Attrition in Randomized Controlled Trials: Using Tracking Information to Correct Bias

TERESA MOLINA-MILLÁN
University of Alicante

KAREN MACOURS
Paris School of Economics and Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE)

I. Introduction

Longitudinal surveys and panel datasets are indispensable tools for the study of economic and demographic dynamics in developing countries. Randomized controlled trials (RCTs) in particular rely on panel datasets to estimate impacts of randomized interventions on the target population. Keeping a panel dataset representative of the study population requires addressing attrition. Potential biases resulting from missing individuals across waves in panel surveys have long been a focus of study for longitudinal data in developed countries.[1] The literature on developing countries is more limited. The many data collection efforts in the context of RCT studies in developing countries, however, point to the

This research would not have been possible without the support of Ferdinando Regalia of the Inter-American Development Bank (IDB) and fellow principal investigators Tania Barham and John Maluccio in the wider research project. We gratefully acknowledge generous financial support from IDB, the Initiative for International Impact Evaluation (3ie; grant no. OW2.216), the National Science Foundation Division of Social and Economic Sciences (SES; grant nos. 11239945 and 1123993), and the French National Research Agency (ANR; grant no. ANR-17-EURE-0001). Teresa Molina-Millán acknowledges financial support from the Spanish Ministry of Science and Innovation (grant no. PID2021-124237NB-I00) and from the Generalitat Valenciana through the plan GenT program (grant no. CDEIGENT/2020/016). We are indebted to Veronica Aguilera, Enoe Moncada, and the survey team from Centro de Investigación de Estudios Rurales y Urbanos de Nicaragua (CIERUNIC) for excellent data collection and for their dogged persistence in tracking. We are grateful for many related discussions and ideas from Tania Barham, Luc Behaghel, John Maluccio, and Joachim De Weerdt as well as comments received during presentations at the Paris School of Economics, the IDB, and the Latin American and Caribbean Economic Association (LACEA), and from two anonymous referees. Data are provided through Dataverse at https://doi.org/10.7910/DVN/2RBAB4. All remaining errors and omissions are our own. Contact the corresponding author, Karen Macours, at karen.macours@psemail.eu.

[1] See, e.g., the special issue of the *Journal of Human Resources* on "Attrition in Longitudinal Surveys" (spring 1998, vol. 33, no. 2).

Electronically published November 12, 2024

Economic Development and Cultural Change, volume 73, number 2, January 2025.

https://doi.org/10.1086/730612

relevance of better understanding attrition in such contexts. This is even more so as more analyses start focusing on the longer-term effects and dynamics after RCTs.

Attrition in developing countries is often driven by migration, and the decision to migrate can be affected by the randomized exposure to particular interventions, contextual and individual characteristics, and the interactions between them. The challenges posed by attrition, therefore, are different from those in developed countries where it is often related to refusals. While a number of longitudinal studies in developing countries document that those who are missing differ in observable characteristics from those who are found (Alderman et al. 2001; Thomas, Frankenberg, and Smith 2001; Falaris 2003; Baird, Hamory, and Miguel 2008; Thomas et al. 2012), the implications for intention-to-treat (ITT) estimations in impact evaluations are not always fully taken into account. Apart from the reduction in the number of observations, and the related loss of statistical power, attrition can reduce internal validity in case it leads to unbalanced samples. It can also have important implications for external validity in the presence of heterogeneous treatment effects, if treatment effects are different for attritors than for the rest of the study population. In addition, take-up of an intervention is likely to be lower for people that migrate prior or during the intervention. Omitting migrants can then lead to an overestimate of the ITT effects. Data collection to estimate impacts of RCTs, however, often does not include protocols to track migrants outside their village or community of origin, sometimes resulting in high attrition rates. Given the mobility of many target populations migrant tracking can be costly, and data collection cost concerns often need to be weighted against the potential consequences of selection bias due to attrition.

This paper exploits different phases of the tracking protocol of four longitudinal impact evaluation surveys to illustrate the potential challenges resulting from nonrandom attrition in RCTs. After showing that commonly used tracking protocols would have led to meaningful over- or underestimation of the treatment effect, even when attrition appeared balanced, we discuss a modification of a common attrition correction estimator that uses information from different stages of the tracking process to account for the remaining attrition.

To motivate the analysis, we reviewed how RCT studies in development economics published in top journals handle attrition.[2] Both survey attrition rates and approaches to address attrition bias concerns vary widely. Average annual attrition rates in studies targeting respondents below 18 years old, for instance,

[2] See app. D, sec. D.2 (apps. A–D are available online) for details on the selection of papers and the different findings.

vary from 0 to 60%. The consideration of the potential attrition bias is often limited, in contrast with the care given ex ante to assure random program placement. Around 23% of studies do not go beyond testing whether attrition rates between treatment arms are different, and 18% of studies do not even show such testing. For studies that address attrition in more detail, a variety of approaches can be found, including studying the correlates of attrition (44%), showing balance after attrition (28%), or analyzing treatment heterogeneous effects on attrition (45%). Overall, only 28% of studies explicitly correct for attrition in the estimations, with nonparametric bounds and inverse probability weights (IPW) the most common methodologies applied. Among studies in which the authors identified nonrandom attrition, around 31% do not apply a sample-selection correction method to correct for attrition. On the other hand, around one-third percent of studies showing balanced attrition rates still apply a sample-selection correction method.[3]

To illustrate the implications of different approaches to attrition for specific cases, this paper analyzes the incidence and implications of attrition on four longitudinal data sets collected for randomized evaluations of cash transfer programs.

We first test whether response rates are balanced between treatment arms at different stages during the tracking process. We then analyze the implications of attrition by estimating ITT effects for different subsamples corresponding to the different phases of the tracking process. We distinguish between a regular tracking phase (RTP), and an intensive tracking phase (ITP). The division between regular and intensive tracking corresponds to normal and high-effort tracking processes, where the regular tracking process is similar to the common protocol in many surveys. Comparing results after the RTP and the ITP allows for discussion of the different identification assumptions. One consequence of attrition can be that the treatment and control samples that were initially balanced are no longer balanced due to the attrition. In this case, it is the internal validity that is compromised. Intensive tracking can reduce the probability and magnitude of such potential imbalances or allow researchers to account for them even when they remain. Another consequence of attrition is that the parameter that is obtained ignoring attrition can no longer be interpreted as the average treatment effect on the initial sample but simply on the sample of respondents. This is true even if attrition is balanced. In this case, the problem is external

[3] While addressing attrition through methods beyond balance tests has become more common in recent literature, qualitatively, the summary of practices in the literature is not very different when considering only more recently published work. See tables D4 and D6 (tables A1–D8 are in the online appendixes).

validity, and it is strongly linked to the heterogeneity of treatment effects. Intensive tracking then helps to assure that the final sample is closer to the initial population of interest and, therefore, to obtain a more policy-relevant estimator. This can be particularly relevant as treatment heterogeneity can be directly linked to migration, so that not including migrants in the final sample can lead to both an underestimate and an overestimate of the average treatment effect.

