Evaluating Nonexperimental Estimators for Multiple Treatments: Evidence from a Randomized Experiment*

Carlos A. Flores†   Oscar A. Mitnik‡

December 22, 2008

Preliminary and Incomplete - Comments Welcome

Abstract

This paper assesses nonexperimental estimators of mean effects of multiple or multivalued treatments by analyzing their effectiveness in adjusting for observable characteristics and eliminating differences in average outcomes among multiple populations. The data we use comes from the National Evaluation of Welfare-to-Work Strategies (NEWWS), a social experiment conducted in the U.S. in the 1990s in which individuals in seven locations were randomly assigned to a control group or to different training programs emphasizing either human capital development or labor force attachment. The prior literature evaluating the performance of nonexperimental methods has focused exclusively on binary treatments. Given the growing interest in evaluating programs in which the treatment is multivalued or there are more than one treatment, it is important to learn about the performance of different estimators in this context. Among the estimators studied, we pay particularly attention to those based on the generalized propensity score or GPS, which equals the probability of receiving a particular treatment (or level of the treatment) conditional on covariates. In addition, we analyze the role of the GPS in identifying units across treatment groups that are comparable in terms of observable characteristics, and provide guidance for its use in practice.

*We thank Chris Bollinger, Guido Imbens, Phil Robins and seminar participants at the University of Kentucky, at the Second Joint IZA/IFAU Conference on Labor Market Policy Evaluation, and at the University of Miami Labor Discussion Group for useful comments and suggestions. Bryan Mueller and Yongmin Zang provided excellent research assistance.

†Department of Economics, University of Miami. Email: caflores@miami.edu.

‡Department of Economics, University of Miami and IZA. Email: omitnik@miami.edu.

1 Introduction

Nonexperimental methods are widely used in economics and other disciplines to evaluate government programs and many types of interventions. In the absence of an experiment (or infeasibility of conducting one), nonexperimental methods are in many situations the only alternative. Among them, those based on selection on observables or unconfoundedness assumptions play an important role (e.g., Imbens, 2004, 2008; Heckman et al., 1999). Most of the focus on nonexperimental methods in the previous two decades has been on estimation of average treatment effects of a binary treatment or intervention on an outcome. In practice, however, individuals are usually exposed to different doses of the treatment or to more than one treatment. As a result, some of the focus has recently shifted to developing methods to evaluate such programs. This paper contributes to this literature by assessing the effectiveness of nonexperimental estimators of mean effects for multiple or multivalued treatments in adjusting for observable characteristics and eliminating differences in average outcomes among multiple populations. The data we use comes from the National Evaluation of Welfare-to-Work Strategies (NEWWS), a social experiment conducted in the U.S. in the 1990s in which individuals in seven locations were randomly assigned to a control group or to different training programs emphasizing either human capital development or labor force attachment.

Since the influential paper by Lalonde (1986) many studies have evaluated the performance of different nonexperimental methods (e.g., Heckman and Hotz, 1989; Friedlander and Robins, 1995; Heckman et al., 1997, 1998; Dehejia and Wahba, 1999, 2002; Michalopoulos et al., 2004; Smith and Todd, 2005; Dehejia, 2005; Mueser et al., 2007). This literature has advanced our understanding of nonexperimental evaluations by specifying conditions under which nonexperimental estimators are more likely to replicate the outcome from a randomized experiment. One of the main conclusions is the importance of comparing "comparable" individuals. For instance, Heckman et al. (1997, 1998ab) stress the importance of comparing treatment and control groups from the same local labor market to which the same questionnaire is administrated, as well as having data on detailed labor market histories. This literature has also highlighted the importance of the propensity score (i.e., the probability of receiving treatment conditional on covariates) to identify regions of the data where treatment and control units are comparable in terms of observed characteristics.

A common characteristic of the current literature evaluating nonexperimental estimators based on a selection-on-observables assumption is its focus on binary treatments: individuals either participate in a program or not. Recently, however, there has been a growing interest in evaluating programs or interventions in which the treatment is multivalued or there are more than one treatment (e.g.,Behrman et al., 2004; Frölich et al., 2004; Flores-Lagunes et al., 2007; Kluve et al., 2007; Plesca and Smith, 2007; Mitnik, 2008), and on different methods to evaluate such programs (e.g., Imbens, 2000; Lechner, 2001; Hirano and Imbens, 2004; Cattaneo, 2007; Flores, 2007). Unfortunately, very little is known about the performance of alternative estimation techniques in terms of reducing the potential selection bias present in nonexperimental evaluations of multiple treatments. To our knowledge, ours is the first study to address this issue.

When the treatment is multivalued or there are more than one treatment we have more parameters of interest than the commonly used average treatment effect (or average treatment effect on the treated) in the binary-treatment case. For instance, one may be interested on pairwise comparisons (e.g., Lechner, 2001), or on finding the level of the treatment (or the particular treatment) that gives the highest average outcome (e.g., Flores, 2007). In this paper, we focus on estimators of what is some times called the "dose-response function" (although this may not be the most appropriate denomination for non-ordered multivalued treatments). It gives the average potential outcome over all possible values of the treatment. In other words, it gives the expected potential outcome at all possible values of the treatment for someone randomly chosen from the population. Since in a nonexperimental evaluation the population is selected into different treatment levels, a major task for estimation of the dose-response function is finding individuals that are comparable simultaneously across all treatment levels.

A general approach to evaluate the performance of nonexperimental estimators in the binary-treatment case consists on using data from a randomized experiment and constructing a nonexperimental control group, for instance, from additional data sets (e.g., Lalonde, 1986) or from different locations (e.g., Friedlander and Robins, 1995). The different nonexperimental estimators are then used on the nonexperimental control group and the experimental treated group and, to asses the performance of the estimators, the results are compared against those from the experiment. One could also apply the estimators to the nonexperimental and experimental control groups, in which case the benchmark is obtaining a zero treatment effect. A special application of this general approach is Heckman et al. (1997), in which their nonexperimental control group consisted of individuals that (i) were eligible to the program being evaluated (the National Job Training Partnership Act, JTPA) but that did not apply, (ii) resided in the same narrowly defined area as the applicants; and, (iii) were administered the same survey as those in the experiment. As stressed in their paper, having a nonexperimental control group in the same local labor market as those receiving treatment, administering the same questionnaire and having detailed labor market history seem to be key for nonexperimental methods to work properly.

