A Review on Principal Stratification with Noncompliance and Missing Outcome Data via Mixture Model

Holly Cui
Duke University
Department of Statistical Science
Durham, NC 27708

Abstract—**This review compares and examines two seminal studies using principal stratification to address noncompliance and missing outcome data via different mixture models, highlighting methodological strengths and offering insights into broken randomized experiments.**

INTRODUCTION

In an ideal setting for causal inference, units are randomized to treatment and control, ensuring that units in both arms have the same expected distribution of all pre-randomization individual characteristics. However, in most observational research involving human subjects, the idealization suffered from post-treatment complications, most notably (1) noncompliance with the randomly assigned treatment, (2) partially defined outcomes, and (3) unintended missing outcomes. The context has been generalized as *broken randomized experiments*, a phrase first coined by Barnard et al. (1998).

To address the complications, the widely accepted approach to formulate causal questions is in terms of *potential outcomes* regardless of the mode of inference, often referred to as the Rubin causal model (RCM; Holland (1986)) for work extending the framework to observational studies. Due to the reason that causal effects are defined by comparisons of potential outcomes on a common set of units, the framework of *principal stratification* (PS; Frangakis and Rubin (2004)) has become prevalent for the challenge of noncompliance based on the technique of instrumental variables (IV; Angrist, Imbens, and Rubin (1996)) in nonrandomized studies in econometrics. Inherently, different PS settings could share the estimation and inference procedures, but the estimands of interests must be carefully crafted to cater for each subgroup under additional assumptions.

A central challenge revolving PS framework resides in the fact that our hypothetical PS memberships are not fully observed for each experimental unit. Moreover, the principal strata are usually nested as a mixture within each group. Therefore, with carefully chosen assumptions, the latent mixture model approach became a hotspot to efficiently extend to general settings, under which proposed the two papers we review. Both authors intend to address the complications in question, where Barnard et al. (2003) utilizes a Bayesian mixture to account for estimates and missingness, and Frumento et al. (2012) adopts a finite mixture model maximized by the EM (expectation-maximization) algorithm (Dempster, n.d.) for granular estimands. By reviewing the two papers proposed, we intend to gain insights into the proper handling of noncompliance and missing outcome data with comparative highlights under PS.

STUDY 1 (BARNARD ET AL. 2003)

Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City

Background

The study evaluates a randomized experiment conducted in New York City by the School Choice Scholarships Foundation (SCSF) Program in 1997. The program randomized the distribution of vouchers that provides 1,300 scholarships to "eligible[1]" children from low-income families to study in private schools. The background characteristics of eligible applicants are collected through a session and a questionnaire, including a pretest administered to eligible children as an additional covariate. (The pre-analysis stage adopts a propensity matched pairs design (PMPD) to balance the covariates. Refer to 3 and 4 in Appendix for final sample size and covariates balance.)

Goal

The overarching goal of this research is to study the effects of the vouchers program (i.e. scholarships) on the school performance of eligible children (excluding kindergarten applicants or multifamilies children). The school performance as an outcome is contextualized as a quantitative measurement, a posttest score (mutivariate in two subjects - math and reading).

However, there are three main challenges with the data. (1) Missingness of background variables: not all families were capable of attending the session in the application stage, and some children did not have a pretest score. (2)

[1]Eligibility required that the children, at the time of application, be living in New York City, entering grades 1–5, currently attending a public school, and from families with in- comes low enough to qualify for free school lunch.

Missingness of outcomes: some children chose to not carry out a posttest after the lottery results of the vouchers have been released. Contrasting with missingness of background, missingness of outcomes that occurs after randomization is not guaranteed to be balanced between the randomized arms. (3) Noncompliance: attendance at a private school was not perfectly correlated with winning a scholarship. For example, some children who won scholarships did not use them (since the amount of scholarships is only $1,400 per person), and some that did not win scholarships were sent to private schools nevertheless.

Method

To address the pain above, Barnard et al. categorizes the situation under *broken randomized experiments* and defines a Bayesian parametric mixture model under PS that accommodates noncompliance, missing covariate information, missing outcome data, and multivariate outcomes. The choice of Bayesian approach is to consider the unobserved (missing) values as unobserved random variables and obtain the posterior predictive distributions of the missing data.

It is worth denoting that the authors also decide to treat the experiment as a *randomized encouragement study*. Specifically, they denoted that the experiment does not really randomize attendance, but rather randomizes the *encouragement*, using financial support, to attend a private school. And the interest of focus is not only on the effect of encouragement itself, but also on the effect of the treatment being encouraged, which further justifies the chosen PS framework for noncompliance.