We estimate the ITT coefficient of the interventions on a primary outcome of each reference paper and show that without conducting an intensive tracking phase, both types of problems arise. Attrition is correlated with many baseline observables, capturing socioeconomic status (SES), demographic composition of the household, family networks, and the potential temporary nature of the baseline residence. Moreover, these characteristics relate differently to attrition in the treatment versus control. A comparison of baseline characteristics by the respondent's status at the end of the follow-up survey (found during RTP/found during ITP/never found) shows that those who were never found are often more similar in baseline characteristics to those in the ITP sample than to those in the RTP sample. We propose a new method to correct for attrition bias exploiting these similarities in observable characteristics between attritors and the intensive tracking sample. We build on Fitzgerald, Gottschalk, and Moffitt (1998) and Wooldridge (2002b) and estimate the probability to be found to construct inverse probability weights. But instead of using baseline information for the complete sample of respondents in the follow-up, we estimate weights using only information on the sample of respondents tracked during the ITP. The underlying assumption is that those found and not found in the ITP sample are more similar, in both observed and unobserved characteristics, than those in the complete tracking phase (CTP) sample. We further show that the observed characteristics have more explanatory power in the ITP sample than in the CTP sample. As such, they allow us to estimate an alternative set of weights based on a different set of assumptions and serve as a useful robustness test to standard IPW using the full sample. This can be particularly useful when bounds are either too large to be informative or the monotonicity assumption underlying Lee bounds is in doubt.

This paper hence illustrates three main points: the usefulness of considering intensive tracking observations, the usefulness of weighting the observations to take into account the characteristics of the attriters, and the potential relevance of the shape of the weight to be informed by attrition in the intensive tracking phase, under plausible assumptions. As such, it relates to the econometric literature on sample selection.[4] It shares with Behaghel et al.'s (2015) selectivity

[4] For a review of this literature, see app. D, sec. D.1.

correction procedure the use of information on those who were difficult to find. We differ from Lee (2002) and Behaghel et al. (2015) by developing a method that allows for nonmonotonic differential attrition, using information from an intensive follow-up. The latter relates to DiNardo, McCrary, and Sanbonmatsu's (2006) and Hull's (2015) use of intensive tracking survey design features. Our approach extends the sample selection correction model based on observable characteristics by identifying, through the tracking protocol, a subsample of respondents similar to the subsample of attritors. That is, we assume that the ITT causal effect is homogeneous among individuals not found after regular tracking, after controlling for a large number of observables through the weights.

This paper further builds on work studying attrition bias in household surveys in developing countries in non-RCT contexts. A number of studies use longitudinal datasets with low attrition rate to analyze differences between movers and stayers and to infer potential attrition bias on the analysis of the outcomes of interest (Thomas, Frankenberg, and Smith 2001; Velasquez et al. 2010; Beegle, DeWeerdt, and Dercon 2011; Thomas et al. 2012).[5] Overall, these studies agree on the fact that attritors differ from those who are found in observable characteristics. Alderman et al. (2001), Falaris (2003), and Fuwa (2011) show that estimates are not necessarily biased even if attritors are different from stayers, but attrition bias can depend on the outcome of interest (Maluccio 2004).[6]

To our knowledge, there is only one other paper specifically studying tracking protocols in the context of an RCT study in a developing country (Baird, Hamory, and Miguel 2008). Analyzing tracking in the Kenya Life Panel Survey, they compare ITT estimates of migrants that were tracked versus populations surveyed in their original locations and find evidence of heterogeneous treatment effects that are correlated with migration. Our paper starts from a similar finding but then uses the information obtained from the intensive tracking phase to correct for attrition bias.

Section II introduces the evaluation designs and the datasets used in the empirical applications. It also illustrates the sensitivity of the ITT estimates with and without inclusion of difficult-to-find respondents. Section III discusses the

[5] For instance, Beegle, De Weerdt, and Dercon (2011) show, with a household fixed effect model, that migrants moving out of their community of origin experienced 36 percentage points more of consumption growth than nonmigrant household members between 1991 and 2004. They would have underestimated the growth in consumption by half of its true increase if they had focused only on individuals residing in their community of origin.

[6] Alderman et al. (2001) did not find any impact of attrition bias on anthropometric indicators in the Kwazulu-Natal Income Dynamics Study (KIDS), but Maluccio (2004) found evidence of attrition bias on expenditures using the same database.

correlates of attrition and compliance to understand the potential biases before introducing the new inverse probability weighting estimator. Section IV concludes and discusses guidelines to evaluate cost trade-offs around intensive tracking.

II. RCT Panel Datasets: Program Design, Evaluation, and Data

We draw on four RCT panel datasets with relatively high attrition after regular tracking, which was followed by an intensive tracking phase. Response rates by tracking phase and treatment group are shown in table 1.[7]

Barham, Macours, and Maluccio (2017) estimate the long-term impact of a conditional cash transfer program, Red de Protección Social (RPS), on long-term labor market outcomes of former beneficiaries in Nicaragua. The 3-year program started in 21 randomly selected treatment localities in the mid-2000 after a baseline census. In 2003, the experimental treatment localities stopped receiving the transfers, and the program started in the 21 experimental control localities and received transfers during the following 3 years. A follow-up survey in 2010 allows us to estimate differential treatment effects for boys aged 9–12 years old at baseline on participation in off-farm employment, which can be seen as the targeted final long-term outcome for the controlled clinical trial (CCT).[8]

During the first phase of data collection, all individuals were tracked in their localities of origin, and some migrants were followed to other localities close by. We refer to this phase as the regular tracking phase, as it is similar to the most used tracking protocol in longitudinal surveys, even if it already includes information on some migrants. A subsequent phase intensively tracked all missing individuals to other regions or to Costa Rica. Survey teams also went back to the localities of origin for regular updates on the destination information and to survey returned temporal migrants (for more detail on the tracking protocol, see

[7] Among the RCTs included in the literature review (see app. D, sec. D.2), five datasets satisfy the basic requirements to apply our methodology, i.e., (i) the endline included an intensive tracking phase and information on the subsample that was tracked intensively and (ii) the baseline data include a vector of variables with enough information to estimate a model of attrition. We maintained those for which we had access to the relevant data.

[8] This cohort had greater program exposure in the early-treatment localities than in the late-treatment localities due to the eligibility criteria for the education transfer and preprogram school dropout patterns. It includes children that were young enough to be eligible for the education transfer if they were living in an early-treatment locality in 2000 but too old to receive the education transfer when the program was introduced in the late-treatment localities in 2003. The sample for the long-term follow-up purposely oversampled children of ages for which the program exposure difference was large, drawing from the full baseline census of eligible households in treatment and control localities. This is accounted for through probability weights.

Off-farm employment is measured as a dichotomous variable that takes the value one if the individual is economically active (in wage or self-employment) outside of the family farm, and zero otherwise.

TABLE 1
RESPONSE RATES BY TRACKING PHASE AND TREATMENT GROUP

	Mean		
	Treatment (1)	Control (2)	Difference in Means (SE) (3)
	A. Red de Protección Social (Barham, Macours, and Maluccio 2017)		
CTP Sample	.882	.897	−.012 (.025)
RTP Sample	.706	.755	−.044 (.038)
ITP Sample	.597	.581	.054 (.073)
	B. Atención a Crisis (Macours, Schady, and Vakis 2012)		
CTP Sample	.952	.951	−.001 (.009)
RTP Sample	.709	.660	.044* (.026)
ITP Sample	.835	.857	−.044 (.031)
	C. Youth Opportunity Program: 4 Years after Baseline (Blattman, Fiala, and Martinez 2014)		
CTP sample	.854	.791	.071*** (.025)
RTP sample	.626	.605	.024 (.024)
ITP sample	.583	.581	.030 (.050)
	D. Youth Opportunity Program: 9 Years after Baseline (Blattman, Fiala, and Martinez 2020)		
CTP sample	.867	.891	−.029 (.024)
RTP sample	.747	.675	.064** (.023)
ITP sample	.322	.521	−.198** (.079)

Note. Survey response rates for the CTP sample in Blattman, Fiala, and Martinez (2014, 2020) correspond to effective response rates. Column 1 reports the mean of the treatment group. Column 2 reports the mean of the control group. Column 3 reports estimates for differences controlling for stratification fixed effects. Standard errors are in parentheses, clustered at locality level in Macours, Schady, and Vakis (2012) and Barham, Macours, and Maluccio (2017), and at the group level in Blattman, Fiala, and Martinez (2014, 2020).
* $p < .10$.
** $p < .05$.
*** $p < .01$.

app. C). Response rates are not significantly different between the early and late treatment groups at the different stages. After RTP, attrition was about 26%, but differences between treatment arms are not significant. Around 60% of those not surveyed after RTP were found during ITP, with the differences between treatment arms again not significant, leading to final balanced attrition of 11%.

