Correcting Attrition Bias using Changes-in-Changes*

Dalia Ghanem, UC Davis
Sarojini Hirshleifer, UC Riverside
Désiré Kédagni, Iowa State
Karen Ortiz-Becerra, University of San Diego

March 26, 2022

Abstract

Attrition is a common and potentially important threat to internal validity in treatment effect studies. We extend the changes-in-changes approach to identify the average treatment effect for respondents and the entire study population in the presence of attrition. Our method can be applied to randomized experiments as well as difference-in-difference designs. A simulation experiment points to the advantages of this approach relative to one of the most commonly used approaches in the literature, inverse probability weighting. Those advantages are further illustrated with an application to a large-scale randomized experiment.

Keywords: nonresponse bias, panel data, randomized experiments, difference-in-differences.
JEL Codes: C21, C23, C93.

*E-mail: dghanem@ucdavis.edu, sarojini.hirshleifer@ucr.edu, dkedagni@iastate.edu, kaortizb@ucdavis.edu.
A Stata ado file to implement the proposed corrections is in progress.

1 Introduction

Attrition is a common and potentially important source of selection bias in a range of treatment effect studies. Attrition has long been recognized as a concern in settings that rely on panel data.[1] In addition, as randomized experiments become a widely implemented methodology in applied economics, attrition tests and corrections are increasingly relevant to empirical practice (Millán and Macours, 2019; Ghanem et al., 2021). The current empirical literature relies on a wide range of approaches to correct for attrition bias. None of them, however, are specifically tailored to the panel data available in many randomized experiments with baseline (pre-treatment) outcome data as well as in difference-in-difference designs.

We propose a novel attrition correction based on the changes-in-changes (CiC) approach (Athey and Imbens, 2006). The correction is suitable for treatment effect settings where baseline outcome data are available for both respondent and attritor subpopulations. The proposed method relies on a key insight: Under the extended CiC conditions, there are two transformations that relate the baseline outcome distribution to the distributions of the treated and untreated potential outcomes in the post-treatment period (Lemma 1). These two transformations are identical for all treatment-response subpopulations and can be identified using the treatment and control respondents. Using these transformations, we can therefore identify not only the counterfactual distribution for the treatment and control respondents, but also the distribution of both potential outcomes for the treatment and control attritors. The identification of the average treatment effects for the respondents (ATE-R) as well as the entire study population (ATE) then follows immediately. Since the CiC assumptions do not require random assignment, our approach is not only suitable for randomized experiments with baseline outcome data, but also difference-in-difference designs.

The CiC conditions require that the outcome is monotonic in a scalar unobservable and that the distribution of this unobservable conditional on treatment and response status is stable over time. Despite this assumption, the distribution of the outcome can vary across time, since it can be a time-varying function of the unobservable. For outcomes that are strictly monotonic in the unobservable, the parameters of interest are point-identified. For outcomes that only satisfy weak monotonicity, bounds can still be obtained for these objects (Athey and Imbens, 2006).

We conduct a simulation study to examine the performance of the CiC corrections in finite samples. We compare their performance with an existing procedure that has the same data requirements, specifically the inverse probability weighting (IPW) correction that assumes unconfoundedness conditional on the baseline outcome. We focus on the IPW approach as a comparator to the CiC approach since it is the most widely-used approach in the literature that also can be used to obtain point estimates of the ATE and ATE-R. In our simulation study, we allow selection into response to depend on the unobservables that determine the potential outcomes, consistent with findings in the empirical literature (Behagel et al., 2015). In all variants of the design we consider, the CiC corrections for the ATE-R and ATE have negligible bias, whereas the performance of the IPW corrections depends on whether the design satisfies missingness-at-random or random assignment conditional on response

[1] See, for example, Fitzgerald et al. (1998), van den Berg and Lindeboom (1998) and Ziliak and Kniesner (1998).

status.[2]

An empirical application revisiting an outcome from the *Progresa* cash transfer program further illustrates the performance of the CiC correction and how it differs from the IPW correction. We study the same outcome across two follow-up surveys that were implemented 12 and 18 months after the study began. In the follow-up at 18 months, we find the CiC-corrected estimate of the ATE is significantly different from both the uncorrected treatment effect estimate as well as the IPW-corrected estimates. Thus, the CiC approach suggests that a correction is required while the IPW approach does not. A test for attrition bias applied to this outcome finds that internal validity for the population is violated.[3] Thus, only the CiC approach is consistent with the findings of the attrition bias tests. In contrast, in the follow-up at 12 months, neither the CiC- nor IPW-corrected estimates are significantly different from each other or the uncorrected treatment effect estimate. Again, the CiC-corrected estimate is consistent with the attrition bias test, since in this case, that test does not reject internal validity.

This paper contributes to the literature on attrition corrections in treatment effect models which build on seminal work on sample selection (Heckman, 1976, 1979). It provides a tractable approach that exploits the presence of baseline outcome data. Millán and Macours (2019) survey the field experiment literature and find that the most widely-used methods include IPW and various bounding approaches.[4] The former approach requires unconfoundedness which is hard to justify in many settings. Horowitz and Manski (2000) provide bounds on the average treatment effect that require minimal assumptions, however the bounds are typically large in practice. Lee (2009) proposes bounds on the average treatment effect for the always-responders, a subset of the respondent subpopulation. Behagel et al. (2015) show that tighter nonparametric bounds on the average treatment effect for a subpopulation of respondents can be obtained by exploiting additional data, specifically the number of calls required to obtain a response. Our approach relies on the presence of baseline outcome data. We show point-identification is possible for continuous outcomes that satisfy the strict monotonicity assumption via the CiC approach, not only for the respondents, but also for the entire study population. Our method can be extended to outcomes that satisfy the weak monotoncity assumption in order to obtain bounds (Remark 2). Another advantage of exploiting the panel structure is that it can be implemented in settings without simple randomization, such as stratified randomization and difference-in-difference designs.

2 Attrition Corrections via Changes-in-Changes

2.1 The model and parameters of interest

Let Y_t and D_t denote the observed outcome and treatment status in period t. The treatment path is denoted by $D = (D_0, D_1)$. To simplify notation, we denote the group membership by G, where $G = 1$ for the treatment group which receives the treatment path $D = (0, 1)$, and $G = 0$ for the control group

[2] This is a byproduct of our design violating the unconfoundedness assumption required for IPW.

[3] We implement the test of attrition bias proposed in Ghanem et al. (2021).

[4] Millán and Macours (2019) also propose a modified version of the inverse probability weighting approach that uses additional data from an intense tracking phase.

which receives the treatment path $D = (0, 0)$, noting that $G = D_1$. We consider the following model.

$$\begin{cases} Y_t &= \mu_0(0, U_0)\mathbb{1}\{t = 0\} + \mu_1(G, U_1)\mathbb{1}\{t = 1\} \\ R &= GR(1) + (1 - G)R(0) \end{cases} \tag{1}$$

where the outcome variable in the baseline period ($t = 0$) is always observed, whereas the outcome variable in the follow-up period ($t = 1$) is observed only if $R = 1$. Specifically, R is an indicator for response in the follow-up period. We assume there are no response issues in the baseline period ($t = 0$).

The variable U_t denotes the unobserved heterogeneity in the outcome, while $R(1)$ and $R(0)$ are the potential responses when the individual is treated or untreated, respectively. As a result, there are four response-types:

$(R(0), R(1)) = (0, 0)$: Never-Responders, $(R(0), R(1)) = (0, 1)$: Treatment-Only Responders,

$(R(0), R(1)) = (1, 0)$: Control-Only Responders, $(R(0), R(1)) = (1, 1)$: Always-Responders.

Let $Y_0(0) = \mu_0(0, U_0)$ denote the untreated potential outcome in the baseline period ($t = 0$) and $Y_1(g) = \mu_1(g, U_1)$ denote the potential outcome for $g = 0, 1$ for the follow-up period ($t = 1$). Let F_Y denote the cumulative distribution function of a random variable Y.

