Received: 13 July 2015 | Revised: 14 March 2017

DOI: 10.1002/jae.2583

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Difference-in-differences when the treatment status is observed in only one period

Irene Botosaru[1] | Federico H. Gutierrez[2]

[1]Department of Economics, Simon Fraser University, Burnaby, BC, Canada

[2]Department of Economics, Vanderbilt University, Nashville, TN, USA

Correspondence
Federico H. Gutierrez, Department of Economics, 415 Calhoun Hall, Vanderbilt University, Nashville, TN 37240, USA.
Email: federico.h.gutierrez@Vanderbilt.Edu

Summary

This paper considers the difference-in-differences (DID) method when the data come from repeated cross-sections and the treatment status is observed either before or after the implementation of a program. We propose a new method that point-identifies the average treatment effect on the treated (ATT) via a DID method when there is at least one proxy variable for the latent treatment. Key assumptions are the stationarity of the propensity score conditional on the proxy and an exclusion restriction that the proxy must satisfy with respect to the change in average outcomes over time conditional on the true treatment status. We propose a generalized method of moments estimator for the ATT and we show that the associated overidentification test can be used to test our key assumptions. The method is used to evaluate JUNTOS, a Peruvian conditional cash transfer program. We find that the program significantly increased the demand for health inputs among children and women of reproductive age.

1 | INTRODUCTION

Researchers in the social sciences often have to rely on observational data in order to evaluate causal effects of policy interventions. Statistical methods used to achieve this goal have to account for the possibility of selection bias, which may arise when beneficiaries of the policy are not randomly selected from the population. Difference-in-differences (DID) is a heavily used method that accounts for selection bias due to time-invariant unobserved covariates. DID identifies the average treatment effect on the treated (ATT) by comparing the difference in post- and pre-program outcomes between two distinct groups: a group that participates in the program (the treated) and a group that does not participate (the control). DID can be applied with either panel data or repeated cross-sectional data as long as the observed outcomes can be classified into treated and control both before and after the program.

The use of repeated cross-sectional data, however, raises issues related to the possibility of doing such a classification. For example, many of the nationally representative cross-sectional household surveys include questions regarding participation in active social programs as well as current (post-program) outcomes. However, they rarely contain retrospective information on pre-program outcomes. Moreover, surveys conducted in pre-program periods never contain information on participation in future (not yet implemented) programs. This lack of information renders the standard DID procedure infeasible.

One way around this missing data issue is to use a proxy for the latent treatment status in order to group pre-program outcomes into treated and control (Abadie, 2005). Since classifying individuals on the basis of a proxy may lead to misclassification, it is important to understand under what conditions the DID estimand using the proxy (instead of the true but latent treatment status) leads to the identification of the ATT. To the best of our knowledge, we are the first to derive such conditions. We believe that it is particularly important to do so since this type of missing data is quite prevalent in applied work.

Our interest in developing a method to deal with this missing data issue is motivated by the evaluation of social programs, where the problem of missing treatment status is particularly prevalent. In this paper we evaluate the impact of the Peruvian conditional cash transfer program JUNTOS on the demand for health inputs among women and children. The source of information to evaluate JUNTOS is the *Encuesta Nacional de Hogares* (ENAHO). The dataset lends itself naturally to our setup since it contains repeated cross-sections with information on the outcomes of interest in both pre- and post-program periods but with information about program participation only after the implementation of JUNTOS. We show how our proposed estimator can be applied to EHANO to study the effects of JUNTOS on the demand for health inputs. As a preview of our results, we find that JUNTOS increased contraceptive use among women of reproductive age by 14 percentage points. This is a relevant result since one of the concerns regarding conditional cash transfers is the possibility that these programs induce women to have more children in order to continue receiving the cash transfers.

An additional motivation for developing a new method rests on the fact that, in the past, researchers have resorted to solutions that, in the presence of misclassification, produce biased results for the ATT. For example, Groen and Polivka (2008) use data from the Current Population Survey (CPS) to perform a DID analysis of the effect of hurricane Katrina on the labor market outcome of evacuees. The treatment status here is being an evacuee due to hurricane Katrina. The identification problem arises because pre-Katrina rounds of the CPS do not contain identifiers of (future) evacuees. The authors address this problem by proxying the true treatment status of individuals in the pre-Katrina rounds with an indicator for living in an affected area. However, since not all individuals who lived in affected areas became evacuees, grouping pre-treatment outcomes by the geographical proxy leads to biased results if misclassification is not accounted for.

As another example, consider Buchmueller, DiNardo, and Valletta (2011), who perform a DID analysis with data from the CPS to study the effect of Hawaii's Prepaid Health Insurance Act on labor market outcomes. The program was implemented in 1974, while the CPS started collecting data on individual health insurance (the treatment status) in 1979, 5 years after the implementation of the program. Because of this timing issue, the authors treat the period 1979–1982 as the pre-treatment period, violating one of the main DID assumptions: that none of the individuals be treated in the pre-treatment period.

In this paper, we propose a new method that identifies the ATT via the DID method when the only available data come from repeated cross-sections with missing individual group membership in one of the two time periods. Our identification strategy is based on formulating the problem as a finite mixture problem.[1] To do so, we require the existence of a set of variables correlated with the treatment status. The set may contain one or multiple such variables, which we call "proxies." The proxies must be observed both before and after the treatment and they must satisfy the following key assumptions: (i) the proxies must not affect the change in average outcomes over time conditional on the true treatment status; and (ii) the propensity score conditional on the proxies must be stationary. Any predictor of the treatment status can be used as a proxy. For example, in our application, the set of proxies contains covariates known to have been used by the program organizers to determine eligibility.

An advantage of our method over methods that rely exclusively on selection on observables is that the set of proxies does not have to be complete. That is, it is not necessary for *all* variables that were used by the program organizers to determine eligibility to be included in the set of proxies, as long the proxies are jointly strongly correlated with the treatment status.[2] When eligibility variables are not observed, any strong predictor of the treatment status may be used. We illustrate this in our simulation studies in the Supporting Information Appendix.

Our identification strategy is constructive, in that the ATT can be estimated from data on the conditional means of pre-program outcomes, post-program outcomes, and the treatment status conditional on the set of proxies. However, given the popularity of linear models in applied work, we specify a linear parametric model for the dependent variable and we cast our identification assumptions as conditional moments. This leads to a generalized method of moments (GMM) estimation procedure for the identified DID estimand. We then propose to use the associated overidentification test for the two key identification assumptions that the proxies must satisfy.

Our paper is related to the literature on treatment effects with misclassification (see Hu, 2008; Lewbel, 2007; Mahajan, 2006). All these papers deal with cross-sections and they focus on the identification of the average treatment effect (ATE) by requiring the existence of both an instrumental variable and a proxy for the mismeasured treatment status. In contrast, our paper takes advantage of the time dimension of the DID setup, which allows us to require just one proxy for the treatment status. Other related papers are Chen et al. (2009), which identifies the ATE by imposing a symmetry assumption on the density of

[1]We are not the first to connect mixture models to treatment effects models. To our knowledge, one of the first papers to do this was Cross and Manski (2002).
[2]Previous papers have analyzed how the misspecification of the propensity score affects the consistency of matching methods (Heckman, Ichimura, Smith, & Todd, 1996; Heckman, Ichimura, Smith, & Todd, 1998; Heckman, Ichimura, & Todd, 1997; Lalonde, 1986; Smith and Todd, 2005). This problem is not present in our method.

the measurement error rather than relying on additional information such as proxies and instrumental variables, and Molinari (2010), which provides partial identification results for the ATE in a cross-sectional setting when the treatment status is missing.

Our paper is also related to de Chaisemartin and D'Haultfoeuille (2015) in that both papers consider the DID approach in the presence of imperfectly observed treatment status. The approach of de Chaisemartin and D'Haultfoeuille can also be applied to a setting where the pre-treatment outcomes cannot be classified into treated and control. However, their method and that of this paper differ in terms of both assumptions and identified parameters. When treatment effects are heterogeneous, the approach of de Chaisemartin and D'Haultfoeuille identifies a local average treatment effect (LATE), whereas our approach identifies the ATT.

An outline of the paper is as follows. Section 2 describes the JUNTOS program and the dataset we use to evaluate it. Section 3 considers identification of the ATT when the treatment status is missing in the pre-treatment period. Section 4 provides a simple parametric model that satisfies all the identification assumptions. Section 5 introduces the GMM estimator and the associated overidentification test. In Section 6 we use our method to evaluate the impact of JUNTOS. Section 7 concludes. Descriptive statistics can be found in the Appendix. In the Supporting Information Appendix, we provide additional proofs, Monte Carlo simulation results, robustness checks for our application, an additional estimator, and an additional illustration to the Mexican cash transfer program, PROGRESA.

