^{obs})(\pi_4+\pi_6+\pi_8+\pi_{16}) - E_{101}(Y^{obs})(\pi_4+\pi_8+\pi_{16})}{\pi_6}.$$

Therefore, the average causal effect $E(Y(T=1) - Y(T=0) \mid G \in \{6, 8, 16\})$ is identified as

$$\frac{E_{111}(Y^{obs})(\pi_4+\pi_6+\pi_8+\pi_{16}) - E_{101}(Y^{obs})(\pi_4+\pi_8+\pi_{16})}{\pi_6} - E_{011}(Y^{obs}). \qquad \square$$

Proof of Proposition 4. Due to Assumptions 5 and 6 and the assumptions of nonexistence of strata 2 and 4, at most 4 strata ($G = 1, 6, 8, 16$) exist. Respondents with $Z = 1$ under treatment and under control are union of the same strata 6, 8 and 16. Due to Assumption 3, we trivially identify $E(Y(T=1) - Y(T=0) \mid G \in \{6, 8, 16\})$ as $E_{111}(Y^{obs}) - E_{011}(Y^{obs})$. $\square$

The perfect instrument assumption

PROPOSITION 5. *Under Assumptions 2 and 3, and the perfect instrument assumption, the population average treatment effect can be identified as*

$$E(Y(T=1) - Y(T=0)) = E(Y(T=1, Z=1) - Y(T=0, Z=1)).$$

Proof. The perfect instrument assumption removes some of the strata ($1-5$,7,$9-13$,15). Those units with $Z = 1$ and $S^{obs} = 1$ in the treatment and control group are comparable since now the strata in both groups are the same and Assumption 2 guarantees that strata distribution is the same in both treatment arms. Assumption 3 implies that: $F(Y(0, 1)) = F(Y(0, 0)) = F(Y(T=0))$ and $F(Y(1, 0)) = F(Y(1, 1)) = F(Y(T=1))$, where $F(U)$ denotes the generic cdf of U. The population average treatment effect (as well as other estimands defined as comparisons of features of the marginal distributions of $Y(T=1)$ and $Y(T=0)$) can be identified as $E(Y(T=1) - Y(T=0)) = E(Y(T=1, Z=1) - Y(T=0, Z=1))$.

This result is related to Manski (2003, pp. 30-31) and Frangakis et al. (2007) where the use of a *key variable*, with the same features as our Z, has been exploited within a principal stratification design, in order to draw valid inference for the marginal distribution of some input data nonignorably missing (missing due to death).

A sufficient condition under which bounds on the average treatment effect for the union of strata 6, 8, and 16 in Equation (11) are strictly tighter than those in Equation (12).

The width of the bounds in Equation (11) is

$$w = E_{111}^{\geq \pi_{6,8,16|111}}(Y^{obs}) - E_{111}^{\leq \pi_{6,8,16|111}}(Y^{obs}),$$

and the width of the bounds in Equation (12), derived under the latent ignorability Assumption 7, is

$$w_{LI} = \min\left\{ w; \max_{\pi_2}\left\{ \frac{\Delta - \pi_2 E_{111}^{\leq \pi_{2|111}}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2} \right\} - \min_{\pi_2}\left\{ \frac{\Delta - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2} \right\} \right\},$$

which depends on π_2. Therefore, the latent ignorability Assumption 7 allows us to tighten the bounds on the causal effect of the treatment for the union of principal strata 6, 8, and 16 if the observable proportions

$P_{s|t,z}$, $s, = 0, 1$, $z, t = 0, 1$, lead to bounds on π_2 such that $w_{LI} < w$. Define

$$\pi_2^m = \arg\min_{\pi_2} \frac{\Delta - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2} \qquad \text{and} \qquad \pi_2^M = \arg\max_{\pi_2} \frac{\Delta - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2}.$$

Then,

$$w_{LI} = \min\left\{ w; \frac{\Delta - \pi_2^M E_{111}^{\leq \pi_{2|111}^M}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2^M} - \frac{\Delta - \pi_2^m E_{111}^{\geq \pi_{2|111}^m}(Y^{obs})}{P_{1|1,1} - P_{1|1,0} - \pi_2^m} \right\}.$$

Therefore, a sufficient condition for $w_{LI} < w$ is that

$$\pi_2^M \left(E_{111}^{\geq \pi_{6,8,16|111}}(Y^{obs}) - E_{111}^{\leq \pi_{2|111}^M}(Y^{obs}) \right) \leq \left(P_{1|1,1} - P_{1|1,0}\right) E_{111}^{\geq \pi_{6,8,16|111}}(Y^{obs}) - \Delta$$

$$\pi_2^m \left(E_{111}^{\geq \pi_{2|111}^m}(Y^{obs}) - E_{111}^{\leq \pi_{6,8,16|111}}(Y^{obs}) \right) \leq \Delta - \left(P_{1|1,1} - P_{1|1,0}\right) E_{111}^{\leq \pi_{6,8,16|111}}(Y^{obs}),$$

with at least one strict inequality.