Extending the logic from the binary-treatment literature, an ideal data for the purpose of evaluating nonexperimental methods of multiple treatments would consist of an experiment in which units are randomized into s different treatments, with $s > 2$. In addition, for each of $s-1$ treatments there would be units that self-select into these same treatments but that are otherwise representative of the population in which the experiment took place (e.g., welfare recipients in a given area and time). These units would form the nonexperimental groups. The data would have to contain detailed information on all units (e.g., background characteristics and previous labor market history), and the same data gathering instrument would have to be used for all units. In this case, we could take the nonexperimental groups plus one of the experimental groups, apply alternative nonexperimental methods to these data, and compare the results to those from the actual experiment. Unfortunately, such a data is not available to the best of our knowledge, and we resort to a different strategy.

In this paper, we resort to the availability of several control groups in different sites of the NEWWS experiment to evaluate alternative nonexperimental estimators of multiple treatments. We use alternative methods to adjust for observable characteristics in order to eliminate differences in average outcomes among members of the control groups in different sites. Relying on individuals from an experiment such as NEWWS has the advantages that i) all individuals regardless of their location are welfare recipients at the time of randomization, which helps to reduce the heterogeneity across sites[1]; and ii) the data and survey instruments gathered for all the individuals are the same. In the case of NEWWS the data available on each individual is extremely rich. By focusing on different geographic locations, however, we have the disadvantage of having to deal with the (potential) structural differences in local labor markets. Indeed, this is one of the issues that appear as very important in this application.

Our strategy of comparing different control groups is similar to that previously used in a binary-treatment context by Friedlander and Robins (1995), Michalopoulos, Bloom, and Hill (2004) and Hotz, Imbens and Mortimer (2005). The key difference in our approach, however, is that while their focus is on pairwise comparisons between controls in different locations, we focus on *simultaneously* comparing the control individuals across *all* locations. This allows us to move beyond binary-treatment methods and evaluate nonexperimental estimators for multiple treatments because we need to adjust for differences in observed characteristics of several groups at the same time.

Finally, among the estimators we evaluate we pay particular attention to those based

[1] As discussed in the following section, in one of the sites in NEWWS (Oklahoma City) randomization took place at the time of application to welfare. Because of this, we exclude this particular site from our analysis.

on the generalized propensity score or GPS (i.e., the probability of receiving a particular treatment conditional on covariates), such as weighting and partial-mean estimators. In addition, we systematically analyze the role of the GPS in identifying units across sites that are comparable in terms of observable characteristics, and provide guidance for its use in practice. We show the crucial role played by the GPS in extending to the multiple-treatment setting the "common support condition" frequently used in the binary-treatment setting.

The paper is organized as follows. Section 2 describes the data used; in section 3 we present the general setup, and in the following section we present the estimators to be used in the paper. The results are presented in Section 5, and Section 6 concludes.

2 Data

The data used in this paper comes from the National Evaluation of Welfare-to-Work Strategies (NEWWS), which is a multi-year study conducted in the early nineties to compare the effects of two alternative strategies for helping welfare recipients (mostly single mothers) to improve their labor market outcomes and leave public assistance. The first strategy emphasized labor force attachment (LFA) by encouraging participants to find employment quickly, and the second focused on human capital development (HCD) by offering academic, vocational and employment-oriented skills training. The programs evaluated in the NEWWS study were operated in seven sites across the U.S.: Atlanta, GA; Columbus, OH; Detroit, MI; Grand Rapids, MI; Oklahoma City; OK; Portland, OR; and Riverside, CA. In Atlanta, Grand Rapids and Riverside both LFA and HCD programs were offered, and individuals were randomly assigned to LFA, HCD or the control group.[2] In the rest of the sites, individuals were randomized to one of the programs (LFA, HCD or a combination of both) or to the control group, which was denied access to the training services offered by the program for a pre-set "embargo" period.

The year in which random assignment took place differs across sites, with the earliest randomization starting in the second quarter of 1991 in Riverside, and the latest in the fourth quarter of 1994 in Portland.[3] The NEWWS data set contains information on labor market

[2] One could use these sites to create alternative nonexperimental groups for those receiving LFA and HCD training. However, as discussed below, since LFA and HCD programs are heterogeneous across sites, this introduces additional biases. For this reason, we focus on comparing average outcomes for control individuals across sites, where everyone is excluded from receiving treatment. This also helps to increase the number of groups considered in our nonexperimental evaluation as the number of sites is greater than the number of alternative treatments.

[3] The dates in which randomization took place in all seven sites are (month/year): Atlanta (01/92-06/93), Columbus (09/92-07/94), Detroit (05/92-06/94), Grand Rapids (09/91-01/94), Oklahoma City (09/91-05/93), Portland (02/93-12/94) and Riverside (06/91-06/93).

outcomes up to 5 years after random assignment, information on individual background characteristics, as well as individual welfare and labor market history up to two years prior to random assignment. We use these characteristics, further described in Section 5, to apply the nonexperimental estimators in which we will focus our analysis.[4]

As we explain in detail in the following section, we employ nonexperimental methods to eliminate differences in control group outcomes across the different locations in the NEWWS experiment. The total number of individuals in the control groups in the seven sites is 17,521. From these, we exclude all men from our analysis (1,303), and also all females with missing values on any of the variables used in our analysis (805). From the remaining observations, we also drop those controls for which it is unknown whether they were embargoed from the program services during the period considered (404). Finally, we exclude two sites from our analysis (5,658), Columbus and Oklahoma City. Columbus has the problem of not having two years of labor market history prior to random assignment. Given the documented importance of controlling for such variables in nonexperimental settings (e.g., Heckman et al., 1997; Hotz et al., 2005) and the fact that it is the only site with that issue, we exclude it from our analysis. We drop Oklahoma City from the analysis because in this site randomization was done to welfare *applicants*, as opposed to welfare *recipients* as in the remaining sites. A big proportion (30%) of those individuals randomized in Oklahoma City did not actually qualify for welfare, and it is hard to believe they would be a reasonable comparison group for individuals that did qualify. In fact, there is evidence in the literature that applicants and recipients are very different in terms of their characteristics and outcomes (e.g., Friedlander, 1988). Hence, in order to have groups across sites that are all formed by welfare recipients at randomization, we drop Oklahoma City from the analysis. The final sample size in our analysis is 9,351 women, with 1,372 women from Atlanta; 2,037 from Detroit; 1,374 from Grand Rapids; 1,740 from Portland and 2,828 from Riverside.