2 | CONDITIONAL CASH TRANSFER PROGRAMS AND JUNTOS

Conditional cash transfers (CCT) are one of the most popular social programs in the world. A common characteristic of CCTs is the transfer of money to poor households with the condition that they invest in human capital. Currently, over 30 countries have some form of CCT. In 2009, the World Bank alone allocated US $2.4 billion to scale up and start CCT operations.[3]

The efficient and equitable allocation of resources demands a constant monitoring of CCTs. However, very few social programs have been implemented with the explicit objective of being evaluated. One exception is the well-known Progresa/Oportunidades, which was randomized among a group of villages by the government of Mexico. The design of such a randomized controlled experiment guarantees internal validity but says little about external validity. The effects of Progresa cannot be extrapolated to other countries where the institutions, culture, and socioeconomic conditions are different. Moreover, even within Mexico, the results of the evaluation obtained from the initial randomization that took place in rural areas cannot be used to infer the impact of Progresa in urban areas. Thus reliable methods that continuously evaluate social programs are important.

Data limitation is an important constraint to evaluating CCTs and other social programs. This forces researchers to use assumptions such as selection on observables (Perova & Vakis, 2012), to identify program effects only on a selected group Levy and Ohls (2003), or to simply not perform the evaluation. In this paper, we propose a new method to identify the ATT when the true treatment status is missing, preventing pre-program outcomes from being classified into treated and control.

2.1 | Description of JUNTOS

JUNTOS is a CCT program, which was implemented in Peru in 2005. The goal of the program is poverty reduction in both the short and the long run via human capital accumulation. JUNTOS consists of a monthly monetary transfer of 100 soles (30 USD) per household. The money is transferred to the mother with the condition that (i) children younger than 5 years old attend health controls, (ii) children who are 6–14 years old attend formal education at least 85% of the academic year, and (iii) pregnant women and lactating mothers receive pre- and postpartum health services (see Perova & Vakis, 2012).[4]

Figure 1 shows the evolution of JUNTOS from 2005 to 2013. The program was first implemented in 70 Peruvian districts in 2005. During 2006 and 2007 the program expanded to 638 districts, reaching more than 420,000 households. From 2008 to 2011 the districts and households enrolled in the program remained relatively constant. In 2012, JUNTOS incorporated new regions and changed the conditionality.

The dataset we use is *Encuesta Nacional de Hogares* (ENAHO), which is a nationally representative survey regularly conducted by the National Institute of Statistics of Peru (INEI). Since 2004, approximately 90,000 individuals have been interviewed annually. All rounds of ENAHO are comparable. Neither the interviewing methodology nor the framing of the questions has

[3]See The World Bank's web page: http://web.worldbank.org/.
[4]Details of JUNTOS can be found at www.juntos.gob.pe.

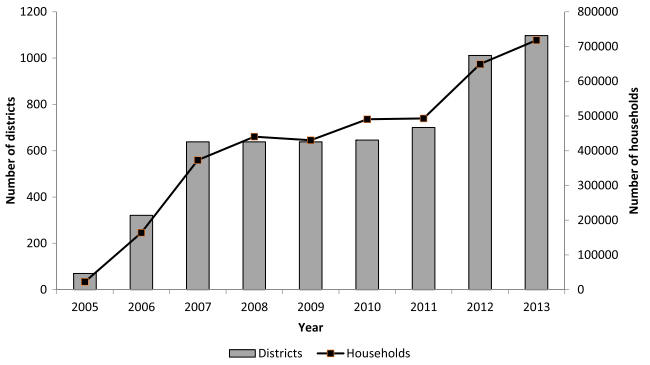

FIGURE 1 Evolution of JUNTOS. Source: www.juntos.gob.pe [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Descriptive statistics for women and children

	Obs.[a]	Mean	SD
Women 15–45 years old			
Beneficiary of JUNTOS[b]	13,911	0.53	0.50
Did not seek medical care when sick	11,805	0.63	0.48
Seek family planning advice (last 3 months)	18,836	0.16	0.37
Use contraceptives (last 3 months)	19,888	0.14	0.35
Children 0–5 years old			
Beneficiary of JUNTOS[+]	8,829	0.64	0.48
Did not seek medical care when sick	7,341	0.44	0.50
Got vaccinated	12,586	0.38	0.48

[a]Observations for which the relevant question was answered.

[b]Defined only for survey years 2008–2011.

changed since 2004. Before 2004, ENAHO had a different methodology with a much reduced set of questions. To avoid comparability problems due to methodological changes, we do not use the previous version of ENAHO.

Despite the fact that there is information on pre-program outcomes, individual treatment status is observed only in post-program surveys. In particular, JUNTOS was implemented in 2005 but 2008 is the first year when treatment status is observed. We restrict the sample to districts where JUNTOS was introduced between 2006 and 2007, excluding 70 districts where the program was operating by the end of 2005. The last post-treatment year we use is 2011, which is the last year before the program conditionality changed. Our sample then contains two pre-treatment years, 2004 and 2005, and four post-treatment years, 2008–2011.

Our goal is to analyze the effect of receiving JUNTOS on the demand for health inputs for women and children. In particular, the outcomes we analyze for women are (i) seeking medical care when sick, (ii) receiving advice in relation to family planning, and (iii) using contraceptives. The outcome we analyze for children are (i) seeking medical care when sick, and (ii) receiving vaccines. Table 1 shows summary statistics for women of reproductive age and for children. The variable JUNTOS is an indicator variable for whether the person lives in a household that is a beneficiary of the program. This variable is observed only from 2008 to 2011. Half of the women and approximately two-thirds of the children in the districts where the program was implemented were living in a household where at least one person was enrolled in JUNTOS. During the sample period, 63% of women and 44% of children did not seek medical care when sick. Contraceptive use and family planning is low, as well as child vaccination rates.

The households eligible for JUNTOS are those considered poor by the government.[5] Poverty status in Peru is determined according to an index that uses housing and household member characteristics. The complete list of variables is collected by the government using the *Unique Socioeconomic Form* (FSU in Spanish). The method for calculating the poverty index is not publicly known. This lack of information does not prevent us from using the method we propose in this paper, which requires the use of proxy variables for the treatment status. We use as proxies observed housing characteristics, such as roof, floor, and

[5]The official web page (www.juntos.gob.pe) indicates that participating households should (i) be classified as poor, (ii) live in the district for more than 6 months, and (iii) have a pregnant woman and/or children between 0 and 14 years old (0 and 19 years old after 2011).

TABLE 2 Descriptive statistics for the JUNTOS proxy variables

	Women 15–45 years old			Children 0–5 years old		
	JUNTOS	Control	Diff.[a]	JUNTOS	Control	Diff.[a]
Roof material						
Reed, tree leaves, mud	0.183	0.087	0.096	0.202	0.124	0.078
Concrete	0.461	0.510	−0.049	0.453	0.499	−0.046
Other	0.356	0.403	−0.047	0.345	0.377	−0.032
Floor material						
Concrete	0.034	0.203	−0.168	0.030	0.162	−0.131
Soil	0.908	0.661	0.247	0.911	0.689	0.222
Other	0.058	0.136	−0.078	0.059	0.149	−0.090
Wall material						
Bricks, concrete	0.005	0.082	−0.076	0.005	0.065	−0.061
Mud	0.890	0.775	0.115	0.864	0.737	0.127
Wood	0.036	0.059	−0.023	0.042	0.086	−0.044
Other	0.069	0.084	−0.015	0.089	0.112	−0.023
Max. education among adults	8.587	10.589	−2.002	7.957	9.522	−1.564
Observations	7,743	6,168		5,779	3,050	

[a]All differences are significant at the 1% level.

wall material, and the highest level of education of adults in the household. These variables are known to be used by program organizers to determine eligibility and are consequently highly correlated with the treatment status. Table 2 shows proxy averages for beneficiaries and non-beneficiaries of JUNTOS in post-treatment years. Women receiving JUNTOS live in houses that are of significantly lower quality. They are twice as likely to have roofs made of reed, tree leaves or mud, 80% less likely to have concrete floors, and 14% more likely to have walls made of mud. In addition to housing conditions, beneficiaries of the program are significantly less educated with two fewer years of formal education. We argue in the next section that these proxies also satisfy the other assumptions needed for identification.