In this paper, we are interested in identifying the average treatment effects for the treated and untreated respondents (ATT-R and ATU-R), for the respondents (ATE-R), and for the study population (ATE), defined as follows:

$$\begin{aligned} \text{ATT-R} &= E[Y_1(1) - Y_1(0)|G = 1, R = 1], \\ \text{ATU-R} &= E[Y_1(1) - Y_1(0)|G = 0, R = 1], \\ \text{ATE-R} &= E[Y_1(1) - Y_1(0)|R = 1], \\ \text{ATE} &= E[Y_1(1) - Y_1(0)]. \end{aligned}$$

We obtain a random sample of the vector (G, R, Y_0, Y_1^*), where all random variables are observed except Y_1^* which is only observed when $R = 1$. In order to identify the above parameters, we use the following assumptions.

Assumption 1 *(Distribution of Unobservable)*

1. $U_0|G, R \stackrel{d}{=} U_1|G, R$.

2. $F_{U_t|G,R}(u|g, 1)$ *is continuous and strictly increasing in* u *for* $g = 0, 1$.

Assumption 2 *(Monotonicity of Structural Function)*

1. $\mu_t(0, u)$ *is strictly increasing in* u *for* $t = 0, 1$.

2. $\mu_1(1, u)$ *is strictly increasing in* u.

Assumption 1.1 states that the distribution of the unobserved heterogeneity U_t conditional on treatment and response status is invariant over time. It is similar to Assumption 3.3 in Athey and Imbens (2006), except that we condition on the nonrandom response status. While Assumption 1.1 requires the distribution of unobserved heterogeneity to be stable across time within treatment-response subpopulations, it allows its distribution to vary arbitrarily across these subpopulations. We further impose Assumption 1.2 to ensure that the distribution function of $Y_t|G = g, R = 1$ is invertible for $g = 0, 1$. Assumption 2.1 implies that for each period, the untreated potential outcome is strictly increasing in the unobserved heterogeneity. It is the same as Assumption 3.2 in Athey and Imbens (2006). Assumption 2.2 requires that the treated potential outcome is strictly increasing in the unobserved heterogeneity. These two monotonicity assumptions are the main driver of our identification results as we show in Lemma 1.

It is important to discuss the implications of Assumptions 1 and 2 for the outcome variable. Despite the time invariance of the distribution of the unobserved heterogeneity (Assumption 1.1), the distribution of the potential outcomes can change over time, since Assumption 2 does not restrict the variability of the structural function across time. The strict monotonicity of the structural function (Assumption 2) together with Assumption 1.2 implies that the outcome of the treatment and control respondents have continuous and strictly increasing distributions. However, the CiC approach can be extended to accommodate discrete outcomes as we discuss in Remark 2.[5]

Assumption 3 *(Random Assignment)* $(Y_0(0), Y_1(0), Y_1(1), R(0), R(1)) \perp G$.

Assumption 3 states that the individuals are randomly assigned to treatment ($G = 1$) and control ($G = 0$) groups. This assumption applies to randomized experiments with simple and cluster randomization designs. Below we provide identification results with and without random assignment.

2.2 Identification results

In this section, we outline how the CiC identification approach can be applied to identify our parameters of interest. We provide results both for the respondent subpopulation and study population. Let $\mathbb{Y}$ denote the support of the random variable Y, and $\mathbb{Y}^{d,t}_{g,r}$ denote the support of $Y_t(d)|G = g, R = r$. Define $F_Y^{-1}(q) = \inf\{y \in \mathbb{Y}|F_Y(y) \geqslant q\}$.

Before we proceed to our main identification results, the following lemma helps us understand how Assumptions 1 and 2 can allow us to "extrapolate" not only to the respondent subpopulations but also to the attritor subpopulations.

Lemma 1 *Suppose that Assumption 1 and 2.1 hold, then:*

1. For $g = 0, 1$, $r = 0, 1$,

$$(i)\ F_{Y_1(0)|G=g,R=r}(y) = F_{Y_0(0)|G=g,R=r}(T_0(y)) \ \text{for}\ y \in \mathbb{Y}^{0,1}_{g,r},$$
$$(ii)\ T_0(y) = F^{-1}_{Y_0|G=0,R=1}(F_{Y_1|G=0,R=1}(y)) \qquad \text{for}\ y \in \mathbb{Y}^{0,1}_{0,1},$$

[5] Theoretically, the strict monotonicity condition can hold in the discrete outcome case if the unobservable is also discrete. However, this would rule out the most well-known models for limited dependent variables, such as linear-index models as discussed in Athey and Imbens (2006).

where $T_0(y) = \mu_0(0, \mu_1^{-1}(0; y))$ *and* $\mu_1^{-1}(0; y)$ *denotes the inverse of* $\mu_1(0, u)$.

2. *Suppose further Assumption 2.2 holds. For* $g = 0, 1$, $r = 0, 1$,

$$(i)\ F_{Y_1(1)|G=g,R=r}(y) = F_{Y_0(0)|G=g,R=r}(T_1(y)) \ \textit{for}\ y \in \mathbb{Y}_{g,r}^{1,1},$$
$$(ii)\ T_1(y) = F^{-1}_{Y_0|G=1,R=1}(F_{Y_1|G=1,R=1}(y)) \qquad \textit{for}\ y \in \mathbb{Y}_{1,1}^{1,1},$$

where $T_1(y) = \mu_0(0, \mu_1^{-1}(1; y))$ *and* $\mu_1^{-1}(1; y)$ *denotes the inverse of* $\mu_1(1, u)$.

The proofs of the above lemma and all propositions are given in Appendix A. Lemma 1.1(i) shows that under the time invariance assumption (Assumption 1.1) and the strict monotonicity of the untreated potential outcome (Assumption 2.1), the distribution of the untreated potential outcome for any treatment-response subpopulation in the follow-up period at a given y equals the distribution of the untreated potential outcome of that subpopulation in the baseline period evaluated at $T_0(y)$, where the transformation, $T_0(\cdot)$, is the same for all treatment-response subpopulations. Since we observe the distribution of the untreated potential outcome of the control respondents in both baseline and follow-up periods, Lemma 1.1(ii) shows that we can identify $T_0(y)$ for $y \in \mathbb{Y}_{0,1}^{0,1}$ using the control respondents by the continuity and strict monotonicity of the outcome distribution (Assumption 1.2).

If we also impose the strict monotonicity assumption on the treated potential outcome (Assumption 2.2), Lemma 1.2(i) shows that the treated potential outcome distribution for any treatment-response subpopulation at a given value y equals the distribution of the untreated potential outcome of that subpopulation in the baseline period evaluated at $T_1(y)$, where the transformation, $T_1(\cdot)$, is the same for all treatment-response subpopulations. Since we observe the untreated potential outcome in the baseline period and the treated potential outcome in the follow-up period for the treatment respondents, we can use them to identify $T_1(y)$ for $y \in \mathbb{Y}_{1,1}^{1,1}$ (Lemma 1.2(ii)).

In sum, Lemma 1 shows that we can use the control and treatment respondents to identify $T_0(y)$ and $T_1(y)$, respectively, on their respective support. Since $T_0(y)$ and $T_1(y)$ are the same for all subpopulations, the identification of the distribution of an unobserved potential outcome for a given treatment-response subpopulation follows immediately assuming that we can observe the baseline outcome distribution for this subpopulation and that additional support conditions hold. We finally note that Lemma 1 does not require random assignment. This allows us to provide identification results for our parameters of interest without random assignment (Assumption 3).