The outcome we analyze in section 5 is an indicator variable equal to one if the individual was ever employed during the two years following randomization, and zero otherwise. We focus on an outcome measured in two years after random assignment because in some sites we cannot be sure that all control individuals were embargoed from receiving program services starting in year three.

3 General Framework

We base our general framework on the potential outcome approach developed by Neyman (1923) and extended by Rubin (1974) to non-experimental settings. Each unit i in our

[4]For further details on the NEWWS study see Hamilton et al. (2001).

sample, $i = 1, 2, \ldots, N$, comes from one of k possible sites. Let $D_i \in \{1, 2, \ldots, k\}$ be an indicator of the location of individual i. We denote the potential outcomes by $Y_i(t_d, d)$, where t_d stands for the treatment and d for the location. Hence, $Y_i(t_d, d)$ is the outcome unit i would obtain if she were located in site d and given treatment t_d. Two differences with respect to the commonly used potential outcomes in program evaluation (e.g., Imbens, 2004) are worth mentioning. First, we let the potential outcome $Y(t_d, d)$ depend on d for notational convenience. Although it may be difficult to think of the site as something we can manipulate (i.e., a "treatment" in Holland's (1986) sense), it is convenient for our purposes as our goal is to simultaneously use individuals from one site as a comparison group for another site. Second, we let t_d depend on d, as not all sites offered LFA and HCD training. For all sites, a value of t of zero denotes the control treatment, which prevents individuals from receiving any program services.

In this paper we focus exclusively on the control groups, so we use only the potential outcomes at zero, or $Y(0, d)$. The reason we focus only on controls is that not every site offered the two programs based on LFA and HCD, and that programs differed across sites in terms of implementation, particular services offered, administration, etc. By focusing on the control treatment we try to minimize treatment heterogeneity across sites, and it allows us to use more sites as they all have a control group.[5]

The data we observe for each unit is (Y_i, D_i, X_i), with X_i a set of pre-treatment variables, and $Y_i = Y(0, D_i)$. Our parameters of interest in this paper are

$$\beta_d = E[Y(0, d)], \text{ for } d = 1, 2, \ldots, k \tag{1}$$

The object in (1) gives the expected outcome under the control treatment in location d for someone randomly selected from our entire sample. In cases where d represents different levels of the treatment (and the zero is omitted from the potential outcome), (1) is the dose-response function.

Even though the treatment is randomly assigned within each site, and therefore $E[Y_i(0, d)|D_i = d]$ is identified from the data for every site, $E[Y_i(0, d)]$ is not identified without further assumptions. In general, it is not possible to use the controls from one location as a comparison group from another because the distribution of the characteristics in all k locations may differ. In order to evaluate nonexperimental methods that adjust for observable characteristics with multiple treatments, we impose the following unconfoundedness or selection-on-observables assumption.

[5] As in Hotz et al. (2005), if one is able to adjust for control group outcomes across sites, the comparison of adjusted outcomes for nominally equal treatments across sites (e.g., LFA programs in different locations) may be interpreted as the effect of program heterogeneity across sites.

Assumption 1 (Unconfounded site) The site an individual belongs to is unconfounded given pre-treatment variables X_i, or

$$D_i \perp \{Y_i(0,d)\}_{d\in\{1,2,\ldots,k\}} | X_i \tag{2}$$

This assumption states that, conditional on a set of covariates, the site an individual belongs to is independent of her potential outcomes. Assumption 1 is similar to that in Hotz et al. (2005) in the binary treatment case.

In addition, we impose an overlap assumption that guarantees that in infinite samples we are able to compare units across all k sites for all values of X.

Assumption 2 (Simultaneous Overlap) For all x and all d

$$0 < \Pr(D_i = d | X = x) \tag{3}$$

By applying iterated expectations we can write $\beta_d = E[E[Y_i(0,d)|X = x]]$, which combined with assumptions 1 and 2 implies we can write β_d as a function of observed data as:

$$\beta_d = E[E[Y_i | D_i = d, X = x]] \tag{4}$$

The goal in this paper is to use the nonexperimental estimators described in the following section to adjust for observable characteristics in order to eliminate differences in average outcomes for controls among the different locations in the NEWWS. As mentioned before, the key difference between our approach and that in the existing literature is that we compare all locations *simultaneously*, as opposed to making pairwise comparison between locations. Hence, the hypothesis we test in section 5 is that

$$\beta_1 = \beta_2 = \ldots = \beta_k \tag{5}$$

The equalities in (5) form the basis of our analysis as they imply that once we control for covariates and integrate over the appropriate distribution of those covariates, the individuals in any of the k locations can be used as a comparison group for all other locations. It is important to note that the outer expectation in (4) is for the distribution of covariates over all the population (i.e., over all locations), and not over the distribution of the covariates for any given location. Hence, (5) does not imply that the average potential outcome for controls in each location is the same across locations –i.e., it does not imply that $E[Y_i(0,d)|D_i = d] = E[Y_i(0,d)|D_i = f]$ for $d \neq f$.

3.1 The Role of Local Market Economic Conditions

The approach previously described to evaluate different estimators based on selection-on-observables assumptions relies on equalizing average outcomes for controls among different locations. However, average outcomes may fail to equalize even after controlling for observable characteristics because of differences in local labor market conditions across sites. Since assumptions 1 and 2 imply that controlling for pre-treatment variables is enough to make individuals comparable across sites, Hotz et al. (2005) also call Assumption 1 the "no macro-effects" assumption. They discuss the role of macro effects in the context of using the outcomes either from the control or the treatment group in one or more locations to predict the corresponding average outcomes in a different location. They note that variables that are constant within a particular location automatically fail the overlap assumption. For instance, in a case where there is only one cohort of control individuals in each location and hence the available local labor market conditions variables (e.g., unemployment rate) are constant for all individuals within a location, these variables do not satisfy the overlap assumption. This happens because, with a fixed number of sites, the probability of finding another site with the same local economic conditions is zero, so the overlap assumption is violated. In addition, as discussed in the introduction, the literature on binary treatments has stressed the importance of comparing treatment and control units from the same local labor market when employing nonexperimental methods (e.g., Friedlander and Robins, 1995; Heckman et al., 1997).