Variables Defined

Under the definition of *potential outcomes*, we thus define:

- i: Child i in our study for $i \in \{1, ..., n\}$
- z: Treatment indicator of child i (1 for private school/scholarship offered; 0 for public school/no scholarship offered)
- $D_i(z)$: Compliance indicator of child i (1 if the child attends private school; 0 if the child attends public school)
- $Y_i(z)$: Potential outcomes of the posttest of child i (if the child were to take the posttest)
- $R_{y_i}(z)$: Outcome indicator of posttest of child i (1 if the child takes the posttest)

Under observed values, we further denote:

- Z_i: Observed treatment assignment indicator of child i
- $D_i = D_i(Z)$: Actual type of school attending of child i
- $Y_i = Y_i(Z_i)$: The outcome to be recorded by the test of child i
- $R_{y_i}(Z_i)$: Whether or not the child takes the test under the observed assignment of child i

More granular definitions of variable notations are provided in 5 in the Appendix. From the definition of D_i, we extend our grouping of subjects into four principal strata C_i:

- $n : \{i : D_i(0) = 0, D_i(1) = 0\}$: Never Takers
- $c : \{i : D_i(0) = 0, D_i(1) = 1\}$: Compliers
- $d : \{i : D_i(0) = 1, D_i(1) = 0\}$: Defiers
- $a : \{i : D_i(0) = 1, D_i(1) = 1\}$: Always Takers

By the nature of PS, the memberships of C_i are not affected by the assigned treatment. And comparisons of potential outcomes under different assignments within principal strata (i.e., principal effects) are well-defined causal effects (Frangakis and Rubin 2004). However, the strata are not fully observed, and the observed (Z, D) groupings consist of a mixture of the strata. For example, children who are observed to attend private school when winning the voucher lottery are a mixture of compliers ($C_i = c$) and always takers ($C_i = a$). Therefore, the mixture model provides enhanced feasibility in handling the nested structure of likelihood and estimates.

Estimands

Based on the defined framework and variables, the study decides to focus on:

ITT: The *intention-to-treat effect*, which is the effect of the randomized encouragement on all subjects.[2]

$$E[Y_i(1) - Y_i(0)|W_i^p, \theta]$$

CACE: The *complier average causal effect*, which is the effect of the randomized encouragement on compliers ($C_i = c$).

$$E[Y_i(1) - Y_i(0)|W_i^p, C_i = c, \theta]$$

Assumptions

To practically implement the proposed framework and address the issue regarding noncompliance and missing data, the study explicitly state the following assumptions:

- *Stable Unit Treatment Value Assumption* (SUTVA): Allow us to assume no interference between family on their outcomes.
- Randomization: Allow us to assume scholarships have been randomly assigned.
- Monotonicity: Allow us to assume there are no defiers ($C_i = d$). This is plausible and reasonable under our context (i.e., no family who would not use a scholarship if they won one, but would pay to go to private school if they did not win a scholarship).
- Compound Exclusion Restriction (ER): The outcomes and missingness of outcomes for never takers ($C_i = n$) and always takers ($C_i = a$) are not affected by treatment assignment, such that:

$$\begin{aligned} &p(Y(1), R_y(1)|X^{obs}, R_x, W, C = n) \\ &= p(Y(0), R_y(0)|X^{obs}, R_x, W, C = n), \quad \text{for never takers} \\ &p(Y(1), R_y(1)|X^{obs}, R_x, W, C = a) \\ &= p(Y(0), R_y(0)|X^{obs}, R_x, W, C = a), \quad \text{for always takers} \end{aligned}$$

By ER, we find that the *ITT* comparison of all outcomes $Y_i(1)$ and $Y_i(0)$ includes the null comparison among the subgroups of never takers and always takers. And by

[2]The ITT is eventually only used for comparison with the simple randomized experiment results. See Assumptions section for detailed justification.

monotonicity, the compliers are the only group of children who would attend private school if and only if offered the scholarship. Thus, the authors took *CACE* to represent the effect of attending public versus private school.

- Latent Ignorability (LI): Allow us to assume the potential outcomes are independent of missingness, given observed covariates conditional on the compliance strata (Unconfoundedness), such that:

$$
\begin{aligned}
&p(R_y(0), R_y(1)|R_x, Y(0), Y(1), X^{obs}, W, C, \theta) \\
&= p(R_y(0), R_y(1)|R_x, X^{obs}, W, C, \theta)
\end{aligned}
$$

Theoretically, the previous three assumptions are already sufficient to identify ITT and CACE. However, according to the authors, it would require a very large sample size and explicit stratification on all subclasses to perform separate modeling analyses. Therefore, LI is further assumed to create more parsimonious modeling approaches that could assist the inference with robustness.

Modeling

The study proceed with a *partial pattern mixture model* approach to missing data problem. The reason is that it focuses the model on the quantities of interest in such a way that parametric specifications for the marginal distributions of R_x, W and X^{obs} can be ignored. Hence, starting from the factorization of the join distribution for the potential outcomes and compliance strata conditional on the covariates and their missing-data patterns, we have:

$$
\begin{aligned}
&p(Y_i(0), Y_i(1), R_{y_i}(0), R_{y_i}(1), C_i|W_i, X_i^{obs}, R_{x_i}, \theta) \\
&= p(C_i|W_i, X_i^{obs}, R_{x_i}, \theta^{(C)}) \\
&\times p(R_{y_i}(0), R_{y_i}(1)|W_i, X_i^{obs}, R_{x_i}, C_i, \theta^{(R)}) \\
&\times p(Y_i(0), Y_i(1)|W_i, X_i^{obs}, R_{x_i}, C_i, \theta^{(Y)})
\end{aligned}
$$

S-model: The compliance principal stratum submodel is defined with two conditional probit models for whether individual i is a complier or a never taker:

$$C_i(n) = 1 \text{ if } C_i(n)^* = g_1(W_i, X_i^{obs}, R_{x_i})'\beta^{(C,1)} + V_i \le 0$$

and

$$
\begin{aligned}
&C_i(c) = 1 \text{ if } C_i(n)^* > 0 \\
&\qquad \text{and } C_i(c)^* = g_0(W_i, X_i^{obs}, R_{x_i})'\beta^{(C,2)} + U_i \le 0
\end{aligned}
$$

where V_i and U_i follows standard normal distribution; g_0 and g_1 are linear link functions in order to balance between the inclusion of all design variables and the maintenance of parsimony; $\beta^{(C,1)}$ and $\beta^{(C,2)}$ are the prior distributions for the compliance submodel following normal distribution with designed hyperparameters.