Macours, Schady, and Vakis (2012) estimate the impact of a different 1-year cash transfer program, Atención a Crisis, on early-childhood development in Nicaragua, with 56 localities randomly allocated to treatment and 50 localities to control. Baseline data was collected in April–May 2005 and endline data between August 2008 and May 2009. The primary outcome is the overall index of cognitive and socioemotional outcomes, available for the sample of children under 5 years old at baseline.[9] After regular data collection, attrition was more than 30% and 4 percentage point higher in the control (significant at the 10%). Attrition in the regular tracking phase was due to nonresponse and inability to locate the children at the moment of the survey visit, in addition to (often temporary) migration. In the intensive tracking phase, the field team visited the target communities repeatedly to recover temporary absence and implemented extensive tracking of migrants, resulting in very similar final attrition rates in treatment and control of 5%. During the ITP, 85% of those not found during the regular tracking phase were found and interviewed.

The other two datasets were collected for the mid- and long-term evaluations of YOP (Youth Opportunities Program), a cash grant program for youth in Uganda (Blattman, Fiala, and Martinez 2014, 2020) for an RCT that randomly allocated 265 youth groups to treatment and 270 to control in 2008. We focus on the impact of the program on income in 2012 and 2017, 4 and 9 years after baseline. The primary outcome is the standardized income family index used in Blattman, Fiala, and Martinez (2020). The index is composed of three measures: monthly net earnings, nondurable consumption, and durable assets. Attrition after regular tracking was between 25% and 40% and was unbalanced in 2017. These studies used intensive tracking on random subsamples, of 38.5% in 2012 and 36% in 2017, selected each year among all those not found during the regular tracking phase. The effective response rates after complete tracking were 84% in 2012 and 87% in 2017. Despite intensive tracking, attrition is unbalanced after complete tracking in 2012 but not in 2017. To account for this unbalanced attrition, the authors used IPW as their main specification.

We compare ITT estimates obtained after regular tracking with those obtained after complete tracking. The former represents the results that would have been obtained if only common tracking rules would have been applied to the survey sample. The later represents the benchmark estimate after exhaustive

[9] The cognitive and socioemotional index is composed of the following measures: the TVIP (the Spanish-speaking version of the Peabody Picture Vocabulary Test), the language and social-personal subtests of the Denver Developmental Screening Test, a short-term memory test, an associative memory test, and the Behavior Problem Index (BPI).

tracking but without further correction for remaining attrition. We also separately show ITT estimates for the subsample tracked during the intensive phase. Results in tables A1–A4 show that point estimates of the effects on the primary outcomes would have been 30%–44% larger in three out of four cases, if the data collection had been stopped after regular tracking. The difference in point estimates is also large, but in the opposite direction, in Blattman, Fiala, and Martinez (2014), who reported a relatively large imbalance in response rates after regular tracking.

III. Inverse Probability Weights

A. Correlates of Attrition

Results in the previous section confirm the importance of understanding the correlates of attrition to make informed assumptions about the nature of selection into the final sample, even when response rates are balanced by treatment group. To do so, we consider both the context and households' reaction to the program. Program participation can induce different types of individuals to migrate and attrit in treatment and control groups, even if the same number of people on average leave the sample. Individuals that have moved out of the study region, before, during or after the program will be harder to find, as are individuals who temporarily migrate for work or family reasons. These migration decisions can be affected either directly or indirectly by the randomized program exposure but could also capture the heterogeneity of the population. Covariates of attrition likely differ from context to context, but broadly speaking, differences in socioeconomic status, existing networks, family structures, and temporary residence are candidates to help explain differential attrition.

Tables A5–A8 show average baseline values of normalized household and individual characteristics for these covariate categories. Tables show values for individuals found after CTP, after RTP, for the subsample of those found during the ITP, and finally, for those missing. The last two columns show differences in means between the CTP sample and ITP sample with respect to the attritors sample. In the four cases, the differences are significantly different from zero for a number of categories, in line with them being either correlates of migration or of the accuracy of the migrant destination information obtained. Tables further show that those found in the intensive phases tend to be more like those not found along many dimensions. This is particularly notable for Blattman, Fiala, and Martinez (2014) and Barham, Macours, and Maluccio (2017), the two cases with samples that were balanced after RTP. As discussed earlier, similar findings have been reported for other panel surveys with intensive tracking protocols.

The differences between the attritors and nonattritors, and the similarities between those found or not found in the intensive phase, are important, as program impacts may well differ along many of these same dimensions. Moreover, it is not a priori obvious in which direction this would affect program impacts. For example, the transfer program may have induced low-SES children to get more education and subsequently migrate when they otherwise would not have or, on the contrary, may reenforce existing differences.

Evidence from the four cases under study shows that even after CTP, attrition remains selective, as baseline characteristics remain significant predictors of attrition. As a result, in the presence of heterogeneous treatment effects, ITT estimates after complete tracking will not reflect ITT estimates for the entire target population. Calculating weights to correct for selective attrition can therefore be valuable, and the information from the intensive tracking phase can be used to obtain an alternative set of weights based on a different set of assumptions than those used for standard IPW.

B. Obtaining Probability Weights Using Information from Tracking

Under the assumption of selection based on observables, unbiased estimates can be obtained using weighted least square regressions (Fitzgerald, Gottschalk, and Moffitt 1998; Wooldridge 2002a). The standard procedure to construct IPW consists of estimating the probability of being surveyed, conditional on a set of covariates, using the complete target population. If the correlates of attrition are significantly different between treatment arms, we obtain separate weights for each experimental group. Applying these weights adjusts for the differences in baseline characteristics between treatment arms that arise because of attrition. Behrman, Parker, and Todd (2009), for instance, construct IPW by treatment arm when estimating medium-term impacts of the PROGRESA/Oportunidades CCT program in Mexico.

When intensive tracking is done, there is additional, potentially valuable information to calculate weights for the attrition selection correction. If those found during the intensive tracking phase are more similar to attritors, we can model selection among those individuals that were hard to find. We start from the insight that individuals tracked during the intensive phase (whether found or not) are similar on observed characteristics. We define the response rate at endline as $R_i = R_i^{\mathrm{RTP}} + (1 - R_i^{\mathrm{RTP}})R_i^{\mathrm{ITP}}$, where R_i^{ITP} is an indicator variable taking the value of 1 if individual i was found after conducting intensive tracking, and R_i^{RTP} is an indicator variable taking the value of 1 if individual i was found after conducting regular tracking. Then, the probability to be found conditional on observable characteristics (X, Z) can be expressed as $P(R_i = 1|X, Z) = P(R_i^{\mathrm{RTP}} = 1|X, Z) + P(R_i^{\mathrm{ITP}} = 1|X, Z, R_i^{\mathrm{RTP}} = 0) \times (1 - P(R_i^{\mathrm{RTP}} = 1|X, Z))$.