Based on this previous literature, local economic conditions are likely to play an important role in our setting even after controlling for observed characteristics. Since in our case we have different cohorts for each site (e.g., footnote 2), we have (potentially) some variation in the local economic conditions within each site. Ideally, we would like to have a large number of cohorts or periods in order to obtain some overlap on the local economic conditions across all sites. In this case, we would be able to exploit the variation in local economic conditions over time to identify "comparable" individuals across sites in terms of their local labor market conditions. Unfortunately, the public-use-data from NEWWS available to us contains only the individual's year of random assignment, so the maximum number of cohorts we can identify per site is three. In addition, as it will be further discussed in section 5, Riverside's local labor market conditions differ significantly from those in the rest of the sites (see, for instance, Figure 1). Hence, we would expect to have difficulties in equalizing the average outcomes for control individuals in Riverside.

Given that local economic conditions are likely to play an important role even after controlling for observed characteristics, in the analysis in section 5 we also present results controlling for them. The specific approach we use to control for these variables is discussed in the following section.

4 Non-experimental Estimators

In this section we discuss the different estimators of β_d we consider in this paper to eliminate differences in control outcomes across all sites. For comparison, we start with the raw mean estimator. Let $1(A)$ be the indicator function, which equals one if event A is true and zero otherwise. This estimator is then given by:

$$\widehat{\beta}_d^{raw} = \frac{\sum_{i=1}^{N} Y_i 1(D_i = d)}{\sum_{i=1}^{N} 1(D_i = d)} \tag{6}$$

This estimator would be unbiased for β_d if the individuals were randomized across different locations. Since the characteristics of the individuals vary across locations, $\widehat{\beta}_d^{raw}$ is a biased estimator of β_d. We use this estimator as a starting point, and we aim at reducing this bias by adjusting for differences in observable characteristics across locations under assumptions 1 and 2.

The result in (4) suggests estimating β_d using a partial mean, which is an average of a regression function over some of its regressors while holding others fixed (Newey, 1994). The regression function of Y on d and X is estimated in a first step, and then we average this function over the covariates holding the site (d) fixed. The most straightforward model for the inner expectation in (4) is a linear regression of the form:

$$E[Y_i|D_i, X_i] = \sum_{j=1}^{k} \alpha_j \cdot 1(D_i = j) + \delta' x_i \tag{7}$$

where δ is the coefficient vector for the covariates. Let the estimated coefficients in (7) be given by $\widehat{\alpha}_j$ and $\widehat{\delta}$. Then, the OLS-based estimator of β_d is given by:

$$\widehat{\beta}_d^{pmX} = \widehat{\alpha}_d + \widehat{\delta}' \left(N^{-1} \textstyle\sum_{i=1}^{N} x_i \right) \tag{8}$$

We also consider a more flexible model of (7) which contains polynomials of the continuous covariates and various interactions. We denote this estimator by $\widehat{\beta}_d^{pmXflex}$.

Recently, part of the focus in the program evaluation literature has been on more flexible ways to control for covariates. The main issue when controlling for the covariates without imposing any structure in the model is that if the dimension of X is large, then nonparametric methods become intractable because of the so-called "curse of dimensionality". The same problem arises in the binary-treatment case. In a seminal paper, Rosenbaum and Rubin (1983) showed that if the two potential outcomes from a binary treatment are independent of the treatment assignment conditional on X, then they are also independent conditional on the propensity score, defined as the probability of being in the treatment group conditional on X. This result implies that we only need to adjust for a scalar variable, as opposed to adjusting for all pretreatment variables. Since the propensity score is rarely known in practice, it is usually estimated using a logit model with interactions and high order terms in X, which can provide a relatively good approximation to the true model (e.g., Rosenbaum and Rubin, 1983; Dehejia and Wahba, 1999).

Imbens (2000) and Lechner (2001) extended the results in Rosenbaum and Rubin (1983) to the multivalued and multiple treatment setting, and Hirano and Imbens (2004) further extended them to the continuous treatment case. The main difference between the approaches in Imbens (2000) and Lechner (2001) is that, while the latter reduces the dimension of the conditioning set from the dimension of X to the dimension of the treatment, Imbens (2000) reduces the dimension to one, just as in the binary case.

Following Imbens (2000), define the generalized propensity score or GPS as the probability of receiving a particular treatment (in our case, belonging to a particular site) conditional on the covariates:

$$r(d,x) = \Pr(D = d|X = x) \tag{9}$$

For the discussion below, it is important to keep in mind the distinction between two different random variables: the probability that an individual gets the treatment she actually received, $R_i = r_i(D_i, X_i)$, and the probability she receives a particular treatment d conditional on her covariates, $R_i^d = r_i(d, X_i)$. Clearly, $R_i^d = R_i$ for those units with $D_i = d$.

Imbens (2000) shows that under unconfoundedness (Assumption 1) we can estimate the average potential outcomes by conditioning solely on the GPS.[6] In particular, in our context the result in Imbens (2000) can be written as:

$$\begin{aligned} (i)\ \gamma(d,r) &\equiv E[Y(0,d)|r(d,X) = r] = E[Y_i|D = d, r(D,X) = r] \\ (ii)\ E[Y(0,d)] &= E[\gamma(d, r(d,X))] \end{aligned} \tag{10}$$

Therefore, the GPS can be used to estimate $\beta_d = E[Y(0,d)]$ by following the two steps in (10). First, one estimates the conditional expectation of Y as a function of D and

[6]Note that, similar to the binary-treatment case, the problem of nonparametrically estimating the regression function of the outcome on the treatment and the covariates is translated to nonparametrically estimating the GPS. In practice, however, it may be preferable to impose restrictions (such as linearity) on the GPS rather than directly on the outcome.

$R = r(D, X)$ (i.e., the probability an individual gets the treatment she actually received). Second, to estimate β_d, we average the conditional expectation $\gamma(d, r)$ over $R^d = r(d, X)$. Hence, the averaging takes place over the values of the propensity score at the location corresponding to the parameter we want to estimate, in this case site d. As stressed in Imbens (2000), note that the second averaging is done over R^d, and not R. In addition, contrary to the binary-treatment case, in the multivalued or multiple treatment setting the conditional expectation $\gamma(d, r)$ does not have a causal interpretation.

The result in (10) suggests estimating β_d using a partial mean. However, contrary to the partial mean estimated using the covariates directly, we now use R_i in the regression function in the first step, and integrate over the distribution of R_i^d in the second step. As before, the regressor that is fixed in the second step is the site.