3 | IDENTIFICATION

In this section, we first discuss the standard DID estimand and show that it identifies the ATT when individual group membership is observed in both time periods. Then we introduce our solution to doing a DID analysis when the treatment status is observed in only one period. We consider the case without observed covariates for simplicity of exposition. We show identification of the ATT with covariates in the Supporting Information Appendix. All assumptions in this section are motivated using our application to JUNTOS.

3.1 | Standard DID

The DID framework can be described as follows. Between periods $t = 0$ and $t = 1$, a fraction of a population is exposed to a treatment. The population is observed in both periods. The variable Y_t describes the individual outcome of interest at time t and the variable D stands for the individual treatment status. Let $D = 1$ if the individual has been exposed to the treatment between $t = 0$ and $t = 1$, and $D = 0$ otherwise. Those with $D = 1$ are called *treated* and those with $D = 0$ are called *control*. Note that D is time invariant.

At each time period, each individual has two potential outcomes: $Y_t(0)$ if the individual is not exposed to the treatment and $Y_t(1)$ if the individual is exposed to the treatment. The realized outcome is then given by

$$Y_t = DY_t(1) + (1 - D)\, Y_t(0). \tag{1}$$

The parameter of interest in this paper is the average treatment effect on the treated. It is defined in terms of potential outcomes as

$$\text{ATT} = E\left(Y_1(1) - Y_1(0) | D = 1\right). \tag{2}$$

The standard DID estimand, which is defined as

$$\theta \equiv \left[E\left(Y_1 | D = 1\right) - E\left(Y_1 | D = 0\right)\right] - \left[E\left(Y_0 | D = 1\right) - E\left(Y_0 | D = 0\right)\right] \tag{3}$$

identifies the ATT under the following assumptions.

Assumption 1. (no anticipation)
There is no anticipatory response for those in the treatment group:

$$E(Y_0(0)|D=1) = E(Y_0(1)|D=1).$$

Assumption 1 may be violated if individuals, anticipating participation in a program change their behavior before the program is implemented.

Assumption 2. (parallel paths)
In the absence of the treatment, the average outcomes for the treated group would have followed the same trend as that for the control group:

$$E(Y_1(0)|D=1) - E\left(Y_0(0)|D=1\right) = E\left(Y_1(0)|D=0\right) - E\left(Y_0(0)|D=0\right).$$

Assumption 2 is the key assumption for the conventional DID. This assumption imposes that, in the absence of the treatment, the average outcomes for the treated should have the same variation over time as the average outcomes for the control. Under this assumption, selection on time-invariant unobservables is allowed but selection on transitory shocks is not.

In our application, Assumption 2 implies that the difference in the demand for health inputs between beneficiaries and non-beneficiaries of JUNTOS would have remained invariant over time had the program not been implemented. This assumption is non-testable. However, in Section 6 we show suggestive evidence for Assumption 2.

Assumption 3-FO. (full observability)
(i) The joint distribution function of (Y_0, D) is observed. (ii) The joint distribution function of (Y_1, D) is observed.

Under Assumption 3-FO individuals can be classified at each time period into treated or control. This assumption is needed in order to be able to identify each of the four conditional means in expression 3.

Theorem 1. *Let Assumptions 1, 2, and 3-FO hold. The DID estimand defined in Equation 3 identifies the ATT defined in Equation 2.*

Proof. Rewriting expression 3, plugging in definition 1, and by Assumptions 1, 2, and A3-FO, gives

$$\begin{aligned}
\theta &= E\left(Y_1|D=1\right) - E\left(Y_0|D=1\right) - \left[E\left(Y_1|D=0\right) - E\left(Y_0|D=0\right)\right] \\
&= E\left(Y_1(1)|D=1\right) - E\left(Y_0(0)|D=1\right) - \left[E\left(Y_1(0)|D=0\right) - E\left(Y_0(0)|D=0\right)\right] \\
&= E\left(Y_1(1)|D=1\right) - E\left(Y_0(0)|D=1\right) - \left[E\left(Y_1(0)|D=1\right) - E\left(Y_0(0)|D=1\right)\right] \\
&= E\left(Y_1(1)|D=1\right) - E\left(Y_1(0)|D=1\right) \\
&= \text{ATT}.
\end{aligned} \tag{4}$$

□

3.2 | DID with missing treatment status

The setting in this paper is such that assumption 3-FO(i) does not hold. In our setup we observe repeated cross-sections drawn from the same population but at time $t = 0$ we observe only Y_0, whereas at time $t = 1$ we observe (Y_1, D). Because of this data limitation, the second term in square brackets in expression 3 is not identified without further assumptions.

To solve this identification problem we require auxiliary information in the form of a set of random variables observed in both time periods and correlated with the treatment status. This set may contain one or multiple variables with discrete or continuous support. We call this set "proxy variables" and we denote the set at time t as Z_t.

Assumption 3-LO. (limited observability)
The joint distribution functions of (Y_0, Z_0) and of (Y_1, D, Z_1) are observed.

This assumption indicates that treatment status is observed only in post-treatment periods, but one or more individual characteristics correlated with treatment status are always observed. The ENAHO dataset we use has such structure. The questionnaires in all repeated cross-section surveys are almost identical, with the same variables available in all years except for program

participation, which is only included after JUNTOS began operating. The set of proxy variables we use, Z_0 and Z_1, is formed of observed housing characteristics, such as roof, floor, and wall material, and the level of education of adults in the household.

Let the support of Z_t be denoted by $\mathcal{Z}_t$, $t = 0, 1$.

Assumption 4. (stationarity)

(a) $\mathcal{Z}_0 = \mathcal{Z}_1 \equiv \mathcal{Z}$;
(b) for all $z \in \mathcal{Z}$:

$$P(D = 1|Z_0 = z) = P(D = 1|Z_1 = z) \equiv e(z).$$

The first part of this assumption requires that the support of the proxies be the same across time periods. The second part of Assumption 4 requires that the proxies have the same effect on the propensity score over time.

Assumption 4 allows us to use information from time period $t = 1$ to recover the (unobserved) propensity score at time period $t = 0$.[6] This assumption holds automatically if the variables in Z_t are time invariant since the joint distribution of (D, Z) is necessarily time invariant. When Z_t includes time-varying variables, Assumption 4 is restrictive. It imposes that individuals who change Z_t over time do so in a specific way such that the proportion of treated individuals in each $z \in \mathcal{Z}$ remains constant.

The set of proxies we use to evaluate JUNTOS are plausibly time invariant. Yet we compute statistical tests that show suggestive evidence for the stationarity of the propensity score conditional on the set of proxies (see Section 6).

Assumption 5. (relevance)
Z_t is informative about D, that is, for $z_1 \neq z_2 \in \mathcal{Z}$:

$$e(z_1) \neq e(z_2).$$

Assumption 5 requires the variables in Z_t and the treatment status D to be correlated. This assumption is testable by using information from the post-program period when D and Z_1 are jointly observed. Table 2 in the Appendix shows that the proxies we use in our application satisfy Assumption 5.

Assumption 6. (exclusion restriction in changes)
Z_t is uninformative about the change over time in average realized outcomes conditional on D, that is:

$$E(Y_1|D, Z_1 = z) - E(Y_0|D, Z_0 = z) = E(Y_1|D) - E(Y_0|D).$$

Assumption 6 is weaker than the assumption usually imposed on proxies, since it requires that the *change*, rather than the level, in the conditional expectation of outcomes be independent of the proxies (see Wooldridge, 2001, p. 63).[7] In our setup, Z_t can have a direct effect on outcomes as long as this effect is time homogeneous.

Within the context of our application, Assumption 6 imposes the following. Since JUNTOS is targeted to the poor (i.e., eligibility is based on a poverty measure), we use housing characteristics, such as roof and wall materials, as proxies for eligibility. These proxies must not predict the evolution over time of the outcome of interest given the true treatment status. For example, among the beneficiaries of the program, the demand for vaccines should have evolved in the same way for those living in houses with brick walls as for those living in houses with wooden walls. However, Assumption 6 allows families living in houses with wooden walls to have a permanently lower demand for vaccines. We show in Section 6 suggestive evidence for this assumption.

We note here that, although it is not possible to test Assumptions 4(b) and 6 individually, it is possible to test them jointly. We present this test in Section 5.

Theorem 2 below shows the identification of the ATT when treatment status is observed in only one period and treatment effects may be heterogeneous.