Y-model: The outcome submodel describes the marginal distribution for the posttest scores Y. Since the outcome is partially missing (with pile-up zeros), the authors defined a censored model as below. The detailed Bayesian specification of hyperparameters and the sampling process can be found in the paper. The model has been checked to ensure a satisfactory level of estimate (similar to MLE gits in likelihood models) and a quick enough mixing rate.

$$
Y_i^{(\text{math})}(z) = \begin{cases} 0 & \text{if } Y_i^{(\text{math}),*}(z) \le 0 \\ 100 & \text{if } Y_i^{(\text{math}),*}(z) \ge 100 \\ Y_i^{(\text{math}),*}(z) & \text{otherwise,} \end{cases}
$$

where

$$
\begin{aligned}
&Y_i^{(\text{math}),*}(z) \mid W_i, X_i^{\text{obs}}, Rx_i, C_i, \theta^{(\text{math})} \\
&\quad \sim \mathrm{N}\big(g_2(W_i, X_i^{\text{obs}}, Rx_i, C_i, z)'\beta^{(\text{math})}, \\
&\qquad \exp\big[g_3(X_i^{\text{obs}}, Rx_i, C_i, z)'\zeta^{(\text{math})}\big]\big)
\end{aligned}
$$

R-model: Apart from the two standard models in a mixture approach, a separate outcome response submodel on $R_{y_i}(z)$ is proposed for missingness of outcomes:

$$
\begin{aligned}
&R_{y_i}(z) = 1 \\
&\qquad \text{if } R_{y_i}(z)^* = g_2(W_i, X_i^{obs}, R_{x_i}, C_i, z)'\beta^R + E_i(z) \ge 0
\end{aligned}
$$

where $E_i(z)$ follows standard normal distribution; $R_{y_i}(0)$ and $R_{y_i}(1)$ are assumed conditionally independent; $\beta^{(R)}$ are the prior distributions for the outcome response submodel following normal distribution with designed hyperparameters.

Results

All results herein are obtained through the same Bayesian mixture structure. The reported estimated ITT and CACE values are not parameters of the model, but rather are functions of parameters and data. The plain numbers are posterior means, and the values in parenthese are 2.5 and 97.5 percentiles of the posterior distribution. They have been adjusted according to the national median and the individual-level estimates were weighted by subgroups.

1. Impact of being offered a scholarship on student outcomes (ITT): For the treatment effect on math scores overall for children from low applicant schools, and also for the subgroup of first graders, the results indicate posterior distributions with mass primarily (>97.5%) to the right of 0. Each effect indicates an average gain of greater than three percentile points for children who won a scholarship while all other intervals cover 0. As a more general pattern, estimates of effects are larger for mathematics than for reading and larger for children from low applicant schools than for children from high-applicant schools.

Grade at application	*Applicant's school: Low*		*Applicant's school: High*	
	Reading	*Math*	*Reading*	*Math*
1	$2.3_{(-1.3, 5.8)}$	$5.2_{(2.0, 8.3)}$	$1.4_{(-4.8, 7.2)}$	$5.1_{(.1, 10.3)}$
2	$.5_{(-2.6, 3.5)}$	$1.3_{(-1.7, 4.3)}$	$-.6_{(-6.2, 4.9)}$	$1.0_{(-4.3, 6.1)}$
3	$.7_{(-2.7, 4.0)}$	$3.3_{(-.5, 7.0)}$	$-.5_{(-6.0, 5.0)}$	$2.5_{(-3.2, 8.0)}$
4	$3.0_{(-1.1, 7.4)}$	$3.1_{(-1.2, 7.2)}$	$1.8_{(-4.1, 7.6)}$	$2.3_{(-3.3, 7.8)}$
Overall	$1.5_{(-.6, 3.6)}$	$3.2_{(1.0, 5.4)}$	$.4_{(-4.6, 5.2)}$	$2.8_{(-1.8, 7.2)}$

NOTE: Year Postrandomization Plain numbers are means, and numbers in parentheses are central 95% intervals of the posterior distribution of the effects on percentile rank.

Fig. 1. ITT Effect of Winning the Lottery on Math and Reading Test Scores

2. Impact of attendinga private school on student outcomes (CACE): The effects of private school attendance follow a pattern similar to that of the ITT effects but with slightly bigger posterior means in absolute value. The intervals have also grown, reflecting that these effects are for only complier subgroups of all children in each cell. As a result, the associated uncertainty for some of these effects (e.g., for fourth graders applying from high-applicant schools) is large.