Hirano and Imbens (2004) implement this approach by estimating the regression function in the first step using a (flexible) parametric regression. Following their approach, we first estimate the regression function

$$E[Y_i|D_i, R_i] = \sum_{j=1}^{k} \alpha_j \cdot 1(D_i = j) + \sum_{j=1}^{k} [\delta_j \cdot 1(D_i = j) \cdot R_i + \eta_j \cdot 1(D_i = j) \cdot R_i^2]$$

Let the estimated coefficients from this regression be denoted by a hat on top of the coefficient. Next, we estimate β_d as:

$$\widehat{\beta}_d^{pmGPS} = E[Y(0)] = \frac{1}{N} \sum_{i=1}^{N} [\widehat{\alpha}_d \cdot 1(D_i = d) + \widehat{\delta}_d \cdot 1(D_i = d) \cdot R_i^d + \widehat{\eta}_j \cdot 1(D_i = d) \cdot (R_i^d)^2]$$

Alternatively, following Newey (1994) and more recently Flores (2007), we consider a more flexible specification in which the first step estimator of the regression function is based on a nonparametric kernel estimator. However, instead of employing the usual Nadaraya-Watson estimator, we use a local polynomial of order one. This estimator has the advantage that it does not have the boundary bias problem the former has. Since in our case the treatment is not continuous as in Flores (2007), the nonparametric regression function of Y_i on D_i and R_i in the first stage is equivalent to having one nonparametric regression function of Y_i on R_i for each site. To formalize the estimator, let $K(u)$ be a kernel function such that $\int K(u)\, du = 1$; let h be a bandwidth satisfying $h \to 0$ and $Nh \to \infty$ as $N \to \infty$; and, let $K_h(u) = h^{-1} K(u/h)$. Then, the nonparametric estimator of $\gamma(d, r)$ in (10), $\widehat{\gamma}(d, r; h)$ is given by:[7]

$$\widehat{\gamma}(d,r;h) = \frac{1}{N}\sum_{i=1}^{N}\frac{\{\widehat{s}_2(r,h) - \widehat{s}_1(r,h)(R_i - r)\} K_h(R_i - r) \cdot Y_i \cdot 1(D_i = d)}{\widehat{s}_2(r,h)\,\widehat{s}_0(r,h) - \widehat{s}_1(r,h)^2} \tag{11}$$

where

$$\widehat{s}_v(r,h) = \frac{1}{N}\sum_{i=1}^{N}(R_i - r)^v K_h(R_i - r) \cdot 1(D_i = d)$$

Based on (11), our nonparametric partial mean estimator of β_d is given by:

$$\widehat{\beta}_d^{pmNPR} = \frac{1}{N}\sum_{j=1}^{N}\widehat{\gamma}\left(d, R_j^d; h\right)$$

In the next section we implement this approach by using an Epanechnikov kernel and select the bandwidth using Silverman's rule: $h = 1.06\min\{\widehat{\sigma}, I/1.34\}N^{-1/5}$, where $\widehat{\sigma}$ is the standard deviation of R_i and I is the interquartile range (e.g., Härdle et al., 2004).

In addition to employing the GPS within a partial mean framework to estimate β_d, the GPS can also be used to control for covariates using a weighting approach. Similar to the binary treatment case, in a multiple or multivalued treatment case one can weight the observations receiving a given treatment level t by the probability of receiving the treatment they actually received conditional on X (i.e., R_i). More specifically, applying the results in Imbens (2000) to our context we can write β_d as a function of the observed data as

$$\beta_d = E\left[\frac{Y_i \cdot 1(D_i = d)}{R_i}\right]$$

where as before, $R_i = r(D_i, X_i)$. Based on this result, a possible estimator of β_d is its sample analogue given by replacing the expectation by the empirical average $N^{-1}\sum_{i=1}^{N}\cdot$. However, similar to the binary case discussed in Imbens (2004), this estimator has the undesirable property that its weights do not necessarily add to one. An alternative is to normalize the weights to add to one. Thus, the estimator we use in this case is given by

$$\widehat{\beta}_d^{ipw} = \sum_{i=1}^{N}\left[\frac{Y_i \cdot 1(D_i = d)}{R_i}\right]\left[\sum_{i=1}^{N}\frac{1(D_i = d)}{R_i}\right]^{-1},$$

where *ipw* stands for inverse probability weight estimator. Similar to the binary-treatment

[7] See, for instance, Wand and Jones (1995).

case, note that $\widehat{\beta}_d^{ipw}$ for $d = 1, \ldots, k$ can also be calculated from the weighted linear regression

$$E\left[Y_i|D_i\right] = \sum_{j=1}^{k} \beta_j^{ipw} \cdot 1\left(D_i = j\right), \tag{12}$$

with weights equal to

$$w_i = \sqrt{\frac{1}{R_i}}$$

Following Imbens (2004), we also consider an inverse probability weight estimator that adds covariates to the weighted regression in (12).[8] Hence, we first estimate the weighted regression

$$E\left[Y_i|D_i, X_i\right] = \sum_{j=1}^{k} \alpha_j \cdot 1\left(D_i = j\right) + \delta' X_i,$$

with weights w_i. Next, we estimate β_d using the estimated coefficients of this weighted regression as:[9]

$$\widehat{\beta}_d^{ipwX} = \widehat{\alpha}_d + \widehat{\delta}' \left(N^{-1} \textstyle\sum_{i=1}^{N} x_i\right). \tag{13}$$

4.1 Implementation Issues

So far we have ignored two important issues in the implementation of the approaches based on the GPS: estimation of the GPS and imposition of the overlap restriction. As in the binary-treatment case, the correct model underlying the GPS is unknown, and a nonparametric approach to its estimation becomes infeasible as the number of covariates grows. In this paper we follow an analogous approach to the binary-treatment setting and estimate the GPS using a flexible multinomial logit that includes interactions and higher order terms of the pretreatment variables.

The overlap condition in Assumption 2 is stronger than that of the binary-treatment case, as it requires that we find comparable individuals across *all sites* for all values of X. In practice, when working with a binary treatment the usual approach is to drop units in the treatment or control group for which it is not possible to find a comparable individual in the other treatment arm, i.e., drop those individuals whose propensity score does not overlap with the propensity score of those in the other treatment arm. Hence, by doing this one redefines the parameter of interest to be conditional on the subpopulation with common

[8] For a discussion of this estimator in the binary-treatment case see, for instance, Imbens and Wooldridge (2008).

[9] In the binary-treatment case this second step is not needed since the weighted regression includes a treatment indicator (and a constant), and the focus is on estimating the treatment effect. Since here the parameter of interest is the average potential outcome, this second step is needed.

overlap on the GPS.