2.2.1 Respondents

The following proposition establishes the identification of the ATT-R and ATU-R under Assumptions 1 and 2. The identification of the ATE-R is implied, since it is a population weighted average of the two. Let $\mathbb{U}_{g,r}$ denote the support of $U_0|G = g, R = r$ for $g = 0, 1$, and $r = 0, 1$

Proposition 1 (Changes-in-Changes on Respondents) *Suppose that Assumptions 1 and 2.1 hold, then:*

1. If $\mathbb{U}_{1,1} \subseteq \mathbb{U}_{0,1}$ *holds, then*

$$F_{Y_1(0)|G=1,R=1}(y) = F_{Y_0|G=1,R=1}(F^{-1}_{Y_0|G=0,R=1}(F_{Y_1|G=0,R=1}(y))) \quad for \quad y \in \mathbb{Y}^{0,1}_{1,1}, \tag{2}$$

$$ATT\text{-}R = E[Y_1|G=1,R=1] - E[F^{-1}_{Y_1|G=0,R=1}(F_{Y_0|G=0,R=1}(Y_0))|G=1,R=1]. \tag{3}$$

2. If Assumption 2.2 and $\mathbb{U}_{0,1} \subseteq \mathbb{U}_{1,1}$ *hold, then*

$$F_{Y_1(1)|G=0,R=1}(y) = F_{Y_0|G=0,R=1}(F^{-1}_{Y_0|G=1,R=1}(F_{Y_1|G=1,R=1}(y)) \quad for \quad y \in \mathbb{Y}^{1,1}_{0,1}, \tag{4}$$

$$ATU\text{-}R = E[F^{-1}_{Y_1|G=1,R=1}(F_{Y_0|G=1,R=1}(Y_0))|G=0,R=1] - E[Y_1|G=0,R=1]. \tag{5}$$

The proof of the above proposition follows from Lemma 1 and is given in Appendix A. The identification of the ATT-R and ATU-R is a direct application of CiC conditional on $R=1$. Similar to the standard CiC identification, the above result does not require random assignment. Indeed, since individuals select into response in the follow-up period, once we condition on response status treatment is no longer randomly assigned.

Remark 1 (Identification of ATE-R) *The above proposition implies that the ATE-R is identified, since*

$$ATE\text{-}R = P(G=1|R=1)ATT\text{-}R + P(G=0|R=1)ATU\text{-}R \tag{6}$$

Remark 2 (Discrete Outcomes) *While we focus our identification results on continuous outcomes, following Athey and Imbens (2006) we can provide bounds for discrete outcomes. Define* $F^{(-1)}_Y(q) = \sup\{y \in \mathbb{Y} \cup \{-\infty\} : F_Y(y) \leqslant q\}$. *Suppose* $\mu_t(d,u)$ *is nondecreasing in* u. *If Assumptions 1 holds and* $\mathbb{U}_{1,1} \subseteq \mathbb{U}_{0,1}$, *then*

$$F^{LB}_{Y_1(0)|G=1,R=1}(y) \leqslant F_{Y_1(0)|G=1,R=1}(y) \leqslant F^{UB}_{Y_1(0)|G=1,R=1}(y)$$

where $F^{LB}_{Y_1(1)|G=0,R=1}(y) = F_{Y_0|G=0,R=1}(F^{(-1)}_{Y_0|G=1,R=1}(F_{Y_1|G=1,R=1}(y))$, *and*
$F^{UB}_{Y_1(1)|G=0,R=1}(y) = F_{Y_0|G=0,R=1}(F^{-1}_{Y_0|G=1,R=1}(F_{Y_1|G=1,R=1}(y))$.

Similar results hold for $F_{Y_1(1)|G=0,R=1}(y)$ *under Assumptions 1, the weak monotonicity of* $\mu_t(1,u)$, *and* $\mathbb{U}_{0,1} \subseteq \mathbb{U}_{1,1}$.

2.2.2 Study Population

In this section, we present two identification results for the study population. The first requires random assignment, whereas the latter does not. As a result, the latter can be used in experimental settings with stratified or other complex randomization schemes as well as in difference-in-difference designs.

Under random assignment, we have $ATE = E[Y_1(1)|G=1] - E[Y_1(0)|G=0]$. Using the law of iterated expectations, we have

$$\begin{aligned} E[Y_1(0)|G=0] &= P(R=1|G=0)E[Y_1|G=0,R=1] + P(R=0|G=0)E[Y_1(0)|G=0,R=0], \\ E[Y_1(1)|G=1] &= P(R=1|G=1)E[Y_1|G=1,R=1] + P(R=0|G=1)E[Y_1(1)|G=1,R=0]. \end{aligned}$$

The only unobservable objects on the right-hand side of the above equations are the average outcomes of the control and treatment attritors, $E[Y_1(0)|G=0,R=0]$ and $E[Y_1(1)|G=1,R=0]$. The following proposition provides sufficient conditions such that we can apply Lemma 1.1 and 1.2 to identify $F_{Y_1(0)|G=0,R=0}$ and $F_{Y_1(1)|G=1,R=0}$, respectively, and thereby their expectations.

Proposition 2 (Identification of the ATE under Random Assignment) *Suppose Assumptions 1, 2 and 3 hold. Suppose further that* $\mathbb{U}_{0,1} = \mathbb{U}_{1,0}$, $\mathbb{U}_{1,0} \subseteq \mathbb{U}_{1,1}$, $\mathbb{U}_{0,0} \subseteq \mathbb{U}_{0,1}$.

$$\begin{aligned} ATE =& P(R=1|G=1)E[Y_1|G=1,R=1] + P(R=0|G=1)E[Y_1(1)|G=1,R=0] \\ & - (P(R=1|G=0)E[Y_1|G=0,R=1) + P(R=0|G=0)E[Y_1(0)|G=0,R=0]) \end{aligned}$$

where

$E[Y_1(1)|G=1,R=0] = E[F^{-1}_{Y_1|G=1,R=1}(F_{Y_0|G=1,R=1}(Y_0))|G=1,R=0]$,

$E[Y_1(0)|G=0,R=0] = E[F^{-1}_{Y_1|G=0,R=1}(F_{Y_0|G=0,R=1}(Y_0))|G=0,R=0]$.

Without random assignment, the law of iterated expectations allows us to write $E[Y_1(d)]$ as follows:

$$\begin{aligned} E[Y_1(d)] \quad = \quad & P(G=1,R=1)E[Y_1(d)|G=1,R=1] + P(G=0,R=1)E[Y_1(d)|G=0,R=1] \\ & + P(G=1,R=0)E[Y_1(d)|G=1,R=0] + P(G=0,R=0)E[Y_1(d)|G=0,R=0]. \end{aligned}$$

For $d=0$, the only terms that are observable on the right-hand side are the probabilities as well as the expected potential outcome without the treatment for the control respondents, $E[Y_1(0)|G=0,R=1]$. Therefore, in order to identify $E[Y_1(0)]$, it remains to identify the distributions of the untreated potential outcome for all remaining subpopulations, $F_{Y_1(0)|G=1,R=0}$, $F_{Y_1(0)|G=1,R=1}$, and $F_{Y_1(0)|G=0,R=0}$. Similarly, for $d=1$, the only terms that are observable on the right-hand side are the probabilities as well as the expected potential outcome with the treatment for the treatment respondents, $E[Y_1(1)|G=1,R=1]$. As a result, in order to identify $E[Y_1(1)]$, it remains to identify the distribution of the treated potential outcome for all remaining subpopulations, $F_{Y_1(1)|G=1,R=0}$, $F_{Y_1(1)|G=0,R=1}$, and $F_{Y_1(1)|G=0,R=0}$.

The next proposition provides sufficient conditions such that we can apply Lemma 1.1 to identify $F_{Y_1(0)|G=1,R=0}$, $F_{Y_1(0)|G=1,R=1}$, and $F_{Y_1(0)|G=0,R=0}$ as well as Lemma 1.2 to identify $F_{Y_1(1)|G=1,R=0}$, $F_{Y_1(1)|G=0,R=1}$, and $F_{Y_1(1)|G=0,R=0}$. The identification of the ATE follows.