Grade at application	*Applicant's school: Low*		*Applicant's school: High*	
	Reading	*Math*	*Reading*	*Math*
1	$3.4_{(-2.0,8.7)}$	$7.7_{(3.0,12.4)}$	$1.9_{(-7.3,10.3)}$	$7.4_{(.2,14.6)}$
2	$.7_{(-3.7,5.0)}$	$1.9_{(-2.4,6.2)}$	$-.9_{(-9.4,7.3)}$	$1.5_{(-6.2,9.3)}$
3	$1.0_{(-4.1,6.1)}$	$5.0_{(-.8,10.7)}$	$-.8_{(-9.5,7.7)}$	$4.0_{(-4.9,12.5)}$
4	$4.2_{(-1.5,10.1)}$	$4.3_{(-1.6,10.1)}$	$2.7_{(-6.3,11.3)}$	$3.5_{(-4.7,11.9)}$
Overall	$2.2_{(-.9,5.3)}$	$4.7_{(1.4,7.9)}$	$.6_{(-7.1,7.7)}$	$4.2_{(-2.6,10.9)}$

NOTE: Plain numbers are means, and parentheses are central 95% intervals of the posterior distribution of the effects on percentile rank.

Fig. 2. CACE Effect of Private School Attendance on Test Scores

STUDY 2 (FRUMENTO ET AL. 2012)

Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data

Background

The study evaluates the effects of a job training program, Job Corps, using the data from a randomized study in 1994 to 1995. The study carried out a national random sample of all eligible applicants, and randomly assigned the sampled individuals to the training program (treatment) or the control group. Interviews were conducted at three subsequent time: week 52, 130, and 208 after the assignment, and employment information at the time of each interview were collected.

Goal

The goal of the study is to evaluate the effects of the training program on both employment status and wages. However, similar to *Study 1*, there are three main challenges with the data: (1) Noncompliance: compliance with assigned treatment was imperfect. Only 68% of those asigned to the program group immediately enrolling (within the first semester after assignment) and participating in the program for at least one week. (2) Partially defined outcomes: wages are not well defined for those who are not employed. (3) Unintended missing outcomes: outcome variables are missing for some participants in the study at some point(s) of the three interviews.

Method

To explicitly address all three challenges simultaneously, the study adopts the *potential outcomes* framework (RCM) as well, and define the causal question under PS. Specifically, for (1) and (2), Frumento et al. proposes the requirements of newly defined causal estimands. For (3), they consider the parameters governing the missingness to be *nuisance*. In terms of defining the principal strata, the authors' approach is quite unique due to the nature of the data: they classify the individuals according to the joint values, whether assigned or not to be trained, of their (a) potential compliances, (2) potential employment statuses, and (3) potential missingness behaviors. Therefore, the potential model complexity grows vastly compared to *Study 1*.

The final modeling of the study adopts a finite mixture model likelihood, which is subsequently maximized by the EM algorithm. It is important to note that there are three time points for the outcomes being collected. In principle, it is feasible to analyze the three weeks jointly under the PS framework. However, according to the authors, it is impractical under the current study design, which already had a number of principal strata for one week. Therefore, they proceeded with analyzing the three weeks independently.

Variables Defined

- X: Pre-assignment observed covariates (A small amount of missing values are `MICE` imputed)
- z: Treatment indicator (1 for training assigned; 0 for no training assigned)
- $D_i(1)$: Compliance indicator for only the treatment group (1 if the unit immediately enrolls in the training; 0 otherwise). Note that if the unit was not assigned the training ($Z = 0$), they will be denied access to the training site, leading to $D_i(0) = 0\ \forall i$ being suppressed
- $S_i(z)$: Potential employment status indicator (1 employed; 0 unemployed)
- $W_i(z)$: Potential outcomes of the wages (if the unit is employed)
- $M_i(z)$: Potential missingness indicator (1 missing; 0 existing)

Note that since wages are well defined only if the unit is employed ($S_i(z) = 1$), further relations are proposed as $W_i(z) = *$ when $S_i(z) = 0$. In our study, W and S are either both observed or both missing, so:

- $M_i(z) = 0$ when $S_i(z)$ and $W_i(z)$ are both observed
- $M_i(z) = 1$ when $S_i(z)$ and $W_i(z)$ are both missing and coded as "?"

In this framework, we could classify units into eight PS (later reduced to six by Assumptions) by $D \times S$:

- $c = \{i : D_i(1) = 1\}$, the subpopulation of compliers;
- $n = \{i : D_i(1) = 0\}$, the subpopulation of noncompliers;
- $EE = \{i : S_i(1) = S_i(0) = 1\}$, those who would be employed regardless of their treatment assignment; for this stratum, $W_i(1)$ and $W_i(0)$ are defined in $\Re^+$;
- $EN = \{i : S_i(1) = 1 \text{ and } S_i(0) = 0\}$, those who would be employed only if assigned treatment; for this stratum, $W_i(1) \in \Re^+$ and $W_i(0) = *$;
- $NE = \{i : S_i(1) = 0 \text{ and } S_i(0) = 1\}$, those who would be employed only if assigned to the control group; for this stratum, $W_i(1) = *$ and $W_i(0) \in \Re^+$; and
- $NN = \{i : S_i(1) = S_i(0) = 0\}$, those who would be nonemployed regardless of their treatment assignment; for this stratum, $W_i(1) = W_i(0) = *$.