The general idea of overlap in the multivalued case is similar to that for the binary case, but since now we want to compare different treatments simultaneously, we need to find comparable individuals across all treatment groups for all different treatments. Let the overlap region with respect to treatment (in our case location) d be given by the subsample

$$Overlap_d = \left\{ i : R_i^d \in \left[\max_{j=1,\ldots,k} \left\{ \min_{\{q:D_q=j\}} R_q^d \right\}, \min_{j=1,\ldots,k} \left\{ \max_{\{q:D_q=j\}} R_q^d \right\} \right] \right\} \tag{14}$$

Then, we define the overlap or common support region as the subsample given by those units that are in the overlap regions for all different sites

$$Overlap = \bigcap_{d=1}^{k} Overlap_d \tag{15}$$

All the estimators based on the GPS are applied within the overlap region given by (15). By restricting our attention to units within the overlap region, we guarantee that we are able to simultaneously find comparable units in terms of observable characteristics in all locations. In order to analyze the importance of comparing "comparable" units in the multivalued or multiple case, we also implement the non-GPS-based estimators discussed in this section using the entire sample as well as only those units in the overlap region.

In addition to the estimators described above, and based on the discussion in section 3.1, we also present estimators that incorporate local economic conditions (LEC) into our analysis. As explained above, we cannot introduce the local economic conditions directly in the estimation of the GPS because we do not have enough variation within sites to identify the effects of LEC separately from the site effects. So, as an alternative to control for these variables, we first regress the outcome of interest on local labor market variables and then we apply the methods discussed in this section using the residuals from this regression as the outcome variable. The idea is that the residuals represent the variation across individuals and sites that is not due to variations in LEC. The local economic condition variables we use in the first-stage regression are the two-year growth rates of the employment to population ratio and of average real earnings. The rates are calculated as the log of the variable in the year of randomization minus the log of the variable two years before. We focus on growth rates because there is more variation on them than in the variables in levels, which helps differentiate the LEC effects from fixed site effects. Also, note that in Figure 1 we present an additional variable of interest, the unemployment rate, which we did not include in the regression because of the very high negative correlation of its growth rate to the growth rate of the employment to population ratio (-0.9). All local labor market variables are measured at the metropolitan statistical area (MSA) level.

In order to make possible the comparison of the results from applying the estimators to different outcomes and to simplify their presentation, we standardize all outcomes with respect to the mean and standard deviation of each outcome across all sites. Again, to insure comparability we perform the standardization of outcomes both before and after imposing the overlap condition (15), so that the mean of all outcomes (even when the overlap condition is applied) across all sites is always zero. Therefore, the target for our estimators when applied to these standardized outcomes is to get as close to zero as possible, if the estimators are successful in controlling for differences in observable characteristics.

Finally, since in our framework we are performing multiple comparisons it is important to have an overall measure of distance of our estimates from this target of zero. Let $\widehat{\beta}_d$ denote a particular estimator of β_d as applied to the standardized data. We use the following two distance measures from zero in the next section. The root mean square distance ($rmsd$) is given by:

$$rmsd = \sqrt{\frac{1}{k}\sum_{d=1}^{k}\widehat{\beta}_d^2}, \tag{16}$$

and the mean absolute distance is given by:

$$mad = \frac{1}{k}\sum_{d=1}^{k}\left|\widehat{\beta}_d\right|. \tag{17}$$

If a particular estimator would succeed in completely eliminating all differences across all sites, then these distances would be zero. Of course, they will never equal be zero, but the closer they get to zero, the better the performance of the estimator. Given that all our outcomes are standardized to be unit-free, we can compare these distance measures across both, estimators and outcomes.

5 Results

In this section we implement the estimators discussed in the previous section using three outcomes: (i) an indicator variable equal to one if the individual was ever employed during the two years following randomization, and zero otherwise; (ii) a "differences" version of the first outcome, in which we subtract an indicator variable for whether the individual was ever employed during the two years prior to randomization; (iii) the residuals from a regression of the first outcome on the two-year growth rates of employment to population ratio and average real earnings. Since the purpose of the last outcome is to control for local economic conditions, we refer to the estimates based on this outcome as "adjusted by LEC".

As discussed in Section 2, we start our analysis by focusing on the control groups in five locations: Atlanta, Detroit, Grand Rapids, Portland and Riverside.

The first five columns of Table 1 show the descriptive statistics of the outcomes and covariates in each of these sites. The covariates include information on demographic and family characteristics, education, housing type and stability, welfare and food stamps use history, and earnings and employment history. In addition, at the end of the table we present the employment to population ratio, average real earnings and unemployment rate in the metropolitan statistical area (MSA) of each site. The table also shows the two-year growth rates of employment to population ratio and average real earnings, which are used to adjust for local economic conditions as described in the previous section.

As expected, there are important and statistically significant differences in the covariates across sites. For instance, while the percentage of blacks in Atlanta is 95 percent, this percentage is only 17 percent in Riverside. Also, individuals in Riverside appear to have better employment attachment and earnings histories, higher education level and less history of dependence on welfare and food stamps aid. The second panel of Table 1 presents the same information, but after the overlap or common support condition in (15) is applied. The bottom of the two panels in the table show that 1,503 out of 9,351 units (about 16%) do not satisfy this condition and are dropped from all analyses where overlap is imposed. In general, it can be seen that for most variables the mean values by site get closer to each other after imposing overlap; however, in most cases these changes are small.

As mentioned in Section 4, we estimate the GPS using a multinomial logit model. All individual level covariates presented in Table 1 were included in the estimation. We use the estimated GPS to study how well covariates are "balanced" across sites given the GPS. We follow two strategies to evaluate balancing. The first one tests, for *each covariate*, if there is joint equality of means across all sites. The results from this test are presented in the first column of panel A in Table 2. Clearly, imposing overlap by itself does not make a difference, since the tests are rejected for all covariates. When we perform the same test weighting each observation by the inverse of the GPS (inverse probability weighting), 10 out of 52 covariates appear as not balanced at the 5% significance level.