Proposition 3 (Identification of the ATE without Random Assignment) *Suppose Assumptions 1 and 2 hold. Suppose further that* $\mathbb{U}_{g,r} = \mathbb{U}\ \forall (g,r) \in \{0,1\}^2$.
Then,

$$ATE = P(R=1,G=1)ATT\text{-}R + P(R=0,G=1)ATT\text{-}A + P(R=1,G=0)ATU\text{-}R + P(R=0,G=0)ATU\text{-}A,$$

where

$ATT\text{-}R = E[Y_1|G=1,R=1] - E[F^{-1}_{Y_1|G=0,R=1}(F_{Y_0|G=0,R=1}(Y_0))|G=1,R=1]$,

$ATT\text{-}A = E[F^{-1}_{Y_1|G=1,R=1}(F_{Y_0|G=1,R=1}(Y_0))|G=1,R=0] - E[F^{-1}_{Y_1|G=0,R=1}(F_{Y_0|G=0,R=1}(Y_0))|G=1,R=0]$,

$ATU\text{-}R = E[F^{-1}_{Y_1|G=1,R=1}(F_{Y_0|G=1,R=1}(Y_0))|G=0,R=1] - E[Y_1|G=0,R=1]$,

$ATU\text{-}A = E[F^{-1}_{Y_1|G=1,R=1}(F_{Y_0|G=1,R=1}(Y_0))|G=0,R=0] - E[F^{-1}_{Y_1|G=0,R=1}(F_{Y_0|G=0,R=1}(Y_0))|G=0,R=0]$.

Proposition 3 has two main practical implications. First, it extends our approach to identify the ATE in settings without (simple) random assignment, such as stratified randomization and difference-in-difference designs. In addition, it allows us to provide a testable implication of the structural assumptions imposed on the potential outcomes under (simple) random assignment. Indeed, under this condition, we must have $F_{Y_1(d)|G=1} = F_{Y_1(d)|G=0}$ for all $d = 0, 1$, that is,

$$\begin{aligned}&P(R=1|G=1)F_{Y_1(d)|G=1,R=1}(y) + P(R=0|G=1)F_{Y_1(d)|G=1,R=0}(y)\\&= P(R=1|G=0)F_{Y_1(d)|G=0,R=1}(y) + P(R=0|G=0)F_{Y_1(d)|G=0,R=0}(y),\end{aligned}$$

for all y and all d. Since each of the distributions involved in the equality is identified, we can test this condition.[6]

2.3 Selection into Response and Time Invariance

In this section, we provide an example of a model for response that implies the time invariance condition required for our identification results.

Suppose the response in the follow-up period depends on the unobservables that determine the outcome in the baseline and follow-up periods as well as treatment status, specifically

$$R = \varphi(U_0, U_1, G).$$

We provide conditions on the response equation and the marginal distribution of (U_0, U_1) that imply time invariance conditional on G and R (Assumption 1.1),

$$P(U_0 \leqslant u|G, R) = P(U_1 \leqslant u|G, R). \tag{7}$$

Note that assuming $P(R = r|G) \neq 0$ for $r = 0, 1$, then for $t = 0, 1$

$$P(U_t \leqslant u|G, R = r) = \frac{P(U_t \leqslant u, \varphi(U_0, U_1, G) = r|G)}{P(\varphi(U_0, U_1, G) = r|G)} \tag{8}$$

Hence, time homogeneity of U_t given G and R holds iff $P(U_t \leqslant u, \varphi(U_0, U_1, G) = r|G)$ is time homogeneous. We have

$$\begin{aligned}P(U_t \leqslant u, \varphi(U_0, U_1, G) = r|G) &= P(\varphi(U_0, U_1, G) = r|U_t \leqslant u, G)F_{U_t|G}(u)\\&= \int_{-\infty}^{u} P(\varphi(U_0, U_1, G) = r|U_t = v, G)dF_{U_t|G}(v),\end{aligned} \tag{9}$$

where $F_{U_t|G}$ denotes the conditional distribution of U_t given G, and the first equality holds from Bayes' rule.

The following proposition provides conditions on $\varphi(\cdot)$ and the distribution of the unobservables that imply the time invariance condition.

Proposition 4 *Let $R = \varphi(U_0, U_1, G)$ and $0 < P(R = r|G) < 1$ for $r = 0, 1$. Suppose that the mapping*

[6] We must also have $F_{Y_0(0)|G=1}(y) = F_{Y_0(0)|G=0}(y)$ for all y, which is the standard balance test.

$(u_0, u_1) \longmapsto \varphi(u_0, u_1, .)$ is symmetric in its arguments (u_0, u_1), and that the joint distribution of U_0 and U_1 given G is exchangeable. Then, $U_0|G, R \overset{d}{=} U_1|G, R$ (Assumption 1.1).

The proof of the above proposition is given in Appendix A. The proposition establishes that for the time invariance condition to hold conditional on G and R, it is sufficient that selection into response be symmetric in U_0 and U_1 and the distribution of (U_0, U_1) conditional on G be exchangeable in U_0 and U_1.[7]

Under random assignment, the exchangeability condition in Proposition 4 can be imposed on the marginal distribution of (U_0, U_1) instead of $(U_0, U_1)|G$. As a result, in the context of randomized experiments, Proposition 4 shows that Assumption 1.1 is consistent with selection into response depending on both U_0 and U_1, albeit in a symmetric way. We emphasize, however, that the conditions in Proposition 4 are merely sufficient, and therefore they do not characterize all selection mechanisms that could be consistent with the time invariance condition required for our attrition correction to be valid.

In the next section, we use these conditions to provide a simulation design that includes a model for selection into response consistent with the conditions required for our CiC attrition corrections to be consistent.

3 Simulation Study

In this section, we examine the finite-sample performance of the CiC attrition corrections. We also report the results for IPW corrections that rely on unconfoundedness conditional on baseline outcome data, $(Y_1(0), Y_1(1)) \perp R|Y_0$, given their wide use in empirical work as well as their connection to CiC (Millán and Macours, 2019; Athey and Imbens, 2006).

The simulation design is presented in Panel A in Table 1. Treatment is randomly assigned with probability 0.5. In accordance with Assumption 2, the potential outcome with and without the treatment are both strictly monotonic in a scalar unobservable, U_t. This scalar unobservable has mean zero and consists of a sum of a time-invariant and a time-varying component, α and $\sigma\eta_t$, respectively, where the latter is identically distributed across time.[8] Since U_t has mean zero, the ATE equals β_1, which is set to be a quarter of a standard deviation of the potential outcome without the treatment. The other parameter in $Y_t(1)$, β_2, determines the extent of treatment effect heterogeneity. Response in period 1, R, is determined by a threshold model in V where the threshold depends on treatment status, G, if $a_0 \neq a_1$, where a_0 and a_1 may be interpreted as the cost to response in the control and treatment group, respectively. The unobservable that determines response, V, is a sum of the average unobservable, $\bar{U}$, and idiosyncratic shock, η. If b equals zero, then missingness is at random. The response equation exhibits monotonicity. Specifically, if we set $a_0 > a_1$, then our population consists of always-responders, treatment-only responders and never-responders. If $a_0 = a_1$, then our population consists of always-responders and never-responders only.

[7] It is worth noting that the exchangeability condition implies the time invariance of the distribution of U_t conditional on G, specifically $U_0|G \overset{d}{=} U_1|G$.

[8] Via numerical evaluation, we ensure that U_t in this design satisfies the time homogeneity assumption conditional on $R(0), R(1)$, which implies Assumption 1.1.

Table 1: Simulation Design

Panel A. Data-Generating Process

Outcome:	$Y_{it}(0) = U_{it}$, $Y_{it}(1) = \beta_1 + \beta_2 U_{it} + U_{it}$ for $t = 0, 1$ where $\beta_1 = 0.25\sigma_{Y_{i1}(0)}$.
Treatment:	$G_i \overset{i.i.d.}{\sim} Bernoulli(0.5)$, $D_{i0} = 0$, $D_{i1} = G$.
Response:	$R_i = 1\{V_i \geqslant a_0(1-G_i) + a_1 G_i\}$, where $V_i = b\bar{U}_i + \epsilon$, $\bar{U}_i = 0.5(U_{i0} + U_{i1})$ and $\epsilon_i \overset{i.i.d.}{\sim} N(0,1)$.
Unobservables:	$\begin{cases} U_{it} = \alpha_i + \sigma\eta_{it},\ t = 0, 1;\ (\alpha_i, \eta_{i0}, \eta_{i1}) \perp \epsilon_i,\ \alpha_i \perp (\eta_{i0}, \eta_{i1}), \\ \alpha_i \overset{i.i.d.}{\sim} N(0,1) \\ \begin{pmatrix} \eta_{i0} \\ \eta_{i1} \end{pmatrix} \overset{i.i.d.}{\sim} N(0, I_2). \end{cases}$