APPENDIX B

Supplementary Material

PROPOSITION 6. *If Assumptions 2–6 hold, then, in addition to the bounds on $E(Y(T = 1) - Y(T = 0)|G \in \{6, 8, 16\})$ given in Equation (11) in Proposition 1, the following bounds can be derived:*

Bounds on the average treatment effect in stratum 6

$$\min_{\pi_2} \left\{ \max\left\{ E_{111}^{\leq \pi_{6|111}}(Y^{obs}); \frac{\Delta_1^6 - \pi_2 E_{111}^{\geq \pi_{2|111}}(Y^{obs})}{\pi_6} \right\} - \min\left\{ E_{011}^{\geq \pi_{6|011}}(Y^{obs}); \frac{\Delta_0^6 - \pi_8 E_{011}^{\leq \pi_{8|011}}(Y^{obs})}{\pi_6} \right\} \right\}$$
$$\leq E(Y(T = 1) - Y(T = 0) \mid G = 6) \leq \tag{A1}$$
$$\max_{\pi_2} \left\{ \min\left\{ E_{111}^{\geq \pi_{6|111}}(Y^{obs}); \frac{\Delta_1^6 - \pi_2 E_{111}^{\leq \pi_{2|111}}(Y^{obs})}{\pi_6} \right\} - \max\left\{ E_{011}^{\leq \pi_{6|011}}(Y^{obs}); \frac{\Delta_0^6 - \pi_8 E_{011}^{\geq \pi_{8|011}}(Y^{obs})}{\pi_6} \right\} \right\}$$

where

$$\Delta_0^6 = E_{011}(Y^{obs})(\pi_6 + \pi_8 + \pi_{16}) - E_{001}(Y^{obs})\pi_{16}$$
$$\Delta_1^6 = E_{111}(Y^{obs})(\pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16}) - E_{101}(Y^{obs})(\pi_4 + \pi_8 + \pi_{16})$$
$$\pi_{6|011} = \text{pr}(G = 6 \mid T = 0, Z = 1, S^{obs} = 1) = \frac{(P_{1|1,1} - P_{1|1,0} + \pi_2)}{P_{1|0,1}}$$

$$\pi_{6|111} = \text{pr}(G = 6 \mid T = 1, Z = 1, S^{obs} = 1) = 1 - \frac{P_{1|1,0} + \pi_2}{P_{1|1,1}}$$
$$\pi_{2|111} = \text{pr}(G = 2 \mid T = 1, Z = 1, S^{obs} = 1) = \frac{\pi_2}{P_{1|1,1}}$$
$$\pi_{8|011} = \text{pr}(G = 8 \mid T = 0, Z = 1, S^{obs} = 1) = \frac{(P_{1|1,0} - P_{1|0,0}) - (P_{1|1,1} - P_{1|0,1}) + \pi_2}{P_{1|0,1}}$$

$\sum_{j=2,4,6,8,16} \pi_j = P_{1|1,1}$, $\sum_{j=4,8,16} \pi_j = P_{1|10}$, $\sum_{j=6,8,16} \pi_j = P_{1|0,1}$, π_6, π_8 *and* π_{16} *are given in Equations (8), (9) and (2), respectively, and bounds for* π_2 *are given in Equation (10).*

Bounds on the average treatment effect in stratum 8

$$\min_{\pi_2} \left\{ \max\left\{ E_{101}^{\leq \pi_{8|101}}(Y^{obs}); E_{111}^{\leq \pi_{8|111}}(Y^{obs}) \right\} - \min\left\{ E_{011}^{\geq \pi_{8|011}}(Y^{obs}); \frac{\Delta_0^8 - \pi_6 E_{011}^{\leq \pi_{6|011}}(Y^{obs})}{\pi_8} \right\} \right\}$$
$$\leq E(Y(T = 1) - Y(T = 0) \mid G = 8) \leq \tag{A2}$$
$$\max_{\pi_2} \left\{ \min\left\{ E_{101}^{\geq \pi_{8|101}}(Y^{obs}); E_{111}^{\geq \pi_{8|111}}(Y^{obs}) \right\} - \max\left\{ E_{011}^{\leq \pi_{8|011}}(Y^{obs}); \frac{\Delta_0^8 - \pi_6 E_{011}^{\geq \pi_{6|011}}(Y^{obs})}{\pi_8} \right\} \right\}$$

where

$\Delta_0^8 = E_{011}(Y^{obs})(\pi_6 + \pi_8 + \pi_{16}) - E_{001}(Y^{obs})\pi_{16}$
$\pi_{8|tz1} = Pr(G = 8 \mid T = t, Z = z, S^{obs} = 1) = \frac{\pi_2 + (P_{1|1,0} - P_{1|0,0}) - (P_{1|1,1} - P_{1|0,1})}{P_{1|t,z}} \quad (t = 0, 1; z = 0, 1)$
$\pi_{6|011} = Pr(G = 6 \mid T = 1, Z = 1, S^{obs} = 1) = \frac{P_{1|1,1} - P_{1|1,0} - \pi_2}{P_{1|0,1}},$

and $\sum_{j=6,8,16} \pi_j = P_{1|0,1}$, π_6, π_8 *and* π_{16} *are given in Equations (8), (9) and (2), respectively, and bounds for* π_2 *are given in Equation (10).*