The second approach we use to check the balancing of covariates based on the GPS consists of a series of pairwise comparisons of the mean of each site versus the mean of the (pooled) remaining sites, as proposed by Hirano and Imbens (2004). The results from this approach are presented in panel B of Table 2. The results shown in this table for the two "raw means" versions (before and after imposing the overlap condition) of this approach correspond to an equality test of these two means. The third version ("Blocking on GPS") consists of dividing the units in a given site (e.g., site d) by the decile of the GPS for their own site (i.e., $R_i = R_i^d$) and, for each decile, calculating the difference between the mean in a given decile and the mean of those individuals in other sites whose estimated GPS for that particular site (i.e., R_i^d) falls in the same decile. For example, for individuals in Atlanta in the first decile of the estimated GPS for individuals living in Atlanta, we chose as comparison group all the individuals living in other sites whose GPS of being in Atlanta is in the same first decile. The weighted (by the number of individuals) average of these difference of means (and the corresponding standard error) are used to test the equality of means of each site versus the other sites. The results in Table 2 regarding balancing of the covariates for the five sites based on blocking on the GPS are mixed. On one hand, for most sites the number of covariates with significant differences decreases with the application of blocking. On the other hand, for some sites (Atlanta and Detroit) the number of unbalanced covariates remains relatively high.[10]

Next we calculate all the estimators presented in Section 4 using the three above mentioned outcomes. Each of the panels in Table 3 corresponds to one of these outcomes. The table presents the point estimates for each site along with its corresponding 95 percent confidence interval based on 500 bootstrap replications. As discussed in the previous section, given that all the outcomes are standardized with respect to its overall mean and standard deviation, the target of these estimators is zero. The table also presents, for each site and estimator, the p-value from a joint equality test that all parameters are equal, as well as the measures of distance, *rmsd* and *mad*, defined in (16) and (17), respectively. Figures 2, 3 and 4 present the same information of Table 3, with each figure corresponding to each of the outcomes.

Regarding the outcome in levels in Figure 2, the raw estimates for Grand Rapids and Riverside start relatively far from zero. The estimates when adjusting for covariates get closer to zero for Grand Rapids; however, for Riverside these estimates remain far from zero. The same holds true for Riverside even after using the outcome in differences. Hence, none of the estimators solely based on the use of individual covariates seem to help in equalizing the mean outcome of Riverside to that of the remaining sites. On the other hand, once we adjust for LEC in the last panel of Table 3 (Figure 4), the average outcomes for Riverside are much closer to those in the remaining sites.

Riverside is a special site in our data set in the sense that its local economic conditions are very different to those in the remaining sites. This can be seen, for instance, in Figure 1,

[10] Appendix Tables 1 an 2 present the variable-by-variable results on the tests used to generate Table 2. Note that in those tables, all covariates are standardized by their mean and standard deviation. In Appendix Table 1 we present (standardized) means of covariates by site (and the p-values for tests of equality of means), while in Appendix Table 2) we present the standardized differences of means between each site and all other sites, and indicate the significance level of those differences.

where the unemployment rates, employment to population ratios and average real earnings are presented for the different randomization cohorts in each site. It is important to note that Riverside is not only different in terms of the levels of these variables (note how much lower is the employment to population ratio, for example), but also in their dynamics: while in the other four sites things are improving in the local markets after randomization, the opposite is true for Riverside. This is not surprising as California had a much stronger recession in the early 1990s than the rest of the country.

In an attempt to improve the results presented in Table 3, we divide our sample into two groups following Hotz et al. (2005): those individuals ever employed and those never employed in the two years prior to random assignment. The idea is that these two groups are too heterogeneous between each other, and thus we could improve the performance of the estimators by dealing with them separately. We repeat the same analysis as before for each group separately. As shown in Table 2, dividing the sample in this way greatly improves the balancing of the covariates for the ever employed group, and also improves it (to a lesser extent) for those never employed. However, when we apply our estimators to the ever employed group the results (shown in Table 4) do not improve much as compared to the case when the groups are pooled. Moreover, for the never employed, consistent with the poor balancing in covariates shown in Table 2, the results (shown in Table 5) worsen. In all, the biggest problem that seems to arise for our estimators, both in the pooled and non-pooled versions of the analysis, is dealing with Riverside. Clearly, Riverside's labor market appears to be very different from the labor markets of the other sites, and there is no adjustment in individuals differences that seems to work properly in this situation.

In order to study the performance of the estimators when applied to locations where the local economic conditions are relatively more similar, we repeat our analysis *excluding* Riverside. Table 2 shows the balancing of the covariates when looking at the four remaining sites. The results suggest that the GPS is doing a very good job in balancing the individual covariates across sites. Table 6 and Figures 5, 6 and 7 present the estimation results for this case. For the outcome in levels, the standardized average outcomes for Atlanta, Detroit and Portland start very close to zero, but the mean for Grand Rapids is relatively high. In this case, the estimators seem to do a good job in equalizing average outcomes across all groups. In most of the cases, the estimators reduce the $rmsd$ and the mad by more than 50 percent. Something similar occurs with the outcome in differences and the one adjusted for LEC.

Finally, Table 7 presents some summary statistics from the 500 bootstrap replications for the estimators applied to the three outcomes when we consider four sites. In particular, we show the p-values resulting from a joint equality test of the parameters across locations based on a Wald test. For each estimator, the first column shows the p-value computed from the original data, which corresponds to the one presented in the fifth column in Table 3. The second column presents the proportion of times the equality test is rejected at a 5 percent level in the bootstrap replications; and the next three columns show some percentiles of the p-values resulting from the replications. As expected, while for the raw estimator the Wald test rejects the null hypothesis of equality in all replications, those models that adjust for covariates are a lot less likely to reject it. In addition, for the *rmsd* and the *mad* of each estimator Table 7 shows: the original estimate as shown in Table 6, the original estimate as a fraction of the value for the raw estimator without imposing overlap, the bias, the standard error, the root mean square error, the minimum, the fifth percentile, the median, the ninety five percentile and the maximum. The patterns that emerge from Table 7 suggest that the GPS based estimators are not doing so well; it probably implies that we need to refine further our estimation of the GPS. Also, once we drop Riverside from the analysis and keep the sites that appear to be more homogeneous in terms of economic condition, we observe that adjusting by LEC does not seem to be so relevant any more.

6 Conclusion

[To be completed]

References

[1] Behrman, Jere R., Yingmei Cheng, and Petra E. Todd. 2004. Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach. Review of Economics and Statistics 86, no. 1 (February): 108-132.

[2] Cattaneo, Matias. 2007. Efficient Semiparametric Estimation of Multi-valued Treatment Effects. University of California, Berkeley, November.

[3] Dehejia, Rajeev. 2005. Practical propensity score matching: a reply to Smith and Todd. Journal of Econometrics 125, no. 1-2: 355-364.

[4] Dehejia, Rajeev H., and Sadek Wahba. 1999. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. Journal of the American Statistical Association 94, no. 448 (December): 1053-1062.

[5] Dehejia, Rajeev H., and Sadek Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. Review of Economics and Statistics 84, no. 1 (February): 151-161.