Panel B. Variants of the Design

	Design I	Design II	Design III
Missing-at-random	No	No	Yes
$(U_0, U_1) \perp (R(0), R(1))$	$b = 1$	$b = 1$	$b = 0$
Differential Attrition Rates	Yes	No	Yes
	$a_0 = \Phi^{-1}\left(\frac{0.3}{\sigma_V}\right)$	$a_0 = \Phi^{-1}\left(\frac{0.25}{\sigma_V}\right)$	$a_0 = \Phi^{-1}\left(\frac{0.3}{\sigma_V}\right)$
	$a_1 = \Phi^{-1}\left(\frac{0.2}{\sigma_V}\right)$	$a_1 = \Phi^{-1}\left(\frac{0.25}{\sigma_V}\right)$	$a_1 = \Phi^{-1}\left(\frac{0.2}{\sigma_V}\right)$

Note: $\sigma_{Y_1(0)} = \sqrt{Var(Y_{i1}(0))}$, $\sigma_V = \sqrt{Var(V)}$, $\Phi^{-1}(q)$ denotes the qth quantile of the standard normal distribution, and I_2 denotes the 2×2 identity matrix.

We consider three variants of our design presented in Panel B of Table 1. All variants feature the same overall attrition rate (25%), but differ in the values of the parameters that determine response (a_0, a_1 and b). In Design I, a_0 and a_1 equal the 30^{th} and 20^{th} percentiles of the distribution of V, respectively, whereas we set $b = 1$. The choices of a_0 and a_1 imply 30% and 20% attrition rates in the treatment and control groups, respectively, yielding an overall attrition rate of 25%. Setting $b = 1$ ensures that V is a function of $\bar{U}$ in addition to ϵ, which violates missing-at-random. As a result, Design I violates internal validity, whether for the respondents or the study population. Design II maintains $b = 1$ and thereby the violation of missing-at-random, but sets both a_0 and a_1 to equal the 25^{th} percentile of the distribution of V. As a result, while violating internal validity for the study population, Design II satisfies internal validity for the respondents, since the population consists of always-responders and never-responders in this design. Finally, in Design III, we set $b = 0$, thereby satisfying missing-at-random, whereas we maintain a_0 and a_1 at the same values as in Design I, which leads to differential attrition rates.

Table 2 reports the simulation results for the CiC and IPW corrections for $n = 2,000$ and $\sigma = 2$ for constant and heterogeneous treatment effects, $\beta_2 = 0$ and $\beta_2 = 1$, respectively. We report the true values of the different objects, since they may vary across designs. For the CiC corrections, we report the simulation results for the four objects we examine in the previous section: ATT-R, ATU-R, ATE-R, and ATE. Note that under the constant treatment effect setting ($\beta_2 = 0$), all of these objects are equal by construction. We report the IPW correction for the ATE-R as well as the ATE following the approach in Huber (2012) with and without trimming. For each correction, we report the mean, bias, standard deviation (SD) and root mean squared error (RMSE). We also report the difference in

Table 2: Simulation Results ($n = 2,000$, $\sigma = 2$)

	Constant Treatment Effects ($\beta_2 = 0$)						Heterogeneous Treatment Effects ($\beta_2 = 1$)					
	True Value	Estim.	Mean	Bias	SD	RMSE	True Value	Estim.	Mean	Bias	SD	RMSE
			Design I						Design I			
		$\widehat{\Delta}_R$	0.34		0.10			$\widehat{\Delta}_R$	0.86	-0.32	0.16	0.36
ATT-R	0.56	CiC	0.56	0.00	0.15	0.15	1.08	CiC	1.08	0.00	0.19	0.19
ATU-R	0.56	CiC	0.56	0.00	0.15	0.15	1.30	CiC	1.30	0.00	0.27	0.30
ATE-R	0.56	CiC	0.56	0.00	0.15	0.15	1.19	CiC	1.18	0.00	0.23	0.23
		IPW1	0.34	-0.21	0.10	0.24		IPW1	0.87	-0.32	0.16	0.35
		IPW2	0.34	-0.21	0.10	0.24		IPW2	0.87	-0.32	0.16	0.35
ATE	0.56	CiC	0.53	-0.03	0.13	0.13	0.56	CiC	0.57	0.01	0.20	0.20
		IPW1	0.29	-0.27	0.11	0.29		IPW1	0.85	0.29	0.17	0.34
		IPW2	0.27	-0.29	0.13	0.32		IPW2	0.79	0.24	0.20	0.31
			Design II						Design II			
		$\widehat{\Delta}_R$	0.56		0.11			$\widehat{\Delta}_R$	1.20	0.00	0.17	0.17
ATT-R	0.56	CiC	0.56	0.00	0.15	0.15	1.20	CiC	1.19	0.00	0.20	0.20
ATU-R	0.56	CiC	0.56	0.00	0.15	0.15	1.19	CiC	1.19	0.00	0.27	0.20
ATE-R	0.56	CiC	0.56	0.00	0.15	0.15	1.19	CiC	1.19	0.00	0.23	0.23
		IPW1	0.56	0.00	0.11	0.11		IPW1	1.20	0.00	0.17	0.17
		IPW2	0.56	0.00	0.11	0.11		IPW2	1.20	0.00	0.17	0.17
ATE	0.56	CiC	0.56	0.00	0.13	0.13	0.56	CiC	0.61	0.05	0.20	0.21
		IPW1	0.56	0.00	0.12	0.12		IPW1	1.25	0.69	0.18	0.72
		IPW2	0.56	0.00	0.13	0.13		IPW2	1.25	0.69	0.20	0.72
			Design III						Design III			
		$\widehat{\Delta}_R$	0.56		0.12			$\widehat{\Delta}_R$	0.56	0.00	0.18	0.18
ATT-R	0.56	CiC	0.56	0.00	0.15	0.15	0.56	CiC	0.56	0.00	0.20	0.20
ATU-R	0.56	CiC	0.56	0.00	0.15	0.15	0.56	CiC	0.57	0.01	0.28	0.20
ATE-R	0.56	CiC	0.56	0.00	0.15	0.15	0.56	CiC	0.56	0.01	0.23	0.23
		IPW1	0.56	0.00	0.12	0.12		IPW1	0.56	0.00	0.18	0.18
		IPW2	0.56	0.00	0.12	0.12		IPW2	0.56	0.00	0.18	0.18
ATE	0.56	CiC	0.56	0.00	0.13	0.13	0.56	CiC	0.57	0.01	0.19	0.20
		IPW1	0.56	0.00	0.12	0.12		IPW1	0.56	0.00	0.18	0.18
		IPW2	0.56	0.00	0.12	0.12		IPW2	0.56	0.00	0.18	0.18

Notes: The simulation results provided above are based on 1,000 simulation replications. The difference in group means between treatment and control respondents in the follow-up period is denoted by $\widehat{\Delta}_R \equiv \frac{\sum_{i=1}^n Y_1 GR}{\sum_{i=1}^n GR} - \frac{\sum_{i=1}^n Y_1 (1-G) R}{\sum_{i=1}^n (1-G) R}$. CiC denotes the Changes-in-Changes estimator of the relevant object. IPW1 (IPW2) refer to the IPW correction without (with) trimming. Following Huber (2012), we trim observations with response propensity score below 5% and treatment propensity score below (above) 5% (95%).

means between the treatment and control respondents, $\hat{\Delta}_R$.

In all three designs we consider, the CiC correction has little or no bias for the relevant object of interest. The performance of the IPW correction however depends on the design in question. In Design I, the IPW correction for both the ATE-R and the ATE exhibit a substantial bias regardless of treatment effect heterogeneity. In Design II, where internal validity for the respondents hold, no correction for the ATE-R is required. As a result, the IPW correction for the ATE-R has negligible bias, but the IPW correction for the ATE remains biased. In Design III, there is missing-at-random, and as a result the IPW corrections exhibit no bias.