Theorem 2. *Let Z_0, Z_1 be random variables with equal supports, $\mathcal{Z}_0 = \mathcal{Z}_1$, of cardinality $K \geqslant 2$. Define the $K \times 2$ matrix P and the $K \times 1$ vector Δ as, respectively,*

$$P \equiv \begin{bmatrix} 1 - e(z_1) & e(z_1) \\ \cdots & \cdots \\ 1 - e(z_K) & e(z_K) \end{bmatrix}, \quad \Delta \equiv \begin{bmatrix} E(Y_1|Z_1 = z_1) - E(Y_0|Z_0 = z_1) \\ \cdots \\ E(Y_1|Z_1 = z_K) - E(Y_0|Z_0 = z_K) \end{bmatrix}. \tag{5}$$

[6] Closely related assumptions are used in the literature on data combination with auxiliary information (see, e.g., Chen, Hong, & Nekipelov, 2011; Chen, Hong, & Tamer, 2005).
[7] For a similar assumption within the context of measurement error with cross-sectional data, see, for example, Lewbel (2007), Hu (2008), and Li (2002).

Let Assumptions 1, 2, 3-LO–5, and 6 hold. Then the ATT is identified, with the two differences of conditional means entering Equation 3 given by

$$\begin{bmatrix} E(Y_1|D=0) - E(Y_0|D=0) \\ E(Y_1|D=1) - E(Y_0|D=1) \end{bmatrix} = \left(P'P\right)^{-1} P'\Delta. \tag{6}$$

Proof. Applying the law of total probability at $t = 0, 1$ gives

$$E(Y_t|Z_t = z) = E(Y_t|D=1, Z_t = z)\, P(D=1|Z_t = z) + E(Y_t|D=0, Z_t = z)\, P(D=0|Z_t = z).$$

By Assumption 5:

$$P(D=1|Z_1 = z) = P(D=1|Z_0 = z) = e(z).$$

Combining the two results above and using Assumption 6 gives

$$\begin{aligned} &E(Y_1|Z_1 = z) - E(Y_0|Z_0 = z) \\ &= [E(Y_1|Z_1 = z, D=1) - E(Y_0|Z_0 = z, D=1)]\, e(z) \\ &\quad + [E(Y_1|Z_1 = z, D=0) - E(Y_0|Z_0 = z, D=0)]\, [1 - e(z)] \\ &= [E(Y_1|D=1) - E(Y_0|D=1)]\, e(z) + [E(Y_1|D=0) - E(Y_0|D=0)]\, [1 - e(z)]. \end{aligned}$$

Evaluating the expression above at $\{z_k\}_{k=1}^K$ gives the following system of equations:

$$\begin{bmatrix} E(Y_1|Z_1 = z_1) - E(Y_0|Z_0 = z_1) \\ \cdots \\ E(Y_1|Z_1 = z_K) - E(Y_0|Z_0 = z_K) \end{bmatrix} = \begin{bmatrix} 1 - e(z_1) & e(z_1) \\ \cdots & \cdots \\ 1 - e(z_K) & e(z_K) \end{bmatrix} \begin{bmatrix} E(Y_1|D=0) - E(Y_0|D=0) \\ E(Y_1|D=1) - E(Y_0|D=1) \end{bmatrix}. \tag{7}$$

Pre-multiplying Equation 7 by P' gives

$$P'\Delta = P'P \begin{bmatrix} E(Y_1|D=0) - E(Y_0|D=0) \\ E(Y_1|D=1) - E(Y_0|D=1) \end{bmatrix}.$$

Since $P'P$ is invertible by Assumption 5, we obtain the following solution:

$$\begin{bmatrix} E(Y_1|D=0) - E(Y_0|D=0) \\ E(Y_1|D=1) - E(Y_0|D=1) \end{bmatrix} = \left(P'P\right)^{-1} P'\Delta.$$

Assumptions 1 and 2 guarantee that the difference

$$E(Y_1|D=1) - E(Y_0|D=1) - [E(Y_1|D=0) - E(Y_0|D=0)]$$

identified above gives the ATT—see the proof of Theorem 1. Hence the result of Theorem 2 follows. □

Remark 1. (covariates)
In applied work, researchers usually condition on observed covariates, X. Our identification strategy carries through when conditioning on such covariates. In our framework, the main differences between proxies, Z_t, and covariates, X, are that (i) covariates need to satisfy a common support assumption, that is, $P(D=1|X) < 1$, for $t = 0, 1$ and for all X, and (ii) covariates do not need to satisfy an exclusion restriction such as Assumption 6. Formal assumptions needed for identification in the presence of observed covariates are included in the Supporting Information Appendix.

Remark 2. (comparison with Wald-DID)
When Z is a time-invariant, binary random variable, our identified expression for the DID estimand is the same as the Wald-DID estimand proposed by de Chaisemartin and D'Haultfoeuille (2015) when no individuals are treated in the pre-treatment period. This expression is given by

$$\frac{E(Y_1|Z=1) - E(Y_0|Z=1) - [E(Y_1|Z=0) - E(Y_0|Z=0)]}{P(D=1|Z=1) - P(D=1|Z=0)}. \tag{8}$$

However, in our setup, expression 8 represents the ATT, whereas in de Chaisemartin and D'Haultfoeuille (2015) it represents a LATE parameter, defined as

$$\text{LATE} = E(Y_1(1) - Y_1(0)|D=1, Z=1).$$

Moreover, the assumptions leading to the identification of Equation 8 are non-nested. For example, de Chaisemartin and D'Haultfoeuille (2015) show that Wald-DID identifies LATE under common trends conditional on Z, and homogeneous treatment effects across groups defined by Z, which restricts the heterogeneity of the treatment effects among treated individuals. On the other hand, our identification strategy relies on the common trends assumption conditional on D, which renders our method inapplicable when transitory shocks are correlated with the treatment status as in Ashenfelter's dip.

Since the assumptions mentioned above are non-nested and they are not testable with binary Z, the identification strategy chosen in practice will depend on both the parameter of interest, ATT versus LATE, and on which set of assumptions is deemed reasonable for the application at hand.

Remark 3. (comparison with ITT)
When the treatment status is missing in both periods or in the post-treatment period, it is common practice to compute the ITT using the eligibility status.[8] Call the eligibility random variable Z. The ITT is then defined as

$$\text{ITT} = E(Y_1|Z=1) - E(Y_0|Z=1) - [E(Y_1|Z=0) - E(Y_0|Z=0)]. \tag{9}$$

The ITT is a useful method, although its meaning is not clear in observational studies (see, e.g., Heckman, Hohmann, Smith, & Khoo, 2000). It is possible to show that

$$\text{ITT} = \text{ATT} \times [P(D=1|Z=1) - P(D=1|Z=0)]$$

under assumptions similar to those in our paper, that is, Assumptions 1, 2, 5, and the assumption below.

Assumption 6′. (exclusion restriction of eligibility)
The effect of treatment on the outcome is due to receiving the treatment, that is:

$$E(Y_t|D,Z) = E(Y_t|D), \quad t = 0, 1$$

Assumption 6′ is conventional in ITT analyses (see, e.g., Imbens & Rubin (2015, chapter 24), and it implies that, conditional on receiving the treatment, eligibility does not affect average outcomes. This assumption is slightly stronger than Assumption 6 since it imposes an exclusion restriction in *levels* rather than in changes over time.

Finally, note that when there is perfect compliance, $P(D=1|Z=1) = 1$, ITT = ATT. However, in the absence of perfect compliance, in order to be able to identify the ITT, one would have to identify $P(D=1|Z=1)$ and $P(D=1|Z=0)$. If the treatment status is observed in at least one post-treatment time period, then Assumption 4 obtains identification of the ITT. Else the ITT identifies only the sign of the ATT under the assumption that $P(D=1|Z=1) > P(D=1|Z=0)$.

4 | A SIMPLE LINEAR MODEL

Consider the following model for potential outcomes, which satisfies the assumptions for Theorem 2. This type of linear model is commonly used in applied research using DID. The notation is as in the previous section but we add the subscript i to emphasize that some parameters are individual specific. For the reasons discussed above, we consider the case where Z is time invariant.

[8] See, for example, Bleakley (2010); Cutler et al. (2010).

$$
\begin{aligned}
Y_{it}(0) &= \alpha_i + \delta_t + h(Z_i) + \varepsilon_{it}, \quad E(\varepsilon_{it}|t, Z_i) = 0, \\
Y_{it}(1) &= \alpha_i + t\beta_i + \delta_t + h(Z_i) + \eta_{it}, \quad E(\eta_{it}|t, Z_i) = 0, \\
\operatorname{cov}(Z_i, D_i) &= \sigma_{zD} \neq 0.
\end{aligned}
$$

When D_i is observed in all periods, this specification leads to the familiar model for the observed outcomes:

$$Y_{it} = \alpha_i + \delta_t + \beta_i t D_i + h(Z_i) + w_{it}, \quad E(w_{it}|t, D_i, Z_i) = 0$$

where $w_{it} = D_i\eta_{it} + (1 - D_i)\varepsilon_{it}$. The DID estimand is given by

$$\text{DID} = E(\beta_i|D_i = 1).$$

Both the parallel paths Assumption 2 and the no anticipation Assumption 1 hold. The stationarity Assumption 4 as well as the relevance Assumption 5 hold by construction. The covariance between Z and D is non-zero and constant over time, which holds automatically since both variables are time invariant. Finally, the exclusion restriction (Assumption 6) is satisfied since $h(.)$ is time invariant.