Bounds on the average treatment effect in stratum 16

$$\max\left\{E_{101}^{\leq \pi_{16|101}}(Y^{obs}); E_{111}^{\leq \pi_{16|111}}(Y^{obs})\right\} - E_{001}(Y^{obs}) \leq$$
$$E(Y(T=1) - Y(T=0) \mid G = 16) \leq \min\left\{E_{101}^{\geq \pi_{16|101}}(Y^{obs}); E_{111}^{\geq \pi_{16|111}}(Y^{obs})\right\} - E_{001}(Y^{obs}) \quad \text{(A3)}$$

where $\pi_{16|1z1} = \mathrm{pr}(G = 16 \mid T = 1, Z = z, S^{obs} = 1) = \frac{P_{1|0,0}}{P_{1|1,z}}$, $z = 0, 1$.

Bounds on the average treatment effect for the union of strata 6 and 8

$$\begin{aligned} &E_{111}^{\leq \pi_{6,8|111}}(Y^{obs}) - \frac{E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16}}{\pi_6+\pi_8} \\ &\quad \leq E(Y(T=1) - Y(T=0) \mid G \in \{6,8\}) \leq \\ &E_{111}^{\geq \pi_{6,8|111}}(Y^{obs}) - \frac{E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16}}{\pi_6+\pi_8} \end{aligned} \quad \text{(A4)}$$

where $\pi_{6,8|111} = \mathrm{pr}(G \in \{6,8\} \mid T = 1, Z = 1, S^{obs} = 1) = \frac{P_{1|0,1} - P_{1|0,0}}{P_{1|1,1}}$, $\pi_6 + \pi_8 + \pi_{16} = P_{1|0,1}$, $\pi_6 + \pi_8 = P_{1|0,} - P_{1|0,0}$ *and* π_{16} *is given in Equation (2).*

Bounds on the average treatment effect for the union of strata 6 and 16

$$\min_{\pi_2}\left\{E_{111}^{\leq \pi_{6,16|111}}(Y^{obs}) - E_{011}^{\geq \pi_{6,16|011}}(Y^{obs})\right\} \leq$$
$$E(Y(T=1) - Y(T=0) \mid G \in \{6,16\}) \leq \max_{\pi_2}\left\{E_{111}^{\geq \pi_{6,16|111}}(Y^{obs}) - E_{011}^{\leq \pi_{6,16|011}}(Y^{obs})\right\} \quad \text{(A5)}$$

where $\pi_{6,16|011} = \mathrm{pr}(G \in \{6,16\} \mid T = 0, Z = 1, S^{obs} = 1) = \frac{P_{1|1,1} - P_{1|1,0} - \pi_2 + P_{1|0,0}}{P_{1|0,1}}$, *and* $\pi_{6,16|111} = \mathrm{pr}(G \in \{6,16\} \mid T = 1, Z = 1, S^{obs} = 1) = 1 - \frac{P_{1|1,0} - P_{1|0,0} + \pi_2}{P_{1|1,1}}$.

Bounds on the average treatment effect for the union of strata 8 and 16

$$\begin{aligned} &\min_{\pi_2}\left\{\max\left\{E_{101}^{\leq \pi_{8,16|101}}(Y^{obs}); E_{111}^{\leq \pi_{8,16|111}}(Y^{obs})\right\} - E_{011}^{\geq \pi_{8,16|011}}(Y^{obs})\right\} \\ &\quad \leq E(Y(T=1) - Y(T=0) \mid G \in \{8,16\}) \leq \\ &\max_{\pi_2}\left\{\min\left\{E_{101}^{\geq \pi_{8,16|101}}(Y^{obs}); E_{111}^{\geq \pi_{8,16|111}}(Y^{obs})\right\} - E_{011}^{\leq \pi_{8,16|011}}(Y^{obs})\right\} \end{aligned} \quad \text{(A6)}$$

where $\pi_{8,16|tz1} = \mathrm{pr}(G \in \{8,16\} \mid T = t, Z = z, S^{obs} = 1) = \frac{\pi_2 + P_{1|1,0} - (P_{1|1,1} - P_{1|0,1})}{P_{1|t,z}}$, $t = 0, 1$, $z = 0, 1$.