In this simulation design, the IPW correction for the ATE-R tends to have a smaller standard deviation relative to the CiC correction for that same object. As a result, in Designs II and III, where a correction for the ATE-R is not warranted, the IPW correction for the ATE-R has a slightly smaller RMSE than the CiC. However, when a correction is warranted as in Design I (II), the CiC correction for the ATE-R and the ATE (ATE-R) has a substantially smaller RMSE relative to IPW, given the latter's sizeable bias.

4 Empirical Illustration

We apply both our proposed CiC correction and the IPW correction to an outcome from a large-scale randomized evaluation of the impact of *Progresa*, a conditional cash transfer program in Mexico. Although this example relies on a randomized controlled trial, as indicated in Section 2.2, our results do not require random assignment. The *Progresa* evaluation, which was implemented in 1997, randomized 506 villages into a treatment group and a control group. These villages were designed to be representative of a larger group of 6396 eligible villages in Mexico. Thus, both the average treatment effect for the respondent subpopulation (ATE-R) and for the study population (ATE) are likely to be of interest in this setting. In the 320 treatment villages, families received a cash transfer conditional on engaging in specific education and health-seeking behaviors if they were below the given threshold on a poverty index. In the control villages, no households were offered a transfer. There is a vast literature that has studied a range of outcomes from the *Progresa* evaluation, with some of the most studied outcomes focusing on education and health (Skoufias, 2001; Schultz, 2004; Attanasio et al., 2012).

The outcome we examine in this application is the value of a productive asset, specifically farm animals.[9] It is a continuous variable as required for point-identification in the CiC approach by Assumption 1 and 2. Although as discussed in Remark 2, it is straightforward to relax these assumptions and obtain bounds. In addition, the baseline outcome data is available, which allows us to apply the CiC correction. The *Progresa* evaluation conducted three follow-up household surveys over 13 months (5, 12, and 18 months after the program began). For this example, we focus on data from the second

[9]Since *Progresa* is a cash transfer targeted largely to rural families who may be in agriculture, they may make productive investments in capital such as production animals. This outcome is first proposed in Gertler et al. (2012). Our findings are not directly relevant to that paper since we focus on the second and third follow-ups individually, while Gertler et al. (2012) focus on the outcome pooled across all three follow-ups. We also restrict our sample to those who appear in the baseline survey.

and third follow-up surveys, since they have significant attrition rates of 11.3% and 12.2% respectively for the outcome of interest.[10] There is modest differential attrition of 2% to 3% in both follow-ups with no clear pattern across treatment and control groups (see Table 3). For the purposes of this analysis, the attrition rate is conditional on appearing in the baseline survey, which included a total of 12,299 households. Thus, the outcome is observed for more that 10,000 households in both follow-up surveys.

We examine the CiC-corrected estimates for the ATE-R and ATE in relation to the the naïve (uncorrected) estimate of the treatment effect, $\widehat{\Delta}_R$.[11] As in our simulations, we also compare the CiC and IPW corrections, since both provide point estimates for the ATE-R and the ATE. IPW is the most widely-used correction in the literature that is directly comparable to the CiC correction.[12] The estimated treatment effect on the value of production animals for the third follow-up is 351.0 pesos, which is significant at the 1% level and is relative to a control mean of 1,096 pesos. The CiC-corrected ATE-R is 283.9 pesos, which is not significantly different from the naïve treatment effect estimate. In contrast, CiC-corrected ATE is 288.9 pesos, which is significantly different from the naïve estimate at the 5% level (Panel B). Specifically, the difference between the CiC-corrected estimate and the naive estimate for the ATE-R is 67.1 with a standard error of 124.7, while for the ATE, the difference is 62.0 with a standard error of 26.0. In contrast, neither of the IPW-corrected estimates, which are 348.6 for the ATE-R and 338.5 for the ATE, are close to be being significantly different from the naïve treatment effect estimate even at the 10% level. Furthermore, for the ATE, the CiC-corrected estimate is significantly different from the IPW-corrected estimate at the 5% level.

Thus, the CiC approach indicates that a correction is required for the ATE while the IPW approach does not. This raises the question as to which of these results is more plausible in this setting. In order to shed some light on this question, we consider whether attrition is likely to be causing a violation of the internal validity that would necessitate a correction. Thus, we implement the attrition tests proposed in Ghanem et al. (2021). We find that the test of internal validity for the respondents (IV-R) is not rejected, whereas the test of interval validity for the study population (IV-P) is rejected. As a result, this empirical application is consistent with the testable prediction of Design II in our simulations, in which internal validity is violated for the ATE, but not the ATE-R. In the simulation results for Design II, the CiC- and IPW-corrected estimates are both consistent for the ATE-R, but only the CiC-corrected estimate is consistent for the ATE.[13]

More broadly, a consideration in interpreting these results is whether the assumptions of the CiC or the IPW corrections are more plausible in this setting. CiC relies on the assumption of time invariance of the *distribution* of unobservables that affect the outcome conditional on treatment and response status. In contrast, IPW relies on the assumption that conditional on baseline outcome, unobservables that affect the response equation are independent of the unobservables that affect the outcome.

In order to illustrate the implications of these assumptions in this setting, we consider two types of

[10] The attrition rate for the first follow-up is 5.6%.

[11] This is simply the difference in the mean outcome between treatment and control respondents.

[12] Another widely used attrition correction are the bounds proposed by Lee (2009). Since this approach only focuses on the average treatment effect on the always-responders, however, it is not directly comparable to our method that focuses on the ATE-R and the ATE.

[13] It may be the case that the test of IV-R fails to reject because it is underpowered.

unobservables that could affect both response and the outcome. First, we consider a case in which the unobservable of concern is constant over time at the individual level ($U_{i0} = U_{i1}$). Since we often assume preferences are relatively stable over time, a relevant example in this case might be a preference for saving. A household's preference for saving would affect the outcome since it is a form of investment, and it could also affect response if people with a higher preference for saving spend more time at home. In this case, both the CiC and IPW assumptions would hold. Since the unobservable is the same over time at the individual level, capturing it at baseline through the outcome can explain response in the follow-up period, which is required for IPW. In addition, the distribution of the unobservable is trivially time-invariant since it is constant at the individual level.

Next, we consider a case in which the unobservables are not the same over time but the distribution is time-invariant ($U_{i0} \neq U_{i1}$, but $U_{i0}|G_i, R_i \stackrel{d}{=} U_{i1}|G_i, R_i$). A range of typical shocks, such as those driven by health, could affect both investment and the likelihood that a household responds to the follow-up survey. The CiC assumption would hold if the outcome is a monotonic transformation of the time-varying unobservable, U_{it}, in both time periods. The IPW assumption does not hold, however, in any case in which response depends on the outcome in the follow-up period. Thus, in our example of a health shock, the IPW assumption requires that response in the follow-up period only depends on the health shock that occurred in the baseline period, but not the health shock in the follow-up period. The CiC assumption would only be violated, however, if the distribution of the outcome changes from the baseline to the follow-up period in such a way that the outcome in the follow-up period is not a monotonic transformation of the baseline outcome.

In order to further confirm that the CiC corrections perform as expected, we also apply the CiC and IPW corrections to the data from the second follow-up survey. In this case, the treatment effect for the second follow-up is 183.8, and this is not significant. The CiC-corrected estimates for the ATE-R and ATE are 130.4 and 142.5 respectively, while the IPW-corrected estimates are 184.1 and 184.6. In contrast to our findings regarding the third follow-up, in this case, neither of the two CiC-corrected estimates are statistically different from either the estimated treatment effect or the IPW-corrected estimates. These findings are again consistent with the results of the application of tests for attrition bias. For this follow-up, we do not reject internal validity for either the respondent subpopulation or the study population, given p-values of 0.514 and 0.658, respectively. Thus, we do not find evidence that an attrition correction is required. That means that the underlying process in this follow-up is consistent with Design III from the simulation section in which neither IV-P nor IV-R are violated. Of course, alternatively, there may simply be insufficient power in this setting to find such violations.