To see that time invariance of $h(.)$ is sufficient for Assumption 6, consider what happens when $h(.)$ varies over time. For any $d = 0, 1$ and any $z \in \mathcal{Z}$:

$$E(Y_{i1} - Y_{i0}|D_i = d) = \delta_1 - \delta_0 + E(h_1(Z_i) - h_0(Z_i)|D_i = d), \tag{10}$$

$$E(Y_{i1} - Y_{i0}|D_i = d, Z_i = z) = \delta_1 - \delta_0 + h_1(z) - h_0(z). \tag{11}$$

If $h_t(.) = h(.)$ for all t, expressions 10 and 11 are equal, and hence Assumption 6 is satisfied. Note that when $h(.)$ is time varying, one cannot disentangle the effect of the treatment from the effect of the proxy on the outcome.

5 | ESTIMATION

In this section, we consider inference for Equation 6. Since the DID estimand in applied work is often estimated using a linear parametric model, we propose a GMM estimator based on such a model. The proposed estimator leads to a GMM overidentification test for our main identification assumptions, 4 and 6. The large-sample properties of our estimator are shown in the Supporting Information Appendix.[9]

5.1 | A GMM estimator

Let $(Y_{it}, Z_{it}, D_i)_{i=1}^n$, $t = 0, 1$, be a random sample. Let $Z_{it} \in \mathbb{R}^K$ be a row vector of proxy variables for individual i. As in the literature on data combination,[10] define an indicator variable for sample membership:

$$T_i = \begin{cases} 1 \text{ if unit } i \text{ is observed at } t = 1 \\ 0 \text{ if unit } i \text{ is observed at } t = 0. \end{cases} \tag{12}$$

Assumption 7.

(i) For each i, T_i is independent of $Y_{1i}(D_i)$, $Y_{0i}(D_i)$, D_i, Z_{i0}, Z_{i1}.

(ii) With probability $q \in (0, 1)$ we draw an individual i at random from $t = 0$ and record i's realizations of (Y_{i0}, Z_{i0}). With probability $1 - q$ we draw an individual i at random from $t = 1$ and record i's realizations of (Y_{i1}, D_i, Z_{i1}).

[9] In the Supporting Information Appendix we derive another GMM estimator that uses the insights of Angrist (1991) on grouped-data estimators. The derivation shows how expression 6 leads naturally to the GMM presented in this section. Additionally, the alternative estimator expands the applicability of our method to a data combination problem. That is, researchers would be able to compute the DID estimand by combining a dataset containing (Y_1, Z), another dataset containing (Y_0, Z), and a third dataset containing (D, Z). Such setting is possible when no round of the household survey collects information about program participation, but the researcher has access to the administrative data of the program organizers.

[10] See, for example, Devereux and Tripathi (2009), Graham (2011), Abrevaya and Donald (2011), and Muris (2013).

In our setup, the true treatment status D_i is observed for all individuals with $T_i = 1$ but it is not observed for any individuals with $T_i = 0$. In conformity with our identification strategy, we use this information as follows.

First, let the propensity score for all i with $T_i = 1$ be known up to $\gamma \in \Gamma \subset \mathbb{R}^{d_\gamma}$ and given by

$$E\left(D_i|Z_{i1} = z, T_i = 1\right) = e\left(z, \gamma\right). \tag{13}$$

Second, let the observed outcome of individual i at time t be specified as

$$Y_{it} = \alpha_i + \delta_t + \beta_i T_i D_i + Z_{it}\kappa + w_{it}, \quad E\left(w_{it}|t, D_i, Z_i\right) = 0, \tag{14}$$

where $\kappa \in \mathbb{R}^K$. Note that Equation 14 is the conventional linear parametric model used in DID analysis (see, e.g., Abadie, 2005).

Specifications 13 and 14 and Assumption 4, which states that $E\left(D_i|Z_{i1} = z, T_i = 0\right) = e\left(z, \gamma\right)$, define the following conditional expectations for the observed outcomes:

$$E\left(Y_{i1}|D_i, Z_{i1}\right) = \delta_1 + \psi_1 D_i + Z_{i1}\kappa, \quad \text{if } T_i = 1, \tag{15}$$

$$E\left(Y_{i0}\ Z_{i0}\right) = \delta_0 + \psi_0 \quad e\left(Z_{i0}, \gamma\right) + Z_{i0}\kappa, \quad \text{if } T_i = 0 \tag{16}$$

where $\psi_0 \equiv [E(\alpha_i|D_i = 1) - E(\alpha_i|D_i = 0)]$ and $\psi_1 \equiv E(\beta_i|D_i = 1) + [E(\alpha_i|D_i = 1) - E(\alpha_i|D_i = 0)]$. The conditioning set for the outcomes in which $T_i = 0$ does not include D_i since the treatment status is not observed in pre-program periods. However, the parameters κ and γ are separately identified even when the propensity score is linear; that is, $e\left(Z_{i0}, \gamma\right) = Z_{i0}\gamma$, since κ is identified from moment condition 15 and γ from 13. As explained in Section 4, κ should be time invariant for Assumption 6 to hold. Then, the DID estimand for this model is given by

$$E(\beta_i|D_i = 1) = \psi_1 - \psi_0.$$

Let $\lambda \equiv (\gamma, \psi_1, \delta_1, \kappa, \psi_0, \delta_0)' \in \mathcal{L}$ be a $d_\gamma + K + 4$ vector, and define

$$e_\gamma(Z_{i1}, \gamma) \equiv \frac{\partial}{\partial \gamma} e\left(Z_{i1}, \gamma\right). \tag{17}$$

The vector of moment conditions is given by

$$g_i(\lambda) \equiv \begin{bmatrix} T_i e_\gamma(Z_{i1}, \gamma)\frac{D_i - e(Z_{i1}, \gamma)}{e(Z_{i1}, \gamma)\left(1 - e(Z_{i1}, \gamma)\right)} \\ T_i \quad (1, Z_{i1}, D_i)'\left(Y_{i1} - \delta_1 - \psi_1 D_i - Z_{i1}\kappa\right) \\ (1 - T_i)\left(1, Z_{i0}\right)'\left(Y_{i0} - \delta_0 - \psi_0 e\left(Z_{i0}, \gamma\right) - Z_{i0}\kappa\right) \end{bmatrix} \equiv \begin{bmatrix} g_{1i}(\gamma) \\ g_{2i}\left(\delta_1, \psi_1, \kappa\right) \\ g_{3i}\left(\delta_0, \psi_0, \gamma, \kappa\right) \end{bmatrix}$$

and it satisfies

$$E\left[g_i\left(\lambda^*\right)\right] = 0, \tag{18}$$

where $\lambda^* \in \mathcal{L}$ is the vector of population parameters. The first moment condition in Equation 18 corresponds to the score function of a likelihood function (the estimation of the propensity score); the second moment condition corresponds to (15) (the linear projection of Y_1 onto the space generated by the treatment status and the proxies); and the third moment condition corresponds to Equation 16.

The optimal GMM estimator of λ^* is given by

$$\widehat{\lambda} = \arg\min_{\lambda \in \mathcal{L}} \left[\frac{1}{n}\sum_{i=1}^{n} g_i(\lambda)\right]' \hat{V}^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} g_i(\lambda)\right], \tag{19}$$

where $\hat{V}$ converges in probability to V, the optimal GMM weighting matrix. Under standard regularity assumptions, $\widehat{\lambda}$ is normally distributed with a variance–covariance matrix defined in the Supporting Information Appendix.

When two or more proxy variables are available, a standard overidentification test can be performed to detect misspecifications in $g_i(\lambda)$ including violations of Assumptions 4 and 6. The overidentification test statistic has a limiting distribution of $\chi^2(K-1)$.

6 | THE EVALUATION OF JUNTOS SOCIAL PROGRAM

In this section we employ our proposed estimator (Equation 19) to evaluate the impact of JUNTOS on the demand for health inputs in two subpopulations: women of reproductive age and children under 5 years old.