Proof. Assumption 2 guarantees that the distribution of G is the same under treatment and under control; therefore, under Assumptions 5 and 6, the principal strata proportions π_1 and π_{16} can be identified by the observed response proportions $P_{1|10}$ and $P_{1|0,0}$, as shown in Equations (1) and (2). In large sample, under Assumptions 2, 5 and 6, we also have $\pi_6 + \pi_8 + \pi_{16} = P_{1|0,1}$, $\pi_4 + \pi_8 + \pi_{16} = P_{1|1,0}$, and $\pi_2 + \pi_4 + \pi_6 + \pi_8 + \pi_{16} = P_{1|1,1}$. These linear combinations of the six observable response proportions imply Equations (7)–(9). Using Equations (1) through (9), we can partially identified π_2 as in Equation (10) and the remaining principal strata proportions, π_j, $j = 4, 6, 8$.

Let $\pi_{g|tz1}$ denote the conditional probability that a unit belongs to the (union of) principal strata g, $g \subseteq \{1, 2, 4, 6, 8, 16\}$, given that the unit belongs to the observed group with $T = t$, $Z = z$ ($t, z = 0, 1$), and $S_i^{\text{obs}} = 1$:

$$\pi_{g|tz1} = \text{pr}(G_i \in g \mid T = t, Z = z, S_i^{\text{obs}} = 1) = \frac{\text{pr}(G_i \in g)}{P_{1|t,z}}.$$

These conditional probabilities cannot be point identified, except for $g = \{1\}$, $g = \{16\}$, $g = \{6, 8, 16\}$, $g = \{4, 8, 16\}$, and $g = \{2, 4, 6, 8, 16\}$. However large sample bounds on them can be easily derived using Equations (7)–(9) and Equation (10).

Bounds on the average treatment effect in stratum 6. Assume that the value of π_2 is known. Under control, $E_{011}(Y^{obs})$ is equal to

$$\frac{E(Y(T=0,Z=1)\,|\,G=6)\pi_6 + E(Y(T=0,Z=1)\,|\,G=8)\pi_8 + E(Y(T=0,Z=1)\,|\,G=16)\pi_{16}}{\pi_6+\pi_8+\pi_{16}},$$

and $E(Y(T=0,Z=1) \mid G=j) = E(Y(T=0,Z=0) \mid G=j)$, $j = 6, 8, 16$, by Assumption 3. Therefore, $E(Y(T=0) \mid G=6)$ can be bounded from below (above) by the expected value of Y for the $\pi_{6|011}$ fraction of smallest (largest) values of Y for units in the observed group with $T=0$, $Z=1$ and $S^{obs}=1$: $E_{011}^{\leq\pi_{6|011}}(Y^{obs}) \leq E(Y(T=0) \mid G=6) \leq E_{011}^{\geq\pi_{6|011}}(Y^{obs})$.

On the other hand, $E_{001}(Y^{obs})$ is equal to $E(Y(T=0,Z=0) \mid G=16)$ and also to, by Assumption 3, $E(Y(T=0,Z=1) \mid G=16)$. Therefore, $E(Y(T=0) \mid G=6)\pi_6 = E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16} - E(Y(T=0,Z=1) \mid G=8)\pi_8$, where $E(Y(T=0,Z=1) \mid G=8) = E(Y(T=0,Z=0) \mid G=8)$ by Assumption 3. The observed group of units with $T=0$, $Z=1$ and $S^{obs}=1$ is the $\pi_{6|011}$, $\pi_{8|011}$, $\pi_{16|011}$ mixture of principal strata 6, 8, and 16, therefore the minimum (maximum) value of $E(Y(T=0) \mid G=8)$ is the expected value of Y for the $\pi_{8|011}$ fraction of smallest (largest) values for units in the observed group with $T=0$, $Z=1$ and $S^{obs}=1$: $E_{011}^{\leq\pi_{8|011}}(Y^{obs}) \leq E(Y(T=0) \mid G=8) \leq E_{011}^{\geq\pi_{8|011}}(Y^{obs})$. Define $\Delta_0^6 = E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16}$, then

$$\frac{\Delta_0^6 - E_{011}^{\geq\pi 8|011}(Y^{obs})\pi_8}{\pi_6} \leq E(Y(T=0) \mid G=6) \leq \frac{\Delta_0^6 - E_{011}^{\leq\pi 8|011}(Y^{obs})\pi_8}{\pi_6},$$

Combining the two bounds on $E(Y(T=0) \mid G=6)$, given π_2, we have

$$\max\left\{E_{011}^{\leq\pi_{6|011}}(Y^{obs}); \frac{\Delta_0^6 - E_{011}^{\geq\pi 8|011}(Y^{obs})\pi_8}{\pi_6}\right\} \leq$$
$$E(Y(T=0) \mid G=6) \leq \min\left\{E_{011}^{\geq\pi_{6|011}}(Y^{obs}); \frac{\Delta_0^6 - E_{011}^{\leq\pi 8|011}(Y^{obs})\pi_8}{\pi_6}\right\}.$$