Each of the above eight PS can be further classified into a mixture of four subgroups, now by M. (Refer to Table I in Appendix)

Estimands

Average treatment effect of Z on program participation D:

$$\Delta^{(ZD)} = E[D_i(1)|\theta] = \Pr(D_i(1) = 1|\theta)$$

Average treatment effect of Z on employment S: By ER assumption, we have:

$$\begin{aligned}\Delta^{(ZS)} &= E[S_i(1)|\theta] - E[S_i(0)|\theta] \\ &= \Pr(G_i = c\&EN|\theta) - \Pr(G_i = c\&NE|\theta)\end{aligned}$$

Average treatment effect of Z on employment S for compliers:

$$\begin{aligned}\Delta^{(DS)} &= E[S_i(1)|D_i(1) = 1; \theta] - E[S_i(0)|D_i(1) = 1; \theta] \\ &= \Pr(G_i = c\&EN|c; \theta) - \Pr(G_i = c\&NE|c; \theta)\end{aligned}$$

Average treatment effect of Z on wage W for the always-employed compliers:

$$\Delta^{(DW)} = E[W_i(1)|G_i = c\&EE; \theta] - E[W_i(0)|G_i = c\&EE; \theta]$$

Assumptions

- SUTVA
- Randomization
- ER for Noncompliers: (i) If $D_i(z) = 0$ $(z = 0, 1)$, then $S_i(0) = S_i(1)$. By this, we could eliminate the $n\&EN$ group and the $n\&NE$ group, and reduce the PS from 8 to 6. (ii) If $D_i(z) = 0$ $(z = 0, 1)$, then $W_i(0) = W_i(1)$.
- *Missing at Random* (MAR): for unintended outcomes missingness.

Modeling

Observed Groups of Units: The PS subgroups above are all under the *potential outcomes* setting. However, for estimation and inference, we need to further define the observed values of each unit into granular subgroups, including latent strata. Due to the drastic amount of groups defined at this stage, refer to Figure 6 in the Appendix for details.

E-step: A multinomial logistic model for the k-dimensional vector of PS memberships:

$$\Pr(G_i = g|X_i; \alpha) = \frac{\exp(X_i \alpha_g)}{\sum_{h=1}^{k} \exp(X_i \alpha_h)} = \pi_{i;g}$$

M-step: A normal distribution for log-wages conditional on covariates X:

$$\begin{aligned}
\text{if } G_i = c\&EE, \quad & \log[W_i(1)] \sim N(\mathbf{X}_i \boldsymbol{\beta}_{c\&EE,1}, \sigma^2_{c\&EE,1}), \\
& \log[W_i(0)] \sim N(\mathbf{X}_i \boldsymbol{\beta}_{c\&EE,0}, \sigma^2_{c\&EE,0}), \\
\text{if } G_i = c\&EN, \quad & \log[W_i(1)] \sim N(\mathbf{X}_i \boldsymbol{\beta}_{c\&EN,1}, \sigma^2_{c\&EN,1}), \\
\text{if } G_i = c\&NE, \quad & \log[W_i(0)] \sim N(\mathbf{X}_i \boldsymbol{\beta}_{c\&NE,0}, \sigma^2_{c\&NE,0}), \\
\text{if } G_i = n\&EE, \quad & \log[W_i(1)] \sim \log[W_i(0)] \\
& \quad \sim N(\mathbf{X}_i \boldsymbol{\beta}_{n\&EE}, \sigma^2_{n\&EE}).
\end{aligned}$$

The final observed data likelihood function $L(\theta_{sci}|D(1), S_{obs}, W_{obs}, Z, X)$ is decomposed into the product of each observed subgroups likelihood, with each subgroup likelihood being the sum of the PS within that group. (7 and 8 in Appendix)

Results

To report the causal effects, we must acknowledge that they are estimated within each PS as functions of the observed data and the MLEs of parameters, averaging over the estimated population distribution of covariates in that PS using the design weights. The expressions of all estimated target estimands can be found in Appendix (Figure 9, 10, 11, and 12).

1. Estimated Treatment Effects: The table presents the three main average treatment effects proposed earlier. For compliers, the effect on employment is negative in the short term; it becomes positive in the long term, but remain small at best. For always employed compliers ($c\&EE$), positive effects on wages are found at all time periods. Note that, here, the authors also tried to compare the MLE under three additional, meaningful assumptions and find out that (1) monotonicity is not supported by the data at any week (λ_M); (2) the average effect of assignment on employment for compliers is absent (λ_{0W}); (3) no effect of assignment on wages for the always-employed compliers is rejected by the data at all weeks (λ_{S0}).