Following Fitzgerald, Gottschalk, and Moffitt (1998), we account for a set of baseline variables that may be driving selection (henceforth, Z), and a vector of baseline covariates that form part of a basic model of interest (henceforth, X).

We modify the IPW estimates and construct weights using only the individuals tracked during the intensive tracking phase. The basic assumption underlying our estimation strategy is that individuals that were hard to find are similar in unobservables to those that ultimately were not found. We assume $y_i^1, y_i^0 \perp R_i^{\mathrm{ITP}} | X, Z, R_i^{\mathrm{RTP}} = 0$, where y_i^1, y_i^0 are potential outcomes.

We obtain the target population density function by weighting the conditional density in the ITP sample using the inverse of the predicted probability to be found during the ITP and weighting the conditional density in the RTP sample using a weight equal to 1. Compared with the standard IPW, the new IPW accounts for differential sample selection during different tracking phases.[10]

Given that the RTP sample is larger than the sample of ITP, the differential characteristics between the ITP sample and the attritors sample are diluted when we apply the standard IPW. This may need to be traded with a smaller sample size in the intensive tracking sample, possibly limiting the number of covariates that can be used in the prediction to obtain intensive-tracking specific weights. Larger intensive tracking samples will allow us to account for more potential drivers of selective attrition without running into overfitting concerns in the prediction model.

To estimate the different weights, we categorize the variables into X and Z. While this categorization is somewhat arbitrary, we think of the SES variables (such as strata fixed effects and household assets) as mostly capturing X, the variables related to the temporary nature of residence, networks, and regions as Z, while the demographic variables can be classified as either X or Z. This gives as potential predictors a wide set of baseline variables capturing the socioeconomic status, the demographics, the baseline networks, and the possible temporary nature of households' baseline residence. As there is a wide set of observed characteristics to consider, and as there are relatively few observations not found after intensive tracking, we follow Doyle et al. (2017) to reduce the set of predictors.[11] We first estimate bivariate regressions in which each potential predictor was tested to determine whether a significant difference existed between those found and those not found. All estimates use population weighted observations and standard errors clustered at the locality or group level,

[10] If the attriters and the nonattriters from the intensive phase are close in terms of X and Z, it may not be necessary to include X and Z in the weight estimates, i.e., $y_i^1, y_i^0 \perp R_i^{\mathrm{ITP}} | X, Z, R_i^{\mathrm{RTP}} = 0$ and $R_i^{\mathrm{ITP}} \perp X, Z | R_i^{\mathrm{RTP}} = 0$ implies $y_i^1, y_i^0 \perp R_i^{\mathrm{ITP}} | R_i^{\mathrm{RTP}} = 0$.

[11] An alternative set of predictions for weights was obtained using the least absolute shrinkage and selection operator (Lasso). See app. B for discussion and Imbens (2015) for a related approach.

depending on the case. The testing was conducted separately by treatment arm. Any measure found to be statistically significant for the early or late treatment group was retained as a potential predictor. We then estimate the probability of being found on this set of baseline predictor variables separately for each experimental group. In order to account for collinearity between measures, the baseline predictor set was further restricted by conducting stepwise selection of variables with backward elimination and using the adjusted R^2 as information criteria. The strata, age, and regional fixed effects were included as fixed predictors in all regressions.

In the last step, we estimate the probability of being found for both treatment arms together, keeping only the predictors as indicated by the stepwise procedure. Following Thomas et al. (2012), we also included interviewers' characteristics if available (fixed effects for the team that first visited a village during the regular tracking) to capture differences between teams in effectiveness of obtaining information for tracking.

Tables 2–5 show the results, which confirm that selection is not random in any of the tracking phases. Although, we observe variation across studies, the estimates for the intensive tracking phases in particular have good predictive power (R^2 between 25 and 51), and the predictive power is higher than for estimates on the full sample or the regular tracking sample. Restricting the sample for estimating weights to the ITP hence helps reduce measurement error in the estimates of the weights. The bottom panel in each table shows F-statistics for joint significance tests by groups of covariates.[12] Drivers of selection among those tracked during ITP differ markedly from those tracked during RTP. Table 2 shows that for RPS Z variables are much more predictive in ITP than in RTP, while X variables are more predictive in RTP than ITP. Tables 4 and 5 show the opposite pattern for YOP, with demographic variables also much more important in RTP than ITP. Differences between covariates' predictive power in RTP and ITP are less pronounced in table 3, in line with the different underlying reasons for attrition (temporary absence) in RTP in this study. The bottom panel of tables 2–5 further show that drivers of sample selection are different for treatment and control groups in different tracking phases. This finding points to a potential threat to the monotonicity assumption needed for other attrition selection corrections (Lee 2009; Behaghel et al. 2015).

C. *Results with Inverse Probability Weights*

Figures 1–4 show weighted least squares (WLS) estimates for assignment to treatment on the main outcome of each study. Each figure shows estimates using

[12] See table A10 for a list of all the covariates retained in each specification.

TABLE 2

LINEAR PROBABILITY MODEL CORRELATES OF ATTRITION—RED DE PROTECCIÓN SOCIAL (BARHAM, MACOURS, AND MALUCCIO 2017)

	CTP Sample		RTP Sample		ITP Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	.691		.103		−.581	
	(.82)		(.29)		(.38)	
Stratification fixed effects (X)	Yes		Yes		Yes	
Three monthly age dummies	Yes		Yes		Yes	
RTP survey supervisor fixed effects	Yes		Yes		Yes	
Observations	1,138		1,138		297	
R^2	.23		.22		.51	
Adjusted R^2	.19		.16		.38	
Joint significance tests with covariates and interaction terms together:						
F-statistic region, residence, and network (Z)	12.131		6.145		11.186	
Degrees of freedom	14		9		16	
F-statistic demographics	3.812		4.708		3.671	
Degrees of freedom	8		16		10	
F-statistic age	4.726		3.017		4.291	
Degrees of freedom	16		16		16	
F-statistic SES and strata (X)	8.021		9.093		5.439	
Degrees of freedom	24		28		18	
Joint significance tests with covariates and interaction terms separately:						
F-statistic region, residence, and network (Z)	16.216	4.242	2.410	4.131	16.032	11.551
Degrees of freedom	7	7	5	5	8	8
F-statistic demographics	2.593	3.837	3.464	2.328	4.808	6.687
Degrees of freedom	4	4	8	8	5	5
F-statistic age	6.364	1.685	3.090	1.518	3.703	2.871
Degrees of freedom	8	8	8	8	8	8
F-statistic SES and strata (X)	6.194	3.319	5.348	4.106	5.292	4.378
Degrees of freedom	12	12	14	14	9	9

Note. Respective values for "CTP Sample," "RTP Sample," and "ITP Sample" are for the test on the interaction between variables included in the model (cols. 1, 3, and 5) and assignment to early treatment (cols. 2, 4, and 6). The full list of controls retained for each of the samples' prediction models following the procedure described in Doyle et al. (2017) is presented in table A10. Standard errors, clustered at locality level, are in parentheses.