Using information provided by the observable means $E_{101}(Y^{obs})$ and $E_{111}(Y^{obs})$, a similar reasoning leads to derive the following bounds on $E(Y(T=1) \mid G=6)$ for a fixed value of π_2:

$$\max\left\{E_{111}^{\leq\pi_{6|111}}(Y^{obs}); \frac{\Delta_1^6 - E_{111}^{\geq\pi 2|111}(Y^{obs})\pi_2}{\pi_6}\right\} \leq$$
$$E(Y(T=1) \mid G=6) \leq \min\left\{E_{111}^{\geq\pi_{6|111}}(Y^{obs}); \frac{\Delta_1^6 - E_{111}^{\leq\pi 2|111}(Y^{obs})\pi_2}{\pi_6}\right\},$$

where $\Delta_1^6 = E_{111}(Y^{obs})(\pi_2+\pi_4+\pi_6+\pi_8+\pi_{16}) - E_{101}(Y^{obs})(\pi_4+\pi_8+\pi_{16})$. Hence, given π_2,

$$\max\left\{E_{111}^{\leq\pi_{6|111}}(Y^{obs}); \tfrac{\Delta_1^6 - E_{111}^{\geq\pi_2|111}(Y^{obs})\pi_2}{\pi_6}\right\} - \min\left\{E_{011}^{\geq\pi_{6|011}}(Y^{obs}); \tfrac{\Delta_0^6 - E_{011}^{\leq\pi 8|011}(Y^{obs})\pi_8}{\pi_6}\right\}$$
$$\leq E(Y(T=1) - Y(T=0) \mid G=6) \leq$$

$$\min\left\{E_{111}^{\geq \pi_{6|111}}(Y^{obs}); \frac{\Delta_1^6 - E_{111}^{\leq \pi_{2|111}}(Y^{obs})\pi_2}{\pi_6}\right\} - \max\left\{E_{011}^{\leq \pi_{6|011}}(Y^{obs}); \frac{\Delta_0^6 - E_{011}^{\geq \pi_{8|011}}(Y^{obs})\pi_8}{\pi_6}\right\}$$

Minimizing (maximizing) the lower (upper) bound over the possible range of π_2 gives the desired bounds on $E(Y(T=1) - Y(T=0) \mid G=6)$ in Equation (A1).

Bounds on the average treatment effect in stratum 8. Assume that the value of π_2 is known. Then, the following bounds for $E(Y(T=0) \mid G=8)$ can be derived, using a similar reasoning as for bounding $E(Y(T=0) \mid G=6)$ (see the proof of bounds on the average treatment effect in stratum 6):

$$\max\left\{E_{011}^{\leq \pi_{8|011}}(Y^{obs}); \frac{\Delta_0^8 - E_{011}^{\geq \pi 6|011}(Y^{obs})\pi_6}{\pi_8}\right\} \leq$$
$$E(Y(T=0) \mid G=8) \leq \min\left\{E_{011}^{\geq \pi_{8|011}}(Y^{obs}); \frac{\Delta_0^8 - E_{011}^{\leq \pi 6|011}(Y^{obs})\pi_6}{\pi_8}\right\},$$

where $\Delta_0^8 = E_{011}(Y^{obs})(\pi_6 + \pi_8 + \pi_{16}) - E_{001}(Y^{obs})\pi_{16}$.

The observed group of units with $T=1$, $Z=1$ and $S^{obs}=1$ is the $\pi_{2|111}$, $\pi_{4|111}$, $\pi_{6|111}$, $\pi_{8|111}$, and $\pi_{16|111}$ mixture of principal strata 2, 4, 6, 8, and 16. Similarly, the observed group of units with $T=1$, $Z=0$ and $S^{obs}=1$ is the $\pi_{4|111}$, $\pi_{8|111}$ and $\pi_{16|111}$ mixture of principal strata 4, 8, and 16. In addition, Assumption 3 implies that $E(Y(T=1, Z=0) \mid G=8) = E(Y(T=1, Z=1) \mid G=8)$. Therefore, in large samples, given π_2, $E(Y(T=1) \mid G=8)$ is at least (at most) the maximum (minimum) between the expected value of Y for the $\pi_{8|101}$ fraction of smallest (largest) values of Y for units in the observed group with $T=1$, $Z=0$ and $S^{obs}=1$, and the expected value of Y for the $\pi_{8|111}$ fraction of smallest (largest) values of Y for units in the observed group with $T=1$, $Z=1$ and $S^{obs}=1$: $\max\left\{E_{101}^{\leq \pi_{8|101}}(Y^{obs}); E_{111}^{\leq \pi_{8|111}}(Y^{obs})\right\} \leq E(Y(T=1) \mid G=8) \leq \min\left\{E_{101}^{\geq \pi_{8|101}}(Y^{obs}); E_{111}^{\geq \pi_{8|111}}(Y^{obs})\right\}$. Hence, given π_2,