6.1 | Estimation with multiple periods and additional tests

The method developed in Sections 3 and 5 considers the two-period case. As explained in Section 2, we have six repeated cross-sections to evaluate the impact of JUNTOS. We modify the moment conditions in Equation 18 in a simple way to allow for multiple periods. We let the coefficients δ_0 and ψ_0 in pre-treatment periods as well as coefficients δ_1 and ψ_1 in post-treatment periods be year specific. Then, each pair of phis can be used to estimate a DID. For instance:

$$\widehat{\psi}_{2008} - \widehat{\psi}_{2005} = \text{DID}_{2005-2008}$$

is used to compute the DID for 2005 and 2008. Having access to multiple pre- and post-treatment periods allows to develop three additional tests related to our identification assumptions. These tests are meant to provide suggestive evidence for our assumptions. We emphasize that, except for Assumptions 4(b) and 6, which we test jointly via the GMM overidentification test in Section 5, the remaining assumptions are individually non-testable.

Pre-treatment trend test (Assumption 2). When a standard DID is computed and data for multiple periods are available, it is common practice to analyze pre-treatment trends to explore the validity of the parallel trends assumption. Since both 2004 and 2005 are pre-treatment periods, our method allows for the possibility of testing the null hypothesis $H_0 : \text{DID}_{2004-2005} = 0$.

Post-treatment propensity score stationarity test (Assumption 4). The propensity score can be computed in each post-treatment period as $P(D = 1|Z) = e(Z, \widehat{\gamma}_t)$. For post-treatment periods, stationarity of the propensity score implies the testable null hypothesis $H_0 : \gamma_t = \gamma_{t+j}$ for all j for which D is observed.

This test follows the same logic as the pre-treatment test discussed above. The invariability of the propensity score between pre- and post-treatment periods (Assumption 4(b)) is untestable, yet one can analyze stationarity over two (or more) post-treatment periods where D is observed and use that as suggestive evidence for the stationarity of the propensity score. Nonetheless, the test we propose should be interpreted with caution. Note that when Z is time invariant the stationarity of the propensity score necessarily holds. In this case, a rejection of the null hypothesis may simply indicate that the composition of beneficiaries changed across post-treatment periods (e.g., the program may have disproportionately expanded over time in relation to some component of Z). Identification is not threatened since all the assumptions in Section 3.2 hold. However, the rejection of the null hypothesis indicates that post-treatment years should not be pooled in a single GMM estimator. Each post-treatment period $s \in \{2008, 2009, 2010, 2011\}$ should be used separately to compute DID_{s-2005}.

Post-treatment exclusion in changes test (Assumption 6). Similarly to the previous test, the conditional expectation of the outcome conditional on both program participation and proxy variables can be computed for all post-treatment periods. In the linear setting considered previously, this means that $E(Y_{it}|D_i, Z_i) = \delta_t + \psi_t D_i + Z_i\kappa_t$, where t denotes post-treatment period. The exclusion restriction of Z from the conditional expectation of the changes in Y (Assumption 6) implies the testable null hypothesis $H_0 : \kappa_t = \kappa_{t+j}$ for all j for which D is observed.

The caveats of the previously discussed stationarity test also apply here. If the composition of beneficiaries changes across different post-treatment periods, then the conditioning set in $E(Y_{it}|D_i, Z_i)$ also changes. In such a case, the null hypothesis may be rejected despite Assumption 6 being satisfied.

The test presented here becomes more meaningful from an empirical point of view when it is analyzed jointly with the post-treatment stationarity test. If the hypothesis of stationarity in the propensity score is not rejected, then there was no significant change in the composition of beneficiaries. In this case, a rejection of the null hypothesis $H_0 : \kappa_t = \kappa_{t+j}$ would indicate a violation of Assumption 6.

6.2 | Results

Figure 2 shows the impact of JUNTOS on the probability of seeking medical care when sick among women of reproductive age. The figure displays the DID point estimate and its 90% confidence interval for each survey year. In all cases, the comparison is in relation to the year 2005. The DID is zero by construction in 2005.

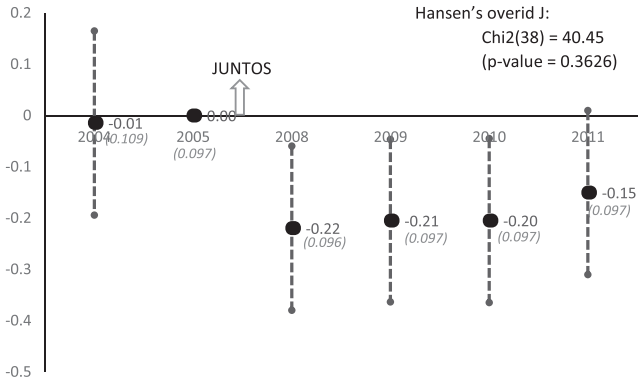

FIGURE 2 Did not seek medical care when sick (women 15–45 years old). Standard errors clustered at the district level are in parentheses. Propensity score F-statistic = 216.45, corr$(D, \hat{D}) = 0.38$. Pooled two-period DID = −0.169 (SE 0.068). Post-treatment propensity score stationarity test $F(24, 1329) = 1.34$ (p-value = 0.1278), excluding year 2010 $F(16, 995) = 1.14$ (p-value = 0.3121). Post-treatment excludability in changes test $F(24, 1329) = 1.43$ (p-value = 0.0837), excluding year 2010 $F(16, 995) = 1.11$ (p-value = 0.3359)

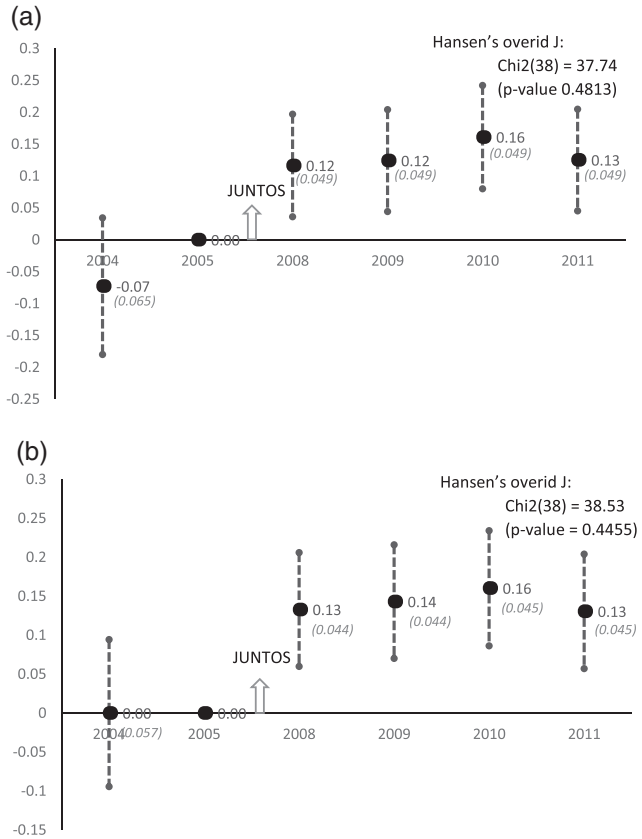

FIGURE 3 Family planning (women 15–45 years old). (a) Family planning consultation. Standard errors clustered at the district level are in parentheses. Propensity score F-statistic = 305.00, corr$(D, \hat{D}) = 0.36$. Pooled two-period DID = 0.151 (SE 0.039). Post-treatment propensity score stationarity test $F(24, 1353) = 1.32$ (p-value = 0.1375). Post-treatment excludability in changes test $F(24, 1353) = 0.78$ (p-value = 0.7640). (b) Contraceptive use. Standard errors clustered at the district level are in parentheses. Propensity score F-statistic = 304.86, corr$(D, \hat{D}) = 0.36$. Pooled two-period DID = 0.138 (SE 0.035). Post-treatment propensity score stationarity test $F(24, 1353) = 1.32$ (p-value = 0.1372). Post-treatment excludability in changes test $F(24, 1353) = 1.16$ (p-value = 0.2665)

The results in Figure 2 indicate that JUNTOS increased the probability that women seek medical care when sick by 22 percentage points in 2008. The beneficial effects of the program remain high, almost at the same level, in subsequent years.

There is suggestive evidence of the parallel trend assumption between beneficiaries and non-beneficiaries (Assumption 2) since the estimated DID between 2004 and 2005 is not significantly different from zero.

Another important piece of information in Figure 2 is the overidentification test. Because we have multiple proxy variables, we can compute this omnibus test that rejects the null hypothesis of correct model specification if any of the assumptions in Section 3 is violated. We see that the test cannot reject the null hypothesis at conventional levels.