Week	$\pi_{c\&EE}$	$\pi_{c\&EN}$	$\pi_{c\&NE}$	$\pi_{c\&NN}$	$\pi_{n\&EE}$	$\pi_{n\&NN}$	$\Delta^{(ZS)}$	$\Delta^{(DS)}$	$\Delta^{(DW)}$	λ_M	λ_{0W}	λ_{S0}
52	0.236	0.032	0.049	0.397	0.127	0.159	−0.017	−0.024	0.276	8.61	1.03	3.67
130	0.293	0.067	0.052	0.298	0.139	0.151	0.015	0.022	0.247	8.06	1.36	2.44
208	0.377	0.044	0.035	0.261	0.162	0.120	0.009	0.013	0.290	4.89	0.92	2.26

NOTE: For each week, we provide the estimated proportions in the principal strata; λ_M is the scaled LRT statistic if the null model assumes monotonicity of employment; λ_{S0}, λ_{0W} represent the scaled LRT statistics for the null model with constraints $\Delta^{(DW)} = 0$ and $\Delta^{(DS)} = 0$, respectively.

2. Estimated Average Wages: The authors also reported the MLE of the average wages among all PS strata under treatment and control are reported. We can see that for compliers to both training and employment, the average wages experience a great leap in short terms and stabilize in long run.

Week	$\bar{W}_{c\&EE,0}$	(SE)	$\bar{W}_{c\&EE,1}$	(SE)	$\bar{W}_{c\&EN,1}$	(SE)	$\bar{W}_{c\&NE,0}$	(SE)	$\bar{W}_{n\&EE}$	(SE)
52	5.52	(0.000)	5.80	(0.000)	7.32	(0.015)	6.80	(0.030)	6.51	(0.001)
130	6.44	(0.000)	6.69	(0.000)	9.22	(0.009)	7.22	(0.022)	7.94	(0.001)
208	7.47	(0.001)	7.76	(0.001)	9.27	(0.026)	8.99	(0.096)	8.97	(0.001)

3. Estimated means of covariates: It is very interesting for the authors to journey back to a thorough estimation of the covariates (Figure 13 in Appendix) after obtaining the causal results. The finding highlights that the distribution of covariates among noncompliers implies the reason why noncompliance happened. For example, the never-employed noncompliers ($n\&NN$) are in general less likely to be white and more likely to be female and to have children; they appear to be the right target of the program, and so their decision to not participate in the program may be partly explained by objective difficulties of participation due to family constraints, suggesting that a more flexible training schedule for them may have satisfied their requirements. This feature of the model may be helpful for policy makers to to redesign the program for better effectiveness, and the nuanced results enhance their ability to do so.

APPENDIX

STUDY 1 (BARNARD ET AL. 2003)

Family size	Treatment	PMPD	Randomized block 1	2	3	4	Subtotal	Total
Single	Scholarship	353	72	65	82	104	323	676
	Control	353	72	65	82	104	323	676
Multi	Scholarship	147	44	27	31	75	177	324
	Control	147	27	23	33	54	137	284
	Total	1,000					960	1,960

Fig. 3. Sample Sizes in the SCSF Program

Variable	Application period 1: Simple random sample	Stratified random sample	PMPD	Periods 2–5: Randomized block
Applicant's school (low/high)	−.98	0	.11	.21
Grade level	−1.63	.03	−.03	−.39
Pretest read score	−.38	.65	.48	−1.05
Pretest math score	−.51	1.17	.20	−1.37
African-American	1.80	1.68	1.59	1.74
Mother's education	.16	.14	.09	1.67
In special education	.31	1.66	−.17	.22
In gifted program	.42	−1.16	−.13	.75
English main language	−1.06	−.02	−1.03	−.44
AFDC	−.28	.49	.83	−1.57
Food stamps	−1.08	−.27	.94	−1.31
Mother works	−1.26	−.30	−1.18	.40
Educational expectations	.50	1.79	.57	.19
Children in household	−1.01	−1.75	.41	−1.02
Child born in U.S.	.49	.73	−1.40	−.69
Length of residence	.42	.71	.66	−.78
Father's work missing	1.09	.70	0	.16
Catholic religion	−1.84	−.19	−.74	−.80
Male	.88	1.22	.76	.53
Income	−.38	−.62	.74	−1.21
Age as of 4/97	−1.57	.18	−.47	−.87

Fig. 4. Design Comparisons in Balance of Background Variables: Single-Child Families. The Numbers Are Z Statistics From Comparing Observed Values of Variables Between Assignments.

Notation	Specifics	General description
Z_i	1 if i assigned treatment 0 if i assigned control	Binary indicator of treatment assignment
$D_i(z)$	1 if i receives treatment under assignment z 0 if i receives control under assignment z	Potential outcome formulation of treatment receipt
D_i	$D_i(Z_i)$	Binary indicator of treatment receipt
C_i	c if $D_i(0)=0$ and $D_i(1)=1$ n if $D_i(0)=0$ and $D_i(1)=0$ a if $D_i(0)=1$ and $D_i(1)=1$ d if $D_i(0)=1$ and $D_i(1)=0$	Compliance principal stratum: c = complier; n = never taker; a = always taker; d = defier
$Y_i(z)$	$(Y_i^{(math)}(z), Y_i^{(read)}(z))$	Potential outcomes for math and reading
Y_i	$(Y_i^{(math)}(Z_i), Y_i^{(read)}(Z_i))$	Math and reading outcomes under observed assignment
$Ry_i^{(math)}(z)$	1 if $Y_i^{(math)}(z)$ would be observed 0 if $Y_i^{(math)}(z)$ would not be observed	Response indicator for $Y_i^{(math)}(z)$ under assignment z; similarly for $Ry_i^{(read)}(z)$
$Ry_i(z)$	$(Ry_i^{(math)}(z), Ry_i^{(read)}(z))$	Vector of response indicators for $Y_i(z)$
Ry_i	$(Ry_i^{(math)}(Z_i), Ry_i^{(read)}(Z_i))$	Vector of response indicators for Y_i
W_i	$(W_{i1}, \ldots, W_{iK})$	Fully observed background and design variables
X_i	$(X_{i1}, \ldots, X_{iQ})$	Partially observed background variables
Rx_i	$(Rx_{i1}, \ldots, Rx_{iQ})$	Vector of response indicators for X_i