$$\max\left\{E_{101}^{\leq \pi_{8|101}}(Y^{obs}); E_{111}^{\leq \pi_{8|111}}(Y^{obs})\right\} - \min\left\{E_{011}^{\geq \pi_{8|011}}(Y^{obs}); \frac{\Delta_0^8 - E_{011}^{\leq \pi 6|011}(Y^{obs})\pi_6}{\pi_8}\right\}$$
$$\leq E(Y(T=1) - Y(T=0) \mid G=8) \leq$$
$$\min\left\{E_{101}^{\geq \pi_{8|101}}(Y^{obs}); E_{111}^{\geq \pi_{8|111}}(Y^{obs})\right\} - \max\left\{E_{011}^{\geq \pi_{8|011}}(Y^{obs}); \frac{\Delta_0^8 - E_{011}^{\geq \pi 6|011}(Y^{obs})\pi_6}{\pi_8}\right\}.$$

Minimizing (maximizing) the lower (upper) bound over the possible range of π_2 gives the desired bounds on $E(Y(T=1) - Y(T=0) \mid G=8)$ in Equation (A2).

Bounds on the average treatment effect in stratum 16. Under control, due to Assumption 3, $E(Y(T=0) \mid G=16)$ can be identified as $E_{001}(Y^{obs})$. Under treatment, using a similar reasoning as for bounding $E(Y(T=1) \mid G=8)$ (see the proof of bounds on the average treatment effect in stratum 8), we can partially identify $E(Y(T=1) \mid G=8)$ as follows: $\max\left\{E_{101}^{\leq \pi_{16|101}}(Y^{obs}); E_{111}^{\leq \pi_{16|111}}(Y^{obs})\right\} \leq E(Y(T=1) \mid G=16) \leq \min\left\{E_{101}^{\geq \pi_{16|101}}(Y^{obs}); E_{111}^{\geq \pi_{16|111}}(Y^{obs})\right\}$. Hence,

$$\max\left\{E_{101}^{\leq \pi_{16|101}}(Y^{obs}); E_{111}^{\leq \pi_{16|111}}(Y^{obs})\right\} - E_{001}(Y^{obs})$$
$$\leq E(Y(T=1) - Y(T=0) \mid G=16) \leq$$
$$\min\left\{E_{101}^{\geq \pi_{16|101}}(Y^{obs}); E_{111}^{\geq \pi_{16|111}}(Y^{obs})\right\} - E_{001}(Y^{obs}).$$

Bounds on the average treatment effect for the union of strata 6 and 8. Under control, the observable mean $E_{011}(Y^{obs})$ is equal to

$$\frac{E(Y(T=0, Z=1) \,|\, G=6)\pi_6 + E(Y(T=0, Z=1) \,|\, G=8)\pi_8 + E(Y(T=0, Z=1) \,|\, G=16)\pi_{16}}{\pi_6 + \pi_8 + \pi_{16}},$$

where $E(Y(T=0,Z=1) \mid G=j) = E(Y(T=0,Z=0) \mid G=j)$, $j=6,8,16$, by Assumption 3. The observable mean $E_{001}(Y^{obs})$ is equal to $E(Y(T=0,Z=0) \mid G=16)$ and also to, by Assumption 3, $E(Y(T=0,Z=1) \mid G=16)$. Therefore, we can, by difference, identify $E(Y(T=0,Z=1) \mid G=6)\pi_6 + E(Y(T=0,Z=1) \mid G=8)\pi_8$ as $E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16}$.
Under treatment, due to Assumption 3, $E(Y(T=1,Z=0) \mid G \in \{6,8\}) = E(Y(T=1,Z=1) \mid G \in \{6,8\})$. The observed group of units with $T=1$, $Z=1$ and $S^{obs}=1$ is the $\pi_{2|111}$, $\pi_{4|111}$, $\pi_{6|111}$, $\pi_{8|111}$, and $\pi_{16|111}$ mixture of principal strata 2, 4, 6, 8, and 16, thus the minimum (maximum) value of $E(Y(T=1) \mid G \in \{6,8\})$ is the expected value of Y for the $\pi_{6,8|111}$ fraction of smallest (largest) values of Y for units in the observed group with $T=1$, $Z=1$ and $S^{obs}=1$: $E_{111}^{\leq \pi_{6,8|111}}(Y^{obs}) \leq E(Y(T=1) \mid G \in \{6,8\}) \leq E_{111}^{\geq \pi_{6,8|111}}(Y^{obs})$. Therefore, the average causal effect, $E(Y(T=1) - Y(T=0) \mid G \in \{6,8\})$ is at least

$$E_{111}^{\leq \pi_{6,8|111}}(Y^{obs}) - \frac{E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16}}{\pi_6+\pi_8},$$

and at most

$$E_{111}^{\geq \pi_{6,8|111}}(Y^{obs}) - \frac{E_{011}(Y^{obs})(\pi_6+\pi_8+\pi_{16}) - E_{001}(Y^{obs})\pi_{16}}{\pi_6+\pi_8}.$$