(i) the main (benchmark) specification from the original study on the RTP sample; (ii) standard IPW on the RTP sample; (iii) the main (benchmark) specification from the original study on CTP; (iv) standard IPW on the CTP sample; and (v) new IPW on the CTP sample. Estimates are shown for the set of covariates used by the authors of each reference paper and alternatively use post-double-Lasso controls (Belloni, Chernozhukov, and Hansen 2014). Standard errors in the main estimates do not account for the imprecision that comes from the fact that weights are estimated in the first step. As weights are estimated using the ITP sample, which by construction has a more limited number of observations

TABLE 3
LINEAR PROBABILITY MODEL CORRELATES OF ATTRITION—ATENCIÓN A CRISIS
(MACOURS, SCHADY, AND VAKIS 2012)

	CTP Sample		RTP Sample		ITP Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−.003		−.108		−.017	
	(.04)		(.14)		(.12)	
Stratification fixed effects	Yes		Yes		Yes	
Age dummies	Yes		Yes		Yes	
Observations	2,870		2,870		843	
R^2	.09		.10		.25	
Adjusted R^2	.06		.07		.19	
Joint significance tests with covariates and interaction terms together:						
F-statistic residence (*Z*)	3.858		3.637		3.295	
Degrees of freedom	6		6		6	
F-statistic demographic (*X*, *Z*)	3.364		1.295		3.194	
Degrees of freedom	12		8		8	
F-statistic SES and strata (*X*)	10.327		17.862		13.666	
Degrees of freedom	42		40		42	
F-statistic age	.622		1.057		.403	
Degrees of freedom	11		11		11	
Joint significance tests with covariates and interaction terms separately:						
F-statistic residence (*Z*)	6.227	2.748	2.354	1.901	4.984	1.270
Degrees of freedom	3	3	5	4	3	3
F-statistic demographic (*X*, *Z*)	2.009	1.092	.065	.754	1.491	.547
Degrees of freedom	6	6	4	4	4	4
F-statistic SES and strata (*X*)	8.220	8.784	17.762	18.145	8.323	9.708
Degrees of freedom	37	37	36	36	37	37

Note. Respective values for "CTP Sample," "RTP Sample," and "ITP Sample" are for the test on the interaction between variables included in the model (cols. 1, 3, and 5) and assignment to early treatment (cols. 2, 4, and 6). The full list of controls retained for each of the samples' prediction models following the procedure described in Doyle et al. (2017) is presented in table A10. Standard errors, clustered at locality level, are in parentheses.

than the full sample, small sample bias is a potential concern. We therefore also present alternative estimates using bootstrapped confidence intervals (in table A9) to account for the first-step prediction. These are broadly similar and indeed often show somewhat tighter confidence intervals.[13] In general, differences in point estimates between estimators are substantial in magnitude, even if confidence intervals overlap.

Figure 1 shows how the ITT, but also the standard IPW after RTP, leads to higher point estimates than the three estimates obtained with the complete sample. Estimates with weights obtained using only the sample from the intensive tracking phase lead to similar results than weights obtained from the

[13] This is in line with Wooldridge (2002b), who shows that computing the asymptotic variance of WLS estimates ignoring that the probability to be surveyed was predicted in a first step leads to conservative standard errors.

TABLE 4
LINEAR PROBABILITY MODEL CORRELATES OF ATTRITION—YOUTH OPPORTUNITY PROGRAM 4 YEARS (BLATTMAN, FIALA, AND MARTINEZ 2014)

	CTP Sample		RTP Sample		ITP Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	.081		2.432***		.248	
	(.20)		(.82)		(.26)	
Stratification fixed effects	Yes		Yes		Yes	
Observations	2,111		2,677		403	
R^2	.29		.13		.35	
Adjusted R^2	.27		.11		.28	
Joint significance tests with covariates and interaction terms together:						
F-statistic residence (*Z*)	79.826		42.713		3.665	
Degrees of freedom	5		5		2	
F-statistic demographics	3.771		5.961			
Degrees of freedom	2		6			
F-statistic SES and strata (*X*)	4.522		3.005		49.466	
Degrees of freedom	38		57		34	
Joint significance tests with covariates and interaction terms separately:						
F-statistic residence (*Z*)	91.430	103.465	58.731	42.803	3.543	3.788
Degrees of freedom	2	2	2	2	1	1
F-statistic demographics	4.874	2.669	5.640	6.283		
Degrees of freedom	1	1	3	3		
F-statistic SES and strata (*X*)	4.434	4.611	2.823	3.213	46.353	52.578
Degrees of freedom	19	19	29	29	17	17

Note. Respective values for "CTP Sample," "RTP Sample," and "ITP Sample" are for the test on the interaction between variables included in the model (cols. 1, 3, and 5) and assignment to early treatment (cols. 2, 4, and 6). The full list of controls retained for each of the samples' prediction models following the procedure described in Doyle et al. (2017) is presented in table A10. Standard errors, clustered at group level, are in parentheses.
*** $p < .01$.

complete tracking sample, despite being based on quite a different prediction model, as shown in table 2. As such they provide a useful robustness test to the standard IPW after CTP. The biggest difference in estimates is obtained when moving from the regular tracking to the full tracking sample, suggesting that putting effort in tracking remains the first best response to limiting attrition bias.

Figure 2 shows how the new IPW estimate compares with the ITT and a standard IPW estimate for Atención a Crisis. As before, results show that ITT estimates declined as a result of intensive tracking but that they are similar for the ITT and the IPW estimates. The similarity in results with the three estimators after complete tracking is not surprising, since with less than 5% of observations missing, selection is minimal. Moreover, given the young ages of the children and the outcomes studied, it seems plausible there is not necessarily much correlation between the remaining migration and program impacts. The similarity between the ITT and the IPW after regular tracking, on the other

TABLE 5

LINEAR PROBABILITY MODEL CORRELATES OF ATTRITION—YOUTH OPPORTUNITY PROGRAM 9 YEARS (BLATTMAN, FIALA, AND MARTINEZ 2020)

	CTP Sample		RTP Sample		ITP Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−.906		−.439		−.833	
	(.79)		(.74)		(.79)	
Stratification fixed effects	Yes		Yes		Yes	
Observations	2,086		2,677		183	
R^2	.10		.13		.33	
Adjusted R^2	.07		.11		.13	
Joint significance tests with covariates and interaction terms together:						
F-statistic residence (Z)	7.680		59.316		7.956	
Degrees of freedom	3		4		2	
F-statistic demographics	5.163		8.470		2.015	
Degrees of freedom	6		6		6	
F-statistic SES and strata (X)	2.094		1.503		8.458	
Degrees of freedom	38		44		33	
Joint significance tests with covariates and interaction terms separately:						
F-statistic residence (Z)	7.608	7.823	75.026	43.606	.003	15.909
Degrees of freedom	2	1	2	2	1	1
F-statistic demographics	5.497	4.829	12.574	4.366	3.322	.708
Degrees of freedom	3	3	3	3	3	3
F-statistic SES and strata (X)	1.748	2.439	.776	2.231	6.865	9.631
Degrees of freedom	19	19	22	22	17	17

Note. Respective values for "CTP Sample," "RTP Sample," and "ITP Sample" are for the test on the interaction between variables included in the model (cols. 1, 3, and 5) and assignment to early treatment (cols. 2, 4, and 6). The full list of controls retained for each of the samples' prediction models following the procedure described in Doyle et al. (2017) is presented in table A10. Standard errors, clustered at group level, are in parentheses.

hand, indicates that the standard IPW did not manage to introduce the relevant sample correction after regular tracking.