Table 3: Value of Production Animal (Mexican pesos)

Panel A. Observed Difference in Means and Attrition Corrections

Follow-up	N	Estimator	Control Mean	$\widehat{\Delta}_R$	CiC				IPW	
					ATT-R	ATU-R	ATE-R	ATE	ATE-R	ATE
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2	10,907	Estimate	1,149	183.8	131.0	129.4	130.4	142.5	184.1	184.6
		S.E.	95.7	121.0	94.3	92.0	93.1	117.5	121.5	120.9
3	10,799	Estimate	1,096	351.0***	285.6***	281.1***	283.9***	288.9***	348.6***	338.5***
		S.E.	86.0	118.9	92.4	101.2	94.6	107.1	113.3	110.2

Panel B. Differences between Estimates

Follow-up	Estimator	(2)-(5)	(2)-(6)	(5)-(6)	(5)-(7)	(6)-(8)	(2)-(7)	(2)-(8)
2	Difference	53.4	41.3	-12.1	-53.7	-42.1	-0.4	-0.8
	S.E.	125.3	37.5	120.7	126.3	38.1	7.0	7.4
3	Difference	67.1	62.0**	-5.1	-64.7	-49.5**	2.5	12.6
	S.E.	124.7	26.0	113.8	116.6	21.7	9.9	12.9

Panel C. Attrition Rates, Baseline Outcome, and Attrition Tests

Follow-up	Attrition Rates			Mean Baseline outcome by Group				Attrition Tests	
	Overall	T	C	TR	CR	TA	CA	IV-R	IV-P
2	11.3	10.6	12.5	1,850.7	1,786.9	1,609.9	1,961.7	0.514	0.658
3	12.2	13.2	10.6	1,905.7	1,850.5	1,294.8	1,455.6	0.713	0.000

Notes: The sample size at baseline consists of 12,299 households. Standard errors refer to bootstrap standard errors. P-values are reported for attrition tests of internal validity of the respondent sub-population (IV-R) and the study population (IV-P). *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

5 Conclusion

In this paper, we propose an attrition correction method for the average treatment effects on the respondents as well as the study population that relies on the changes-in-changes framework. We achieve identification through two main assumptions: continuity and strict monotonicity of each potential outcome in a scalar unobservable, and time invariance of the distribution of the unobservable conditional on treatment and response statuses. We illustrate the methodology through both simulations and an empirical example. In particular, we demonstrate that there are cases in which the CiC corrections are consistent with attrition tests, and the IPW corrections are not.

In practice, including covariates can render the identifying assumptions required for the proposed corrections more plausible. The extension of our identification results to include covariates is straightforward. There are several estimation approaches available in the literature. Athey and Imbens (2006) provide a nonparametric estimation approach as well as a parametric one relying on separability assumptions. More recently, Melly and Santangelo (2015) propose a semiparametric approach which avoids the curse of dimensionality of the former while relaxing the separability restrictions of the latter.

A Mathematical Appendix

Proof (Lemma 1)
1. For $(g,r) \in \{0,1\}^2$, Assumptions 1 and 2.1 imply the following for y in the support of $Y_1(0)|G = g, R = r$

$$\begin{aligned} F_{Y_1(0)|G=g,R=r}(y) &= P(\mu_1(0,U_1) \leqslant y|G=g,R=r) = P(U_1 \leqslant \mu_1^{-1}(0;y)|G=g,R=r) \\ &= P(U_0 \leqslant \mu_1^{-1}(0;y)|G=g,R=r) = P(\mu_0(0,U_0) \leqslant \mu_0(0,\mu_1^{-1}(0;y))|G=g,R=r) \\ &= F_{Y_0(0)|G=g,R=r}(\mu_0(0,\mu_1^{-1}(0;y))) \equiv F_{Y_0(0)|G=g,R=r}(T_0(y)) \quad for \quad y \in \mathbb{Y}_{g,r}^{0,1} \end{aligned} \tag{10}$$

where the second equality holds by Assumption 2.1, the third equality follows from Assumption 1.1 and the fourth equality follows by applying the transformation $\mu_0(0,.)$ which is valid by Assumption 2.1. The result in (i) follows by the definition of $T_0(y) = \mu_0(0,\mu_1^{-1}(0;y))$.

The identification of $T_0(y)$ on $y \in \mathbb{Y}_{0,1}^{0,1}$ follows from (10), the fact we observe $F_{Y_t(0)|G=0,R=1}$ for $t = 0, 1$, and the strict monotonicity of $F_{Y_0(0)|G=0,R=1}(\cdot)$ by Assumptions 1.2 and 2.1, which imply

$$T_0(y) = F_{Y_0|G=0,R=1}^{-1}(F_{Y_1|G=0,R=1}(y)) \quad for \quad y \in \mathbb{Y}_{0,1}^{0,1}. \tag{11}$$

This completes the proof of (i) and (ii).
2. By similar arguments, Assumptions 1, 2.1 and 2.2 imply the following for $(g,r) \in \{0,1\}^2$,

$$\begin{aligned} F_{Y_1(1)|G=g,R=r}(y) &= P(\mu_1(1,U_1) \leqslant y|G=g,R=r) = P(U_1 \leqslant \mu_1^{-1}(1;y)|G=g,R=r) \\ &= P(U_0 \leqslant \mu_1^{-1}(1;y)|G=g,R=r) = P(\mu_0(0,U_0) \leqslant \mu_0(0,\mu_1^{-1}(1;y))|G=g,R=r) \\ &= P(Y_0(0) \leqslant \mu_0(0,\mu_1^{-1}(1;y))|G=g,R=r) \\ &\equiv F_{Y_0(0)|G=g,R=r}(T_1(y)) \quad for \quad y \in \mathbb{Y}_{g,r}^{1,1} \end{aligned} \tag{12}$$

The identification of $T_1(y)$ on $y \in \mathbb{Y}_{1,1}^{1,1}$ follows from (12), the fact we observe $F_{Y_0(0)|G=1,R=1}$ and $F_{Y_1(1)|G=1,R=1}$, and the strict monotonicity of $F_{Y_0(0)|G=1,R=1}(\cdot)$ by Assumptions 1.2 and 2.1, which imply

$$T_1(y) = F_{Y_0|G=1,R=1}^{-1}(F_{Y_1|G=1,R=1}(y)) \quad for \quad y \in \mathbb{Y}_{1,1}^{1,1}. \tag{13}$$

This completes the proof of (i) and (ii). □

Proof (Proposition 1)
1. Since all conditions required for Lemma 1.1 are imposed, $T_0(y)$ is identified for $y \in \mathbb{Y}_{0,1}^{0,1}$ (Lemma 1.1(ii)). The imposed support condition, $\mathbb{U}_{1,1} \subseteq \mathbb{U}_{0,1}$, together with the strict monotonicity of the untreated potential outcome (Assumption 2.1) implies that $\mathbb{Y}_{1,1}^{0,1} \subseteq \mathbb{Y}_{0,1}^{0,1}$. As a result, $T_0(y)$ is identified for $y \in \mathbb{Y}_{1,1}^{0,1}$. By Lemma 1.1(i)-(ii), the equality in (2) follows,

$$F_{Y_1(0)|G=1,R=1}(y) = F_{Y_0|G=1,R=1}(F_{Y_0|G=0,R=1}^{-1}(F_{Y_1|G=0,R=1}(y))) \quad for \quad y \in \mathbb{Y}_{1,1}^{0,1}. \tag{14}$$

The identification of the ATT-R in (3) is immediate from (2).

2. Since all conditions required for Lemma 1.2 hold, $T_1(y)$ is identified for $y \in \mathbb{Y}_{1,1}^{1,1}$. The support condition imposed here, $\mathbb{U}_{0,1} \subseteq \mathbb{U}_{1,1}$, together with the strict monotonicity of the treated potential outcome (Assumption 2.2) implies that $\mathbb{Y}_{0,1}^{1,1} \subseteq \mathbb{Y}_{1,1}^{1,1}$. As a result, $T_1(y)$ is identified for $y \in \mathbb{Y}_{0,1}^{1,1}$. By Lemma 1.2(i)-(ii), the equality in (4) follows,

$$F_{Y_1(1)|G=0,R=1}(y) = F_{Y_0|G=0,R=1}(F^{-1}_{Y_0|G=1,R=1}(F_{Y_1|G=1,R=1}(y)) \quad for \quad y \in \mathbb{Y}_{0,1}^{1,1}. \tag{15}$$

The identification of the ATU-R in (5) is immediate from (4).