Bounds on the average treatment effect for the union of strata 6 and 16. Assume that the value of π_2 is known. By Assumption 3, $E(Y(T=t,Z=0) \mid G \in \{6,16\}) = E(Y(T=t,Z=1) \mid G \in \{6,16\})$, $t=0,1$. The observed group of units with $T=0$, $Z=1$ and $S^{obs}=1$ is the $\pi_{6|011}$, $\pi_{8|011}$, and $\pi_{16|011}$ mixture of principal strata 6, 8, and 16, thus the minimum (maximum) value of $E(Y(T=0) \mid G \in \{6,16\})$ is the expected value of Y for the $\pi_{6,16|011}$ fraction of smallest (largest) values of Y for units in the observed group with $T=0$, $Z=1$ and $S^{obs}=1$: $E_{011}^{\leq \pi_{6,16|011}}(Y^{obs}) \leq E(Y(T=0) \mid G \in \{6,16\}) \leq E_{011}^{\geq \pi_{6,16|011}}(Y^{obs})$. Analogously, the observed group of units with $T=1$, $Z=1$ and $S^{obs}=1$ is the $\pi_{2|111}$, $\pi_{4|111}$, $\pi_{6|111}$, $\pi_{8|111}$, and $\pi_{16|111}$ mixture of principal strata 2, 4, 6, 8, and 16, thus the minimum (maximum) value of $E(Y(T=1) \mid G \in \{6,16\})$ is the expected value of Y for the $\pi_{6,16|111}$ fraction of smallest (largest) values of Y for units in the observed group with $T=1$, $Z=1$ and $S^{obs}=1$: $E_{111}^{\leq \pi_{6,16|111}}(Y^{obs}) \leq E(Y(T=1) \mid G \in \{6,16\}) \leq E_{111}^{\geq \pi_{6,16|111}}(Y^{obs})$. Therefore, the average causal effect, $E(Y(T=1) - Y(T=0) \mid G \in \{6,16\})$, given π_2 is at least $E_{111}^{\leq \pi_{6,16|111}}(Y^{obs}) - E_{011}^{\geq \pi_{6,16|011}}(Y^{obs})$, and at most $E_{111}^{\geq \pi_{6,16|111}}(Y^{obs}) - E_{011}^{\leq \pi_{6,16|011}}(Y^{obs})$. Minimizing (maximizing) the lower (upper) bound over the possible range of π_2 gives the desired bounds on $E(Y(T=1) - Y(T=0) \mid G \in \{6,16\})$ in Equation (A5).

Bounds on the average treatment effect for the union of strata 8 and 16. Assume that the value of π_2 is known. Under control, using a similar reasoning as for bounding $E(Y(T=0) \mid G \in \{6,16\})$ (see the proof of bounds on the average treatment effect in strata 6 and 16), we have that, given π_2, $E(Y(T=0) \mid G \in \{8,16\})$ is at least $E_{011}^{\leq \pi_{8,16|011}}(Y^{obs})$ and at most $E_{011}^{\geq \pi_{8,16|011}}(Y^{obs})$. Analogously, under treatment, using a similar reasoning as for bounding $E(Y(T=1) \mid G=16)$ (see the proof of bounds on the average treatment effect in stratum 16), we have that, given π_2, $E(Y(T=1) \mid G \in \{8,16\})$ is at least (at most) the maximum (minimum) between the expected value of Y for the $\pi_{8,16|101}$ fraction of largest values of Y for units in the observed group with $T=1$, $Z=0$ and $S^{obs}=1$, and the expected value of Y for the $\pi_{8,16|111}$ fraction of largest values of Y for units in the observed group with $T=1$, $Z=1$ and $S^{obs}=1$: $\max\left\{E_{101}^{\leq \pi_{8,16|101}}(Y^{obs}); E_{111}^{\leq \pi_{8,16|111}}(Y^{obs})\right\} \leq E(Y(T=1) \mid G \in \{8,16\}) \leq \min\left\{E_{101}^{\geq \pi_{8,16|101}}(Y^{obs}); E_{111}^{\geq \pi_{8,16|111}}(Y^{obs})\right\}$. Hence, given π_2,

$$\max\left\{E_{101}^{\leq \pi_{8,16|101}}(Y^{obs}); E_{111}^{\leq \pi_{8,16|111}}(Y^{obs})\right\} - E_{011}^{\geq \pi_{8,16|011}}(Y^{obs})$$
$$\leq E(Y(T=1) - Y(T=0) \mid G \in \{8,16\}) \leq$$
$$\min\left\{E_{101}^{\geq \pi_{8,16|101}}(Y^{obs}); E_{111}^{\geq \pi_{8,16|111}}(Y^{obs})\right\} - E_{011}^{\leq \pi_{8,16|011}}(Y^{obs})$$