The caption to Figure 2 shows two measures of how well the proxy variables predict the treatment status obtained by estimating the propensity score. The first measure is the correlation between the observed treatment status and the predicted treatment status. The drawback of this measure is that it does not incorporate the gain in precision when the sample size increases. Hence we also report the F-statistic, paralleling the IV literature.

Most importantly, the Figure 2 caption shows the results of the post-treatment propensity score stationarity test and the post-treatment excludability in changes test. These two tests are computed using auxiliary regressions (i.e., outside the GMM system) where the propensity score and the conditional expectation of the outcomes are estimated allowing the coefficients of the proxies Z to be year specific. Then, the tests are implemented by computing a standard F-test for the null hypothesis that the coefficients associated with Z are jointly the same across years.

Results of the tests show that the hypothesis of propensity score stationarity cannot be rejected at conventional levels. However, the hypothesis of excludability in changes is rejected at the 10% level. Further investigation indicates that this result is driven by one proxy variable (wall material) in year 2010. When this year is excluded the null hypothesis fails to be rejected.

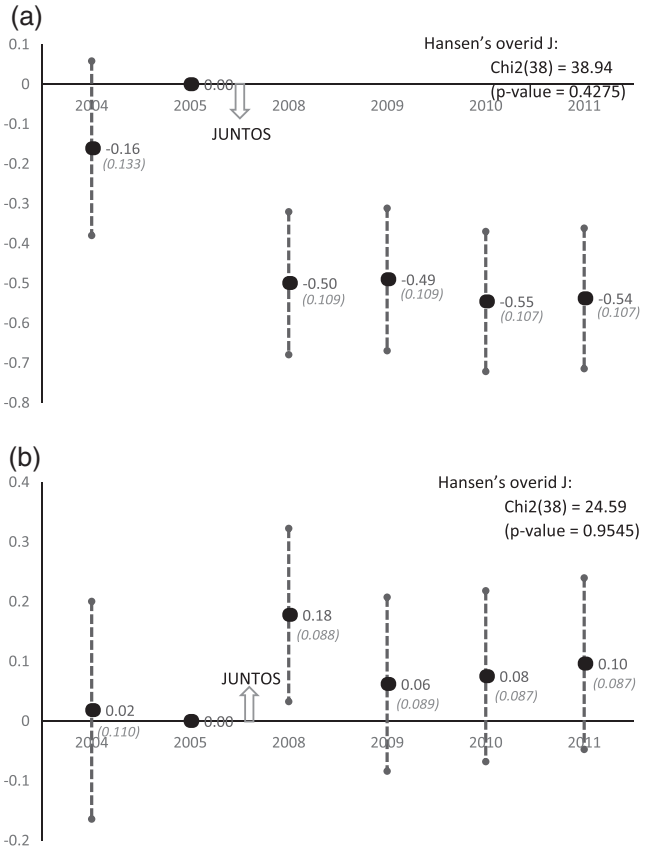

FIGURE 4 Health inputs (children 0–5 years old). (a) Did not seek medical care when sick. Standard errors clustered at the district level are in parentheses. Propensity score F-statistic = 93.37, corr$(D, \hat{D}) = 0.35$. Pooled two-period DID = −0.438 (SE 0.092). Post-treatment propensity score stationarity test $F(24, 1179) = 1.35$ (p-value = 0.1229), excluding years 2010–2011 $F(8, 576) = 0.91$ (p-value = 0.5080). Post-treatment excludability in changes test $F(24, 1179) = 1.75$ (p-value = 0.0142), excluding years 2010–2011 $F(8, 576) = 0.98$ (p-value = 0.4529). (b) Vaccines. Standard errors clustered at the district level are in parentheses. Propensity score F-statistic = 153.83, corr$(D, \hat{D}) = 0.33$. Pooled two-period DID = 0.096 (SE 0.072). Post-treatment propensity score stationarity test $F(24, 1316) = 1.67$ (p-value = 0.0222), excluding year 2011 $F(16, 977) = 0.69$ (p-value = 0.8035). Post-treatment excludability in changes test $F(24, 1316) = 0.71$ (p-value = 0.8438), excluding year 2011 $F(16, 977) = 0.88$ (p-value = 0.5936)

Even a small deviation from Assumption 6 can make the post-treatment excludability in changes test reject the null hypothesis when the power of the test is sufficiently high. This does not mean that such deviation affects the results significantly. We re-estimate the model allowing the coefficients associated with Z in 2010 (κ in system 18) to be different from their counterparts in other years. The point estimates (shown in the Supporting Information Appendix) changed only 1 percentage point, from -20% to -19% in 2010, suggesting that the deviation from Assumption 6, although statistically significant, is negligible in the final results.

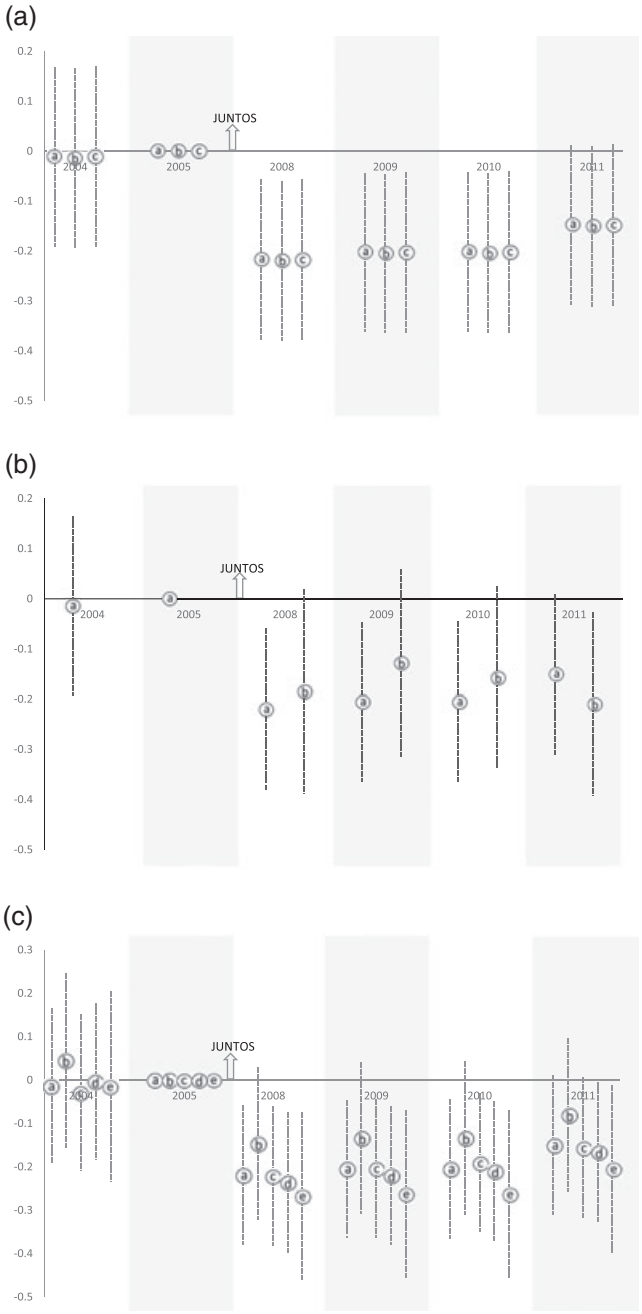

FIGURE 5 Did not seek medical care when sick (women 15–45 years old). (a) Functional forms for the propensity score. Standard errors clustered at the district level are in parentheses: ⓐ Propensity score linear probability model; ⓑ Propensity score probit; ⓒ propensity score logit. (b) Joint estimation versus separate regressions. Series ⓐ: values for all years jointly estimated in a single GMM with a common propensity score. Series ⓑ: each DID estimated separately ignoring other years. Overidentification test fails to reject the null hypothesis of correct moment specification in all cases. (c) Alternative set of proxy variables. Standard errors clustered at the district level in parentheses: ⓐ no proxy variable omitted; ⓑ roof material omitted in the proxy variables set, overidentification test p-value = 0.46; ⓒ floor material omitted in the proxy variables set, overidentification test p-value = 0.38; ⓓ wall material omitted in the proxy variable set, overidentification test p-value = 0.42; ⓔ max. education of adults omitted in the proxy variables set, overidentification test p-value = 0.37

Figure 3(a) is analogous to Figure 2, although the variable studied in this case is whether women seek family planning consultation. This outcome is particularly relevant when CCTs are studied. Because money transfers are made conditional on having children living in the household, there is a concern that CCTs increase the demand for children above the optimal level. In such circumstances, this type of social program may induce households to remain poor due to excess fertility, generating the opposite effect to that which was intended. On the other hand, fertility may decline as a consequence of the program in environments where the rate of unwanted births is high. By increasing the frequency of medical care, women may be more likely to obtain information about family planning and acquire contraceptives. This seems to be the case in Peru.