Figures 3 and 4 compare the ITT and WLS estimates using the standard IPW with estimates with the new IPW method for YOP.[14] To estimate the IPW, we use the set of covariates used by Blattman, Fiala, and Martinez (2020) to model sample selection on the full sample. To estimate the new IPW, we use a vector of covariates selected following the steps in section III.B. As before, the point estimates obtained after RTP, whether using standard IPW or not, are quite different from the estimates after complete tracking. In this case, the data added during the intensive tracking leads to higher estimates, and this holds using sample weights, standard IPW, or the new IPW. The point estimates of the last

[14] Blattman, Fiala, and Martinez (2020) do not report ITT estimates, given the unbalanced attrition, but they are included here for comparison purposes.

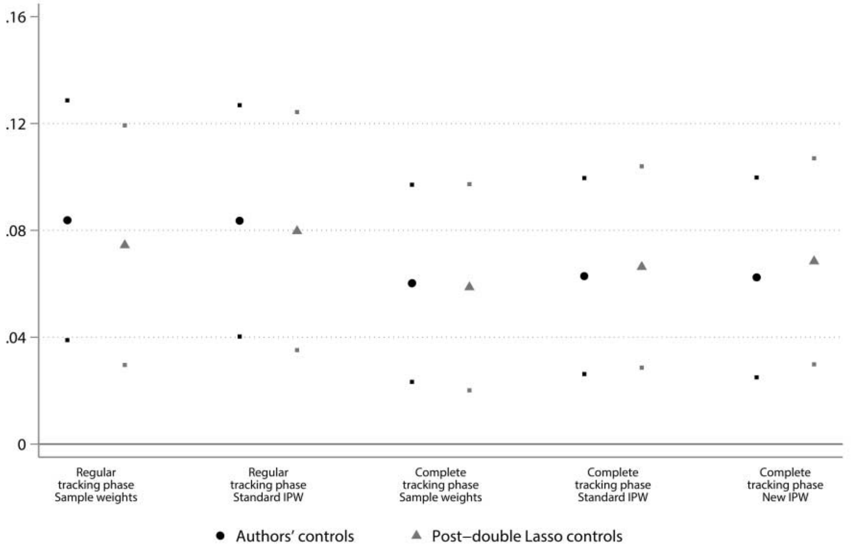

Figure 1. WLS estimates for assignment to treatment on the main outcome of the study on off-farm employment (Barham, Macours, and Maluccio 2017). Estimates are based on WLS regressions using sampling weights, standard IPW, and new IPW. Each estimate comes from a regression of the dependent variable on an indicator for assignment to treatment, strata fixed effects, three monthly age fixed effects, a set of dummies indicating whether individual had 1, 2, 3, or at least 4 years of education at baseline, and region fixed effects. Specification in markers 2, 4, 6, 8, and 10 includes controls selected using the post-double-selection Lasso procedure. The sampling strata, the three monthly age dummies, the education dummies, and the region dummies are partialled out. Confidence intervals are built using statistical significance at the 10% level. Standard errors are clustered at locality level.

being the highest, possibly suggesting it helped further correct for the remaining selection after CTP.

A similar pattern emerges with results from the 9-year evaluation in figure 4 but with the bias after RTP working in the opposite direction. Applying the new IPW drives the point estimate close to zero. Hence, in both the 4-year and the 9-year studies, the correction introduced by overweighting those that are hard to find using the new IPW was larger than that of the standard IPW.

Given that the reason and the selective nature of attrition, as well as the extent to which it interacts with possible heterogeneous treatment effects, will differ between studies, the new IPW approach proposed in this paper will logically not always lead to a sizable correction as compared with the ITT estimates. Overall, we conclude that using the information from the intensive tracking phase to overweight observations that were hard to find can help correct for differential selective attrition and that this offers a valuable robustness test when studying interventions that can directly affect the decision to migrate or when considering impacts on mobile cohorts, as is the case in Blattman, Fiala, and Martinez (2014, 2020) and Barham, Macours, and Maluccio (2017). This method can be used also when intensive tracking is only done on

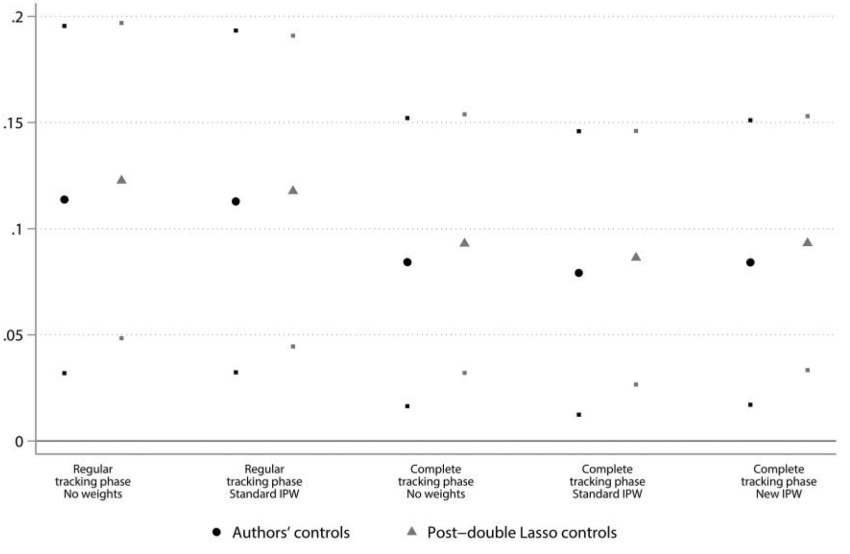

Figure 2. WLS estimates for assignment to treatment on the main outcome of the study on cognitive and socioemotional index (Macours, Schady, and Vakis 2012). Estimates are based on WLS regressions using sampling weights (markers 1 and 2, and markers 5 and 6), standard IPW (markers 3 and 4, and markers 7–8), and new IPW (markers 9 and 10). The estimation sample in markers 1–4 consists of individuals found during the RTP; the estimation sample in markers 5–10 consists of individuals found during the CTP. Each estimate comes from a regression of the dependent variable on an indicator for assignment to treatment, strata fixed effects, six monthly age fixed effects, and gender fixed effects. Specification in markers 2, 4, 6, 8, and 10 includes controls selected using the post-double-selection Lasso procedure. The sampling strata, the six monthly age dummies, and the gender dummies are partialled out. Confidence intervals are built using statistical significance at the 10% level. Standard errors are clustered at locality level.

a random subsample, though the estimation of the weights will be facilitated by having a relatively large subsample of individuals tracked during the ITP to estimate the model of attrition.

IV. Conclusion

Attrition can affect external and internal validity of any impact evaluation, and this is particularly relevant for studies involving youth and other mobile populations in developing countries. This paper analyzes attrition bias in four randomized experiments and shows the sensitivity of the ITT estimates to different assumptions regarding attrition and related data-collection strategies even when attrition rates are balanced. It shows that tracking protocols commonly found in the literature that focus efforts on the region of study and limited number of revisits can lead to substantial under- or overestimates.