Fig. 5. Notation for the ith Subject

STUDY 2 (FRUMENTO ET AL. 2012)

$M_i(1)$	$M_i(0)$	Outcome
0	0	never missing
1	0	missing only under treatment
0	1	missing only under control
1	1	always missing

TABLE I

PRINCIPAL STRATA FURTHER CONDITIONED ON MISSING INDICATOR M

Observed subgroups $O(Z_i, D_i(1), S_{i,\text{obs}})$	Latent strata
$O(1,1,1) = \{i : Z_i = 1, D_i(1) = 1, S_{i,\text{obs}} = 1\}$	$c\&EE$, $c\&EN$
$O(1,1,0) = \{i : Z_i = 1, D_i(1) = 1, S_{i,\text{obs}} = 0\}$	$c\&NN$, $c\&NE$
$O(1,0,1) = \{i : Z_i = 1, D_i(1) = 0, S_{i,\text{obs}} = 1\}$	$n\&EE$
$O(1,0,0) = \{i : Z_i = 1, D_i(1) = 0, S_{i,\text{obs}} = 0\}$	$n\&NN$
$O(1,?,1) = \{i : Z_i = 1, D_i(1) =?, S_{i,\text{obs}} = 1\}$	$n\&EE$, $c\&EE$, $c\&EN$
$O(1,?,0) = \{i : Z_i = 1, D_i(1) =?, S_{i,\text{obs}} = 0\}$	$n\&NN$, $c\&NN$, $c\&NE$
$O(0,?,1) = \{i : Z_i = 0, D_i(1) =?, S_{i,\text{obs}} = 1\}$	$c\&EE$, $c\&NE$, $n\&EE$
$O(0,?,0) = \{i : Z_i = 0, D_i(1) =?, S_{i,\text{obs}} = 0\}$	$c\&EN$, $c\&NN$, $n\&NN$
$O(1,1,?) = \{i : Z_i = 1, D_i(1) = 1, S_{i,\text{obs}} =?\}$	$c\&EE$, $c\&EN$, $c\&NE$, $c\&NN$
$O(1,0,?) = \{i : Z_i = 1, D_i(1) = 0, S_{i,\text{obs}} =?\}$	$n\&EE$, $n\&NN$
$O(0,?,?) = \{i : Z_i = 0, D_i(1) =?, S_{i,\text{obs}} =?\}$	$c\&EE$, $c\&EN$, $c\&NE$, $c\&NN$, $n\&EE$, $n\&NN$

Fig. 6. Correspondence between observed subgroups and latent strata

$$
\begin{aligned}
&L(\boldsymbol{\theta}_{\text{sci}} \mid \mathbf{D(1)}, \mathbf{S}_{\text{obs}}, \mathbf{W}_{\text{obs}}, \mathbf{Z}, \mathbf{X}) \\
&\propto \prod_{i \in O(1,1,1)} \left[\pi_{i:c\&EE} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{c\&EE,1}, \sigma^2_{c\&EE,1}\right)\right. \\
&\left. + \pi_{i:c\&EN} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{c\&EN,1}, \sigma^2_{c\&EN,1}\right)\right] \\
&\times \prod_{i \in O(1,1,0)} \left[\pi_{i:c\&NE} + \pi_{i:c\&NN}\right]
\end{aligned}
$$

Fig. 7. Observed data likelihood function

$$
\begin{aligned}
&\times \prod_{i \in O(1,0,1)} \left[\pi_{i:n\&EE} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{n\&EE}, \sigma^2_{n\&EE}\right)\right] \times \prod_{i \in O(1,0,0)} \left[\pi_{i:n\&NN}\right] \\
&\times \prod_{i \in O(1,?,1)} \left[\pi_{i:c\&EE} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{c\&EE,1}, \sigma^2_{c\&EE,1}\right)\right. \\
&+ \pi_{i:c\&EN} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{c\&EN,1}, \sigma^2_{c\&EN,1}\right) \\
&\left. + \pi_{i:n\&EE} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{n\&EE,1}, \sigma^2_{n\&EE,1}\right)\right] \\
&\times \prod_{i \in O(1,?,0)} \left[\pi_{i:c\&NE} + \pi_{i:c\&NN} + \pi_{i:n\&NN}\right] \\
&\times \prod_{i \in O(0,?,1)} \left[\pi_{i:c\&EE} N_i(\mathbf{X}_i \boldsymbol{\beta}_{c\&EE,0}, \sigma^2_{c\&EE,0})\right. \\
&+ \pi_{i:c\&NE} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{c\&NE,0}, \sigma^2_{c\&NE,0}\right) \\
&\left. + \pi_{i:n\&EE} N_i\left(\mathbf{X}_i \boldsymbol{\beta}_{n\&EE}, \sigma^2_{n\&EE}\right)\right] \\
&\times \prod_{i \in O(0,?,0)} \left[\pi_{i:c\&EN} + \pi_{i:c\&NN} + \pi_{i:n\&NN}\right] \\
&\times \prod_{i \in O(1,1,?)} \left[\pi_{i:c\&EE} + \pi_{i:c\&EN} + \pi_{i:c\&NE} + \pi_{i:c\&NN}\right] \\
&\times \prod_{i \in O(1,0,?)} \left[\pi_{i:n\&EE} + \pi_{i:n\&NN}\right].
\end{aligned}
$$