□

Proof (Proposition 2) Note that by random assignment

$$\begin{aligned}
&E[Y_1(1) - Y_1(0)] = E[Y_1(1)|G=1] - E[Y_1(0)|G=0], \\
=&P(R=1|G=1)E[Y_1(1)|G=1,R=1] + P(R=0|G=1)E[Y_1(1)|G=1,R=0] \\
&- P(R=1|G=0)E[Y_1(0)|G=0,R=1] - P(R=0|G=0)E[Y_1(0)|G=0,R=0].
\end{aligned}$$

All of the quantities on the RHS of the last equality are observed except $E[Y_1(1)|G=1,R=0]$ and $E[Y_1(0)|G=0,R=0]$. Under the maintained assumptions, the conditions for Lemma 1 hold. Therefore, we can identify the distribution of $Y_1(1)$ for the treatment attritors as well as the distribution of $Y_1(0)$ for the control attritors. Specifically, we have

$$\begin{aligned}
&F_{Y_1(1)|G=1,R=0}(y) = F_{Y_0|G=1,R=0}(F^{-1}_{Y_0|G=1,R=1}(F_{Y_1|G=1,R=1}(y))), \\
&E[Y_1(1)|G=1,R=0] = E[F^{-1}_{Y_1|G=1,R=1}(F_{Y_0|G=1,R=1}(Y_0))|G=1,R=0],
\end{aligned}$$

and

$$\begin{aligned}
&F_{Y_1(0)|G=0,R=0}(y) = F_{Y_0|G=0,R=0}(F^{-1}_{Y_0|G=0,R=1}(F_{Y_1|G=0,R=1}(y))), \\
&E[Y_1(0)|G=0,R=0] = E[F^{-1}_{Y_1|G=0,R=1}(F_{Y_0|G=0,R=1}(Y_0))|G=0,R=0].
\end{aligned}$$

As a result, the ATE is identified. □

Proof (Proposition 3) We have for all $d = 0, 1$,

$$\begin{aligned}
E[Y_1(d)] \quad = \quad & P(G=1,R=1)E[Y_1(d)|G=1,R=1] + P(G=0,R=1)E[Y_1(d)|G=0,R=1] \\
& + P(G=1,R=0)E[Y_1(d)|G=1,R=0] + P(G=0,R=0)E[Y_1(d)|G=0,R=0].
\end{aligned}$$

Then

$$\begin{aligned}
E[Y_1(1)] - E[Y_1(0)] \quad = \quad & P(G=1,R=1)E[Y_1(1) - Y_1(0)|G=1,R=1] + P(G=0,R=1)E[Y_1(1) - Y_1(0)|G=0,R=1] \\
& + P(G=1,R=0)E[Y_1(1) - Y_1(0)|G=1,R=0] + P(G=0,R=0)E[Y_1(1) - Y_1(0)|G=0,R=0].
\end{aligned}$$

Under the maintained assumptions, the conditions for Lemma 1 hold. Therefore, the distributions $F_{Y_1(1)|G=1,R=0}$, $F_{Y_1(1)|G=0,R=1}$, $F_{Y_1(1)|G=0,R=0}$, $F_{Y_1(0)|G=1,R=0}$, $F_{Y_1(0)|G=1,R=1}$, and $F_{Y_1(0)|G=0,R=0}$ are

identified. The distributions, $F_{Y_1(1)|G=1,R=1}$ and $F_{Y_1(0)|G=0,R=1}$, as well as the probability weights, $P(G=g,R=r)$ for all $g=0,1$ and $r=0,1$, are identified from the data. Therefore, all objects in the definition of the ATE given in the proposition are identified. It follows that the ATE is identified. □

Proof (Proposition 4) Let $F_{U_0,U_1|G}$ denote the joint distribution of (U_0,U_1) conditional on G, and let $F_{U_t|G}$ denote the marginal distribution of U_t conditional on G. By exchangeability, it follows that

$$F_{U_0|G}(v) = F_{U_1|G}(v) \Longrightarrow dF_{U_0|G}(v) = dF_{U_1|G}(v),$$

Furthermore, since $F_{U_0|G=g}(u_0) = F_{U_1|G=g}(u_0)$ for $u_0 \in \mathbb{U}_g$, the support of $U_0|G=g$, we obtain

$$F_{U_1|U_0=u_0,G=g}(u_1) = \frac{F_{U_0,U_1|G=g}(u_0,u_1)}{F_{U_0|G=g}(u_0)} = \frac{F_{U_0,U_1|G=g}(u_1,u_0)}{F_{U_1|G=g}(u_0)} = F_{U_0|U_1=u_0,G=g}(u_1)$$

for $u_0 \in \mathbb{U}_g$ and $g=0,1$, where the second equality follows the exchangeability condition.

On the other hand, we have

$$\begin{aligned} P(\varphi(U_0,U_1,G)=r|U_0=v,G) &= P(\varphi(U_1,U_0,G)=r|U_0=v,G), \\ &= P(\varphi(U_1,v,G)=r|U_0=v,G), \\ &= P(\varphi(U_0,v,G)=r|U_1=v,G), \\ &= P(\varphi(U_0,U_1,G)=r|U_1=v,G), \end{aligned}$$

where the first equality holds under symmetry of $\varphi(.,.,G)$, the third holds because $F_{U_1|U_0,G} = F_{U_0|U_1,G}$.

Therefore, from Equation (9), we conclude that $P(U_t \leqslant u, R=r|G)$ is time homogeneous. Since R and G are time-invariant random variables, it follows that $P(U_t \leqslant u|G,R)$ is also time homogeneous. □

References

Athey, Susan and Guido W. Imbens, "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 2006, *74* (2), 431–497.

Attanasio, Orazzio P., Costas Meghir, and Ana Santiago, "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA," *The Review of Economic Studies*, 2012, *79* (1), 37–66.

Behagel, Luc, Bruno Crépon, Marc Gurgand, and Thomas Le Barbanchon, "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models," *Review of Economics and Statistics*, 2015, *97*, 1070–1080.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt, "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *The Journal of Human Resources*, 1998, *33* (2), 251–299.

Gertler, Paul J., Sebastian W. Martinez, and Marta Rubio-Codina, "Investing Cash Transfers to Raise Long-Term Living Standards," *American Economic Journal: Applied Economics*, January 2012, *4* (1), 164–92.

Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Becerra, "Testing Attrition Bias in Field Experiments," 2021. CEGA Working Paper No. 113.

Heckman, James J, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," in Sanford V. Berg, ed., *Annals of Economic and Social Measurement*, Vol. 5, National Bureau of Economic Research, 1976, pp. 475–492.

Heckman, James J., "Sample Selection Bias as A Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

Horowitz, Joel L. and Charles F. Manski, "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 2000, *95* (449), 77–84.

Huber, Martin, "Identification of Average Treatment Effects in Social Experiments Under Alternative Forms of Attrition," *Journal of Educational and Behavioral Statistics*, 2012, *37* (3), 443–474.

Lee, David S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Melly, Blaise and Giulia Santangelo, "The changes-in-changes model with covariates," 2015. Unpublished Manuscript.

Millán, Teresa Molina and Karen Macours, "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias," 2019. Unpublished Manuscript.

Schultz, Paul, "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program," *Journal of Development Economics*, 2004, *74* (1), 199–250. New Research on Education in Developing Economies.

Skoufias, Emmanuel, "PROGRESA and Its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI," Working Paper, International Food Policy Research Institute April 2001.

van den Berg, Gerard J. and Maarten Lindeboom, "Attrition in Panel Survey Data and the Estimation of Multi-State Labor Market Models," *The Journal of Human Resources*, 1998, *33* (2), 458–478.

Ziliak, James P. and Thomas J. Kniesner, "The Importance of Sample Attrition in Life Cycle Labor Supply Estimation," *The Journal of Human Resources*, 1998, *33* (2), 507–530.