The paper further proposes an alternative method to account for sample selection driven by observable characteristics that takes advantage of information obtained through intensive tracking and the observation that those found after intensively tracked are often most similar to those not found. Building

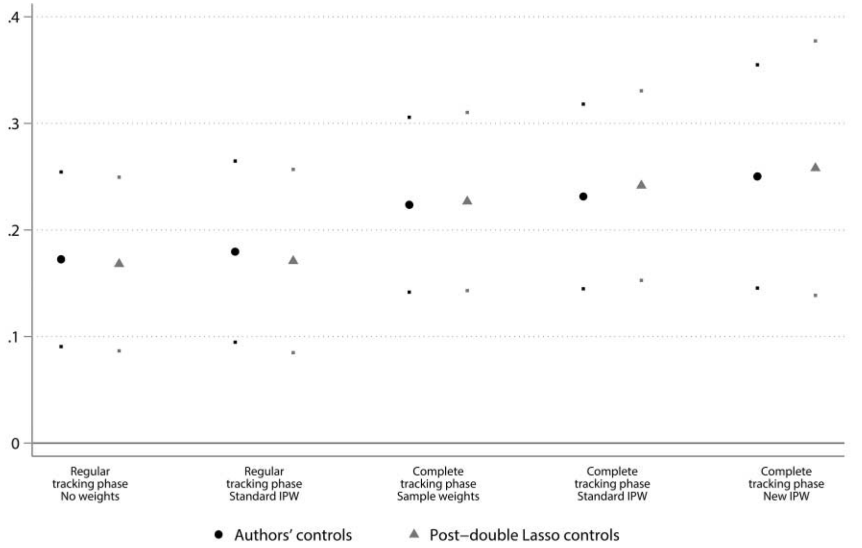

Figure 3. WLS estimates for assignment to treatment on the standardized income index: treatment effects after 4 years (Blattman, Fiala, and Martinez 2014). Estimates are based on OLS regressions in markers 1 and 2 and based on WLS regressions using standard IPW (markers 3 and 4, and markers 7 and 8), sample weights (markers 5 and 6), and new IPW (markers 9 and 10). The estimation sample in markers 1–4 consists of individuals found during the RTP; the estimation sample in markers 5–10 consists of individuals found during the CTP. Each estimate comes from a regression of the dependent variable on an indicator for assignment to treatment and district fixed effects. Specification in markers 1, 3, 5, 7, and 9 includes a vector of 38 baseline covariates (for details, see Blattman, Fiala, and Martinez 2020). Specification in markers 2, 4, 6, 8, and 10 includes controls selected using the post-double-selection Lasso procedure. The sampling strata are partialled out. Confidence intervals are built using statistical significance at the 10% level. Standard errors are clustered at group level.

on the literature that proposes attrition corrections through reweighting, we assign a weight equal to one to those found during a regular tracking phase, and we predict the probability to be found for those found during an intensive tracking phase and those not found. We compare estimates using these new inverse probability weights, with ITT estimates that do not account for attrition and with standard WLS estimates. We show that following regular tracking practices similar to those in most empirical studies would have led to substantial overestimates or underestimates and that inferences from regular attrition correction methods would not necessarily have helped to predict the direction of the bias, possibly due to violation of the monotonicity assumption and heterogeneous treatment effects.

Conducting intensive tracking reduces attrition bias leading to more robust estimates and allows for accounting for heterogeneous treatment effects. In three of the four studies analyzed, we find that not including those found during the ITP leads to an overestimate of the ITT effects on the outcomes of interest. The results highlight the importance of studying attrition bias even in projects with balanced attrition rates.

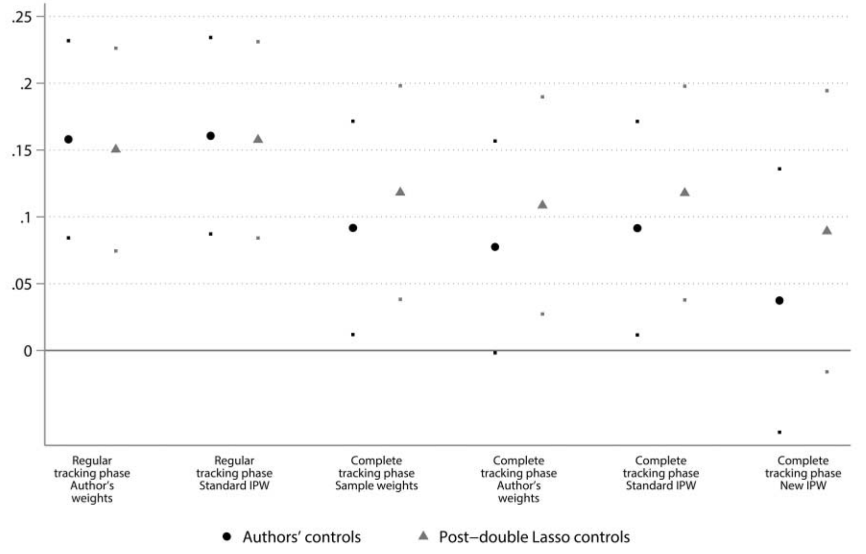

Figure 4. WLS estimates for assignment to treatment on standardized income index: treatment effects after 9 years (Blattman, Fiala, and Martinez 2020). Estimates are based on WLS regressions using authors' IPW (markers 1, 2, 7 and 8), standard IPW (markers 3 and 4, and markers 9 and 10), sample weights (markers 5 and 6), and new IPW (markers 11 and 12). The estimation sample in markers 1–4 consists of individuals found during the RTP; the estimation sample in markers 5–12 consists of individuals found during the CTP. Each estimate comes from a regression of the dependent variable on an indicator for assignment to treatment and district fixed effects. Specification in markers 1, 3, 5, 7, 9, and 11 includes a vector of 38 baseline covariates (for details, see Blattman, Fiala and Martinez 2020). Specification in markers 2, 4, 6, 8, 10, and 12 includes controls selected using the post-double-selection Lasso procedure. The sampling strata are partialled out. Confidence intervals are built using statistical significance at the 10% level. Standard errors are clustered at group level.

As opportunities for long-term follow-ups of RCTs increase, the trade-off between tracking costs and attrition bias is likely to become relevant for an increasing number of studies. Using detailed data on the tracking process from Barham, Macours, and Maluccio (2017), we evaluate the costs of in-person tracking in terms of the number of enumerator working days (i.e., the number of days the team worked times the number of enumerators in the team at each moment). During the regular tracking phase, the team worked 91 days, and the cost to find and interview the RTP sample was 1,486 enumerator days. This allowed us to obtain information on more than 70% of the target sample, as well as information on migrants' destination. To track and interview the ITP sample, a smaller enumerator team worked an additional 218 days and a total 905 additional enumerator days (see fig. 5), increasing the share of the target sample successfully interviewed to 89%. Hence, each individually interviewed in the intensive tracking phase was more than twice as expensive in enumerator time alone, while also being more costly because of higher transport and coordination costs involved