Fig. 8. Observed data likelihood function (cont.)

$$\hat{\Delta}^{(ZD)} = \hat{\pi}_{c\&EE} + \hat{\pi}_{c\&EN} + \hat{\pi}_{c\&NE} + \hat{\pi}_{c\&NN} = \hat{\pi}_c$$

Fig. 9. Estimated causal effect of Z on D

$$\hat{\Delta}^{(ZS)} = \hat{\pi}_{c\&EN} - \hat{\pi}_{c\&NE}$$

Fig. 10. Estimated causal effect of Z on S

$$\hat{\Delta}^{(DS)} = \frac{\hat{\pi}_{c\&EN} - \hat{\pi}_{c\&NE}}{\hat{\pi}_c}$$

Fig. 11. Estimated causal effect on employment for compliers

$$\hat{\Delta}^{(DW)} = \frac{\sum_{i=1}^{N} \omega_i \hat{\pi}_{i:c\&EE} \exp\left\{\mathbf{X}_i \hat{\boldsymbol{\beta}}_{c\&EE,1} + \frac{1}{2}\hat{\sigma}^2_{c\&EE,1}\right\}}{\sum_{i=1}^{N} \omega_i \hat{\pi}_{i:c\&EE}} - \frac{\sum_{i=1}^{N} \omega_i \hat{\pi}_{i:c\&EE} \exp\left\{\mathbf{X}_i \hat{\boldsymbol{\beta}}_{c\&EE,0} + \frac{1}{2}\hat{\sigma}^2_{c\&EE,0}\right\}}{\sum_{i=1}^{N} \omega_i \hat{\pi}_{i:c\&EE}},$$

Fig. 12. Estimated causal effect on wages for always-employed compliers

Variable	*c&EE*	*c&EN*	*c&NE*	*c&NN*	*n&EE*	*n&NN*
Week 52						
Female	0.41	0.26	0.25	0.44	0.40	0.45
Age at baseline	18.9	19.0	19.3	18.4	19.5	18.8
White	0.34	0.40	0.34	0.20	0.33	0.23
With a partner	0.07	0.03	0.04	0.04	0.10	0.08
Has children	0.17	0.09	0.11	0.17	0.21	0.25
Education	10.2	10.2	10.1	9.8	10.6	9.9
Ever arrested	0.24	0.29	0.32	0.24	0.28	0.31
Mother's education	11.73	11.68	11.64	11.41	11.63	11.51
Father's education	11.68	12.11	11.69	11.41	11.60	11.51
Household income > 6000	0.58	0.61	0.58	0.47	0.60	0.48
Personal income > 6000	0.10	0.14	0.07	0.05	0.14	0.06
Have job	0.29	0.29	0.30	0.14	0.32	0.16
Had job, prev. yr.	0.75	0.76	0.72	0.55	0.80	0.58
Months in job, prev. yr.	4.97	5.08	5.06	2.83	5.57	3.08
Earnings, prev. yr.	3889.6	4112.4	4379.8	1973.4	4780.8	2508.2

Fig. 13. Estimated means of covariates within PS, computed using design weights and estimated membership probabilities, week 52

REFERENCES

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55. https://doi.org/10.1080/01621459.1996.10476902.

Barnard, John, Jiangtao Du, Jennifer L. Hill, and Donald B. Rubin. 1998. "A Broader Template for Analyzing Broken Randomized Experiments." *Sociological Methods & Research* 27 (2): 285–317. https://doi.org/10.1177/0049124198027002005.

Barnard, John, Constantine E Frangakis, Jennifer L Hill, and Donald B Rubin. 2003. "Principal Stratification Approach to Broken Randomized Experiments." *Journal of the American Statistical Association* 98 (462): 299–323. https://doi.org/10.1198/016214503000071.

Dempster, AP. n.d. "Laird, NM, and Rubin, DB (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. J. r." *Statist. Soc. B 39, 1* 38.

Frangakis, Constantine E., and Donald B. Rubin. 2004. "Principal Stratification in Causal Inference." *Biometrics* 58 (1): 21–29. https://doi.org/10.1111/j.0006-341X.2002.00021.x.

Frumento, Paolo, Fabrizia Mealli, Barbara Pacini, and Donald B. Rubin. 2012. "Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data." *Journal of the American Statistical Association* 107 (498): 450–66. http://www.jstor.org/stable/23239583.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. https://doi.org/10.1080/01621459.1986.10478354.