The results in Figure 3(a) suggest that JUNTOS induced women to increase by 12–16 percentage points the frequency of family planning consultation. Figure 3(b) shows results in the same direction. The use of contraceptives increased by a similar magnitude.

Figure 4 shows children's outcomes. One of the main objectives of CCTs is to end the transmission of poverty across generations. To achieve this goal, these programs create incentives for parents to invest in children's human capital. The rationale is consistent with the literature on human development, which indicates that early life health significantly influences a variety of adult outcomes (Almond & Currie, 2010). We investigate the demand for health inputs among children younger than 5 years.

Figure 4(a) indicates that JUNTOS increased the demand for medical care when the child became sick by about 50 percentage points. The null hypothesis of the post-treatment excludability test is rejected at 5% unless years 2010 and 2011 are excluded. For this reason, results for the last 2 years should be interpreted with caution. Nonetheless, robustness analysis as that performed for Figure 2 show similar results (Supporting Information Appendix).

Figure 4(b) shows the impact of JUNTOS on children's vaccination. These results suggest that the program had a strong effect in the initial years (18 percentage points) and later declined. This pattern is expected. Once the program was implemented, it created the incentive to vaccinate newborn children on a regular schedule as well as older children who where behind schedule. After a few years of operation, it is likely that children who where behind regularized their condition. Then, the impact of the program reduced since parents started to follow the regular vaccination schedule.

Perova and Vakis (2012) previously evaluated the impact of JUNTOS using both an instrumental variable approach and matching techniques. Overall, we obtain much more precise estimates. Among the outcomes we analyze, they can only reject the zero effect hypothesis at the 10% level for contraceptive use and children seeking medical care. Nonetheless, their point estimates for these two outcomes do not differ much from ours (12 vs. 13 percentage points for contraceptive use and 55 vs. 50 percentage points for children seeking medical care).

6.3 | Robustness

Figure 5(a) shows how results in relation to women seeking medical care change when the functional form for the propensity score alternates between a linear probability model, a probit model, and a logit model. The figure shows that results are strongly robust to alternative functional forms.

The DID results for all years in Figure 2 are jointly estimated in a single GMM. This computation imposes that (i) the set of coefficients estimated for the propensity score be time invariant (Assumption 4) and (ii) the impact of proxy variables on outcomes be constant over time as well (Assumption 6). The post-treatment tests as well as the overidentification test indicate that these two assumptions hold. Nonetheless, in Figure 5(b) we also show the results of estimating the DID for each year separately. Thus the propensity score and the impact of proxy variables on outcomes are specific to each year.

Figure 5(c) shows results when alternative sets of proxy variables are used. Series ⓐ correspond to Figure 2, included for comparison reasons. All sets of proxies give similar results.

7 | CONCLUSION

In this paper, we introduce a new method which allows practitioners to use the conventional DID methodology when data have a repeated cross-sectional structure and the individual treatment status is observed in only one period. We show that the ATT is identified when a proxy for the latent treatment status exists and is observed in both time periods. Our main identifying assumptions are: (i) the stationarity of the propensity score conditional on the proxy; and (ii) the non-correlation between the proxy and the changes over time in average outcomes conditional on the true treatment status. We propose a simple GMM estimator for the ATT. We use our method to estimate the impact of JUNTOS, a conditional cash transfer implemented in Peru, on the demand for health input. We find that the program had an important positive effect on the probability of seeking medical

care when needed among both children and women of reproductive age. Our results indicate that the probability that women use contraceptives increased by approximately 14 percentage points.

ACKNOWLEDGEMENTS

We are grateful to the Editor Thierry Magnac and three anonymous referees for their comments and suggestions, which have greatly improved this paper. We would like to thank Sylvain Chabé-Ferret, Yanquin Fan, Jean-Pierre Florens, Xavier D'Haultfoeuille, Yuichi Kitamura, Pascal Lavergne, Chris Muris, and Shu Shen for helpful discussions, and seminar audiences at TSE, Bristol Econometric Study Group 2013, AMES Singapore, Wisconsin–Madison, UW Seattle, UBC, SFU, UVic, Vanderbilt, Yale, LAMES Mexico City, Latin American Workshop of Econometrics Montevideo, CEA Toronto, the 5th French Econometrics Conference (Toulouse), and the 11th World Congress of the Econometrics Society (Montreal) for comments and suggestions.

REFERENCES

Abadie, A. (2005). Semiparametric difference-in-difference estimators. *Review of Economic Studies*, *72*, 1–19.

Abrevaya, J., & Donald, S. G. (2011). A GMM approach for dealing with missing data on regressors and instruments. (Working paper), Austin, TX.

Almond, D., & Currie, J. (2010). Human capital development before age five. In Card, D., & Ashenfelter, O. (Eds.), *Handbook of Labor Economics*, Vol. *4B*. Amsterdam, Netherlands: Elsevier, pp. 1315–1486.

Angrist, J. D. (1991). Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics*, *47*(2–3), 243–266.

Bleakley, H. (2010). Malaria eradication in the Americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, *2*(2), 1–45.

Buchmueller, T. C., DiNardo, J., & Valletta, R. G. (2011). The effect of an employer health insurance mandate on health insurance coverage and the demand for labor: Evidence from Hawaii. *American Economic Journal: Economic Policy*, *3*, 25–51.

Chen, X., Hong, H., & Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, *49*(4), 901–937.

Chen, X., Hong, H., & Tamer, E. (2005). Measurement error models with auxiliary data. *Review of Economic Studies*, *72*, 343–366.

Chen, X., Hu, Y., & Lewbel, A. (2009). Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information. *Statistica Sinica*, *19*, 949–968.

Cross, P. J., & Manski, C. F. (2002). Regressions, short and long. *Econometrica*, *70*(1), 357–368.

Cutler, D., Fung, W., Kremer, M., Singhal, M., & Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India. *American Economic Journal: Applied Economics*, *2*(April), 72–94.

de Chaisemartin, C., & D'Haultfoeuille, X. (2015). *Fuzzy differences-in-differences (Warwick Economics Research Paper Series)*. Warwick, UK: University of Warwick.

Devereux, P. & Tripathi, G. (2009). Optimally combining censored and uncensored datasets. *Journal of Econometrics*, *151*, 17–32.

Graham, B. S. (2011). Efficiency Bounds for missing data models with semiparametric restrictions. *Econometrica*, *79*(2), 437–452.

Groen, J. A., & Polivka, E. (2008). The effect of Hurricane Katrina on the labor market outcomes of evacuees. *American Economic Review*, *98*(2), 43–48.

Heckman, J., Ichimura, J., Smith, J., & Todd, P. (1996). Sources of selection bias in evaluating social programs: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. *Proceedings of the National Academy of Sciences USA*, *93*(23), 13416–13420.

Heckman, J., Ichimura, J., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*(5), 1017–1098.

Heckman, J., Ichimura, J., & Todd, P. (1997). Matching as an econometric evaluation estimator. *Review of Economic Studies*, *64*(4), 605–654.

Heckman, J., Hohmann, N., Smith, J., & Khoo, M. (2000). Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics*, *115*(2), 651–694.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, *144*, 27–61.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, *76*, 604–620.

Levy, D., & Ohls, J. (2003). *Evaluation of Jamaica's PATH program: Methodology report*. Washington, DC: Mathematica Policy Research.

Lewbel, A. (2007). Estimation of Average treatment effects with misclassification. *Econometrica*, *75*, 537–551.

Li, T. (2002). Robust and consistent estimation of nonlinear errors-in-variables models. *Journal of Econometrics*, *110*, 1–26.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica*, *74*, 631–665.

Molinari, F. (2010). Missing treatments. *Journal of Business and Economic Statistics*, *28*(2), 82–95.

Muris, C. (2013). Efficient GMM estimation with general missing data patterns. (Working paper), Burnaby, Canada.

Perova, E., & Vakis, R. (2012). 5 years in Juntos: New evidence on the program's short and long-term impacts. *Revista Economía*, *35*(69), 53–82.

Smith, J., & Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators. *Journal of Econometrics*, *125*, 305–353.

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Botosaru I, Gutierrez FH. Difference-in-differences when the treatment status is observed in only one period. *J Appl Econ*. 2017;1–18 https://doi.org/10.1002/jae.2583