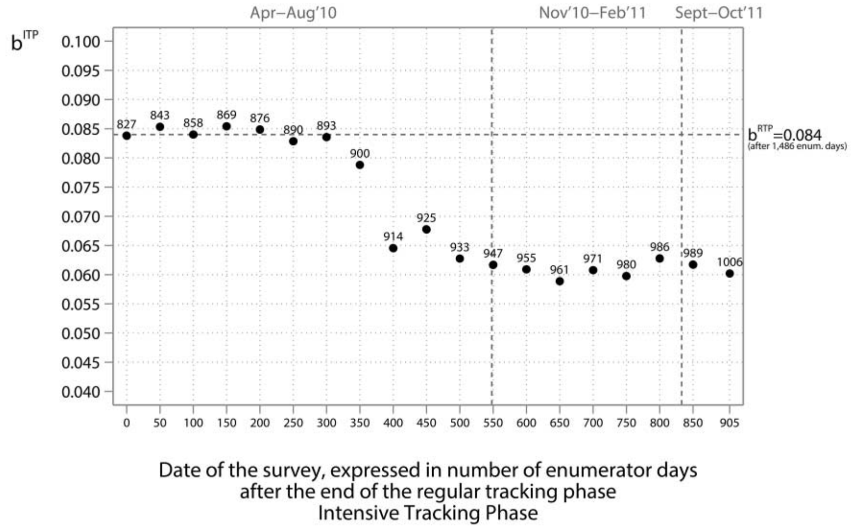

Figure 5. Cost analysis: ITT estimates on off-farm employment during the ITP program by Red de Protección Social (Barham, Macours, and Maluccio 2017). Estimates are based on WLS on the dependent variable on an indicator for assignment to treatment, strata fixed effects, three monthly age fixed effects, a set of dummies indicating whether the individual had 1, 2, 3, or at least 4 years of education at baseline, and region fixed effects. All estimates use the survey sample weights. The figures plot ITT estimates (β) on the sample of respondents found during the RTP, plus those found during the ITP after *X* days of enumerators days. An enumerator day is defined as any working day in which the team of enumerators worked after RTP (March 23, 2010) times the number of enumerators in the team at each date. Numbers above the dots show the number of male respondents between the ages of 9 and 12 years found until that day. The horizontal line shows the value of the ITT estimate after conducting RTP. The vertical lines mark the different phases of the ITP (see app. C).

with tracking.[15] Although the cost of obtaining the hard-to-find individuals was hence clearly nonnegligible, the evidence in this paper suggests it can be hard to predict the direction of attrition bias without such intensive tracking. The paper further shows that having data on a subset of individuals that was hard to find provides an alternative estimator to correct for attrition bias, which can help test sensitivity of the results to assumptions underlying the standard IPW. The new IPW method proposed in this paper can be applied if only a random subset is tracked intensively, as long as the sample size of those sampled to be intensively tracked is large enough.

The costs and benefits involved in intensively tracking a random subset or the full sample of those missing after a regular tracking phase will, more generally,

[15] Accounting for survey breaks, the regular tracking phase spanned a 5-month period, and the intensive tracking phase spanned 1.5 years. As fig. 5 shows, most of the observations in the intensive tracking phase were collected in the 5 months directly following the regular tracking phase. This difference in timing between the regular and most of the observations of the intensive tracking phases is small compared with the 10-year period since baseline and is unlikely to be driving the results (as also suggested by the stability of the point estimates in fig. 5 from August 2010 onward).

be a function of a number of factors. Intensively tracking more individuals will become more important when there are many possible factors driving selective attrition in a given context (and hence the set of covariates to include in the selection estimation is large), when there are reasons to believe treatment heterogeneity is large, or when power calculations indicate that minimizing attrition is needed to have sufficient precision of the estimates. On the other hand, when costs of tracking each individual missing after the regular tracking phase are high, when sample sizes are large, or when potential drivers of selective attrition are limited, tracking only a random subset allows us to reduce costs and still obtain reliable predictions and weights. Tracking costs will depend on a number of context-specific factors, including the spatial dispersion of migrants and the related costs of locating an individual, conditional on having located others and on the technology and accessibility that can help facilitate tracking. Phone surveys can limit some of these costs and, depending on the primary outcomes of interest, can offer an alternative to obtain high tracking rates for reasonable costs. These different considerations can help provide guidance for intensive tracking decisions. Using the information from hard-to-find individuals to calculate attrition weights can then allow us to get more mileage out of such intensive tracking and, as such, help improve the trade-off between costs and potential bias.

References

Alderman, Harold, Jere Behrman, Hans-Peter Kohler, John A. Maluccio, and Susan Watkins. 2001. "Attrition in Longitudinal Household Survey Data: Some Tests from Three Developing Countries." *Demographic Research* 5:79–124.

Baird, Sarah, Joan Hamory, and Edward Miguel. 2008. "Tracking, Attrition, and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)." CIDER Working Paper no. C08–151, University of California, Berkeley.

Barham, Tania, Karen Macours, and John A. Maluccio. 2017. "Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings after 10 Years." CEPR Discussion Paper no. 11937, Center for Economic and Policy Research, Washington, DC.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. 2011. "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey." *Review of Economics and Statistics* 93, no. 3:1010–33.

Behaghel, Luc, Bruno Crépon, Marc Gurgand, and Thomas Le Barbanchon. 2015. "Please Call Again: Correcting Non-Response Bias in Treatment Effect Models." *Review of Economics and Statistics* 97, no. 5:1070–80.

Behrman, Jere R., Susan W. Parker, and Peter Todd. 2009. "Medium Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." In *Poverty, Inequality and Policy in Latin America*, 219–70. Cambridge, MA: MIT Press.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *Review of Economic Studies* 81, no. 2:608–50.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics* 129, no. 2:697–752.

———. 2020. "The Long Term Impacts of Grants on Poverty: 9-Year Evidence from Uganda's Youth Opportunities Program." *American Economic Review: Insights* 2, no. 3:287–304.

DiNardo, John, Justin McCrary, and Lisa Sanbonmatsu. 2006. "Constructive Proposals for Dealing with Attrition: An Empirical Example." Working paper, National Bureau of Economic Research, Cambridge, MA.

Doyle, Orla, Colm Harmon, James J. Heckman, Caitriona Logue, and Seong Hyeok Moon. 2017. "Early Skill Formation and the Efficiency of Parental Investment: A Randomized Controlled Trial of Home Visiting." *Labour Economics* 45:40–58.

Falaris, Evangelos M. 2003. "The Effect of Survey Attrition in Longitudinal Surveys: Evidence from Peru, Cote d'Ivoire and Vietnam." *Journal of Development Economics* 70:133–57.

Fitzgerald, John, Peter Gottschalk, and Robert A. Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources* 33, no. 2:251–99.

Fuwa, Nobuhiko. 2011. "Should We Track Migrant Households When Collecting Household Panel Data? Household Relocation, Economic Mobility, and Attrition Biases in the Rural Philippines." *American Journal of Agricultural Economics* 93, no. 1:56–82.

Hull, Peter. 2015. "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons." Working paper, Massachusetts Institute of Technology, Cambridge.

Imbens, Guido W. 2015. "Matching Methods in Practice: Three Examples." *Journal of Human Resources* 50, no. 2:373–419.

Lee, David S. 2002. "Trimming for Bounds on Treatment Effects With Missing Outcomes." NBER Working Paper no. 0277, National Bureau of Economic Research, Cambridge, MA.

———. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76, no. 3:1071–102.

Macours, Karen, Norbert Schady, and Renos Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* 4, no. 2: 247–73.

Maluccio, John A. 2004. "Using Quality of Interview Information to Assess Nonrandom Attrition Bias in Developing-Country Panel Data." *Review of Development Economics* 8, no. 1:91–109.

Thomas, Duncan, Elizabeth Frankenberg, and James P. Smith. 2001. "Lost but Not Forgotten: Attrition and Follow-Up in the Indonesia Family Life Survey." *Journal of Human Resources* 36:556–92.

Thomas, Duncan, Firman Witoelar, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, Cecep Sumantri, and Wayan Suriastini. 2012. "Cutting the Costs of Attrition: Results from the Indonesia Family Life Survey." *Journal of Development Economics* 98, no. 1:108–23.

Velasquez, Andrea, Maria E. Genoni, Luis Rubalcava, Graciela Teruel, and Duncan Thomas. 2010. "Attrition in Longitudinal Surveys: Evidence from the Mexican Family Life Survey." Proceedings of the Northeast Universities Development Consortium Conference, October 2010.

Wooldridge, Jeffrey M. 2002a. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

———. 2002b. "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification." *Portuguese Economic Journal* 1, no. 2:117–39.