Minimizing (maximizing) the lower (upper) bound over the possible range of π_2 gives the desired bounds on $E(Y(T=1)-Y(T=0) \mid G \in \{8,16\})$ in Equation (A6). □

Proposition 2 shows how Assumption 7 allows us to tighten the bounds on the causal estimands $E(Y(T=1)-Y(T=0) \mid G \in \{6,8,16\})$ given in Proposition 1. Analogously, the bounds in Proposition 6 on the causal estimands $E(Y(T=1)-Y(T=0) \mid G \in g)$, for $g = \{6\},\{8\},\{16\},\{6,8\},\{6,16\},\{8,16\}$, can be tightened using alternative latent ignorability assumptions for nonresponse. Here, we omit these results, which are available upon request from the authors.

REFERENCES

ANGRIST J. D., & IMBENS, G. W. & RUBIN, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *J. Am. Statist. Assoc.* **91**, 444–472.

BURGHARDT, J., MCCONNELL, S. & SCHOCHET, P. & JOHNSON, T. & GRITZ, M. & GLAZERMAN, S. & HOMRIGHAUSEN, J. & JACKSON, R. (2001). Does Job Corps work? Summary of the National Job Corps Study. Document No. PR01-50, Princeton, NJ: Mathematica Policy Research, Inc.

CHEN, H. & GENG, Z. & ZHOU, X-H. (2009). Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data (with discussion). *Biometrics* **65**, 675–691.

CHENG, J. & SMALL, D. S. & TAN, Z. & TEN HAVE, T. R. (2009). Efficient nonparametric estimation of causal effects in randomized trials with noncompliance. *Biometrika* **96**, 19–36.

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.

FRANGAKIS C. E., & RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 191–199.

FRANGAKIS C. E., & RUBIN, D. B. & AN, M. & MACKENZIE, E. (2007). Principal stratification designs to estimate input data missing due to death. *Biometrics* 63 (2007), 641–649.

FRÖLICH, M. (2007) Nonparametric IV estimation of local average treatment effects with covariates. *J. Econometrics* **139**, 35–75.

FUMAGALLI, L. & LAURIE, H. & LYNN, P. (2010). Experiment with methods to reduce attrition in longitudinal surveys. *ISER Working Paper* 2010-4.

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.

IMAI, K. (2008). Sharp bounds on the causal effects in randomized experiments with truncation-by-death. *Statist. Probab. Lett.* **78**, 144–149.

IMBENS, G. W. & MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica* **72**, 1845–1857.

JANSSENS, W. & VAN DER GAAG, J. & RINKE DE WIT, T. (2008). Pitfalls in the estimation of HIV prevalence. *Amsterdam Institute for International Development Research Series*, RS 08-03.

KRUEGER, A. B. (1999). Experimental estimates of education production functions. *Quart. J. Econ.* **114**, 497–532.

LEE, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Rev. Econ. Stud.* **76**, 1071–1102.

LEPKOWSKI, J. M. & COUPER, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In *Survey Nonresponse*, Eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, New York: John Wiley.

LITTLE, R. J. A. & RUBIN, D. B. (2002). *Statistical Analysis with Missing Data.* 2nd Edition, New York: John Wiley.

MANSKI, C. F. (2003). *Partial Identification of Probability Distributions.* Springer-Verlag.

MEALLI, F. & PACINI, B. (2008). Comparing principal stratification and selection models in parametric causal inference with nonignorable missingness. *Computational Statistics and Data Analysis* **53**, 507–516.

NICOLETTI, C. (2010). Poverty analysis with missing data: Alternative estimators compared. *Empirical Econ.* **38**, 1–22.

NICOLETTI, C. & PERACCHI, F.(2005). Survey response and survey characteristics: Micro-level evidence from the European Community Household Panel. *J. R. Statist. Soc. A* **168**, 1–19.

ROSENBAUM, P. & RUBIN D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psycol.* **66**, 688–701.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

RUBIN, D. B. (1978). Bayesian inference for causal effects. *Ann. Statist.* **6**, 34–58.

RUBIN, D. B. Discussion of "Randomization analysis of experimental data: the Fisher randomization test" by D. Basu, *J. Am. Statist. Assoc.* **75**, 591–593.

RUBIN, D. B. (1998). More powerful randomization-based p-values in double-blind trials with noncompliance (with discussion). *Statist. Med.* **17**, 371–389.

SCHRÄPLER, J. P. (2004). Respondent behavior in panel studies - A case study for income nonresponse by means of the German Socio-Economic Panel (SOEP)" *Sociological Methods & Research* **33**, 118–156.

ZHANG, J. & RUBIN, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J. Educ. Behav. Statist.* **28**, 353–368.

ZHANG, J. & RUBIN, D. B. & MEALLI, F. (2009). Likelihood-based analysis of causal effects via principal stratification: New Approach to evaluating job-training programs. *J. Am. Statist. Assoc.* **104**, 166–176.