[6] Flores, Carlos A. 2007. Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment. University of Miami, Department of Economics, November.

[7] Flores-Lagunes, Alfonso, Arturo Gonzalez, and Todd C. Neumann. 2007. Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps. IZA Discussion Paper, no. 2846 (June).

[8] Friedlander, Daniel (1988). Subgroups Impacts and Performance Indicators for Selected Welfare Employment Programs. New York: Manpower Demonstration Research Corporation.

[9] Friedlander, Daniel, and Philip K. Robins. 1995. Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods. The American Economic Review 85, no. 4 (September): 923-937.

[10] Frölich, Markus, Almas Heshmati, and Michael Lechner. 2004. A microeconometric evaluation of rehabilitation of long-term sickness in Sweden. Journal of Applied Econometrics 19, no. 3: 375-396.

[11] Hamilton, G.; Freedman, S.; Gennetian, L.; Michalopoulos, C.; Walter, J.; Adams-Ciardullo, D.; Gassman-Pines, A.; McGroder, S.; Zaslow, M. Brooks. J. and Ahluwalia S. (2001). How Effective are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs. Washington, D.C.: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation and Administration for Children and Families, and U.S. Department of Education.

[12] Härdle, W.; Müller, M.; Sperlich, S. and Werwatz, A. (2004). Nonparametric and Semiparametric Models. Springer Series in Statistics.

[13] Heckman, James J., Hidehiko Ichimura„ and Petra E. Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. The Review of Economic Studies 64, no. 4: 605-654.

[14] Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1998a. Matching As An Econometric Evaluation Estimator. The Review of Economic Studies 65, no. 2: 261-294.

[15] Heckman, James J., Hidehiko Ichimura, J. Smith and Petra E. Todd. 1998b. Characterizing Selection Bias Using Experimental Data. Econometrica, 66, 1017-1098.

[16] Heckman, J. and J. Hotz (1989). Alternative Methods for Evaluating the Impact of Training Programs, (with discussion), Journal of the American Statistical Association, 84, 804, 862-874.

[17] Heckman, J., LaLonde, R. and Smith, J. (1999) "The Economics and Econometrics of Active Labor Market Programs" in O. Ashenfelter and D. Card (eds.) Handbook of Labor Economics. Elsevier Science North Holland, 1865-2097.

[18] Hirano, K. and Imbens, G. W. (2004). “The Propensity Score with Continuous Treatments”, in Andrew Gelman and Xiao-Li Meng eds., Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. John Wiley and Sons, 73-84.

[19] Holland, P. W. (1986). “Statistics and Causal Inference”(with Discussion and Reply), Journal of the American Statistical Association, 81(396), 945-970.

[20] Hotz, V. Joseph, Guido W. Imbens, and Julie H. Mortimer. 2005. Predicting the efficacy of future training programs using past experiences at other locations. Journal of Econometrics 125, no. 1-2: 241-270.

[21] Imbens, Guido W. 2000. The Role of the Propensity Score in Estimating Dose-Response Functions. Biometrika 87, no. 3 (September): 706-710.

[22] Imbens, G. (2004) "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review" Review of Economics and Statistics, 84, 4-29.

[23] Imbens G. and Wooldridge, J. (2008). Recent Developments in the Econometrics of Program Evaluation, National Bureau of Economic Research Working Paper #14251.

[24] Kluve, Jochen, Hilmar Schneider, Arne Uhlendorff, and Zhong Zhao. 2007. Evaluating Continuous Training Programs Using the Generalized Propensity Score. IZA Discussion Paper, no. 3255.

[25] LaLonde, Robert J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. The American Economic Review 76, no. 4: 604-620.

[26] Lechner, Michael. 2001. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In Econometric Evaluation of Labour Market Policies, ed. Michael Lechner and Friedhelm Pfeiffer, 43-58. ZEW Economic Studies 13. New York: Springer-Verlag.

[27] Michalopoulos, Charles, Howard S. Bloom, and Carolyn J. Hill. 2004. Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? Review of Economics and Statistics 86, no. 1: 156-179.

[28] Mitnik, Oscar A. 2008. Intergenerational transmission of welfare dependency: The effects of length of exposure. University of Miami, Department of Economics, March.

[29] Mueser, Peter; Troske, Kenneth and Gorislavsky, Alexey. 2007. Using State Administrative Data to Measure Program Performance. Review of Economics and Statistics 89, no. 4: 761-783.

[30] Newey, W. K. (1994). "Kernel Estimation of Partial Means and a General Variance Estimator", Econometric Theory, 10, 233-253.

[31] Neyman, J. (1923) "On the Application of Probability Theory to Agricultural Experiments: Essays on Principles" Translated in Statistical Science, 5, 465-80.

[32] Plesca, Miana and Smith, Jeffrey. 2007. Evaluating Multi-treatment Programs: Theory and Evidence from the U.S. Job Training Partnership Act Experiment. Empirical Economics 32. 491-528.

[33] Rubin, D. (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies" Journal of Educational Psychology, 66, 688-701.

[34] Smith, Jeffrey A., and Petra E. Todd. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125, no. 1-2: 305-353.

[35] Wand, M. P. and Jones, M. C. (1995). Kernel Smoothing, Chapman and Hall, London, New York.

Table 1. Descriptive Statistics NEWWS Data

Variables	Before Imposing Overlap Conditon					After Imposing Overlap (5 sites)					After Imposing Overlap (4 sites)			
	ATL	DET	GRP	POR	RIV	ATL	DET	GRP	POR	RIV	ATL	DET	GRP	POR
Outcomes														
Ever employed in 2 years after RA	0.59	0.59	0.70	0.59	0.46	0.59	0.59	0.69	0.58	0.44	0.58	0.59	0.68	0.58
	(0.49)	(0.49)	(0.46)	(0.49)	(0.50)	(0.49)	(0.49)	(0.46)	(0.49)	(0.50)	(0.49)	(0.49)	(0.47)	(0.49)
Ever employed in 2 years after RA (Diff)	0.07	0.15	0.03	0.05	-0.07	0.07	0.15	0.04	0.06	-0.06	0.07	0.15	0.06	0.08
	(0.57)	(0.57)	(0.55)	(0.57)	(0.58)	(0.57)	(0.57)	(0.56)	(0.58)	(0.58)	(0.57)	(0.57)	(0.56)	(0.59)
Covariates														
Demographic & Family Characteristics														
Black	0.95	0.89	0.41	0.20	0.17	0.95	0.88	0.46	0.24	0.24	0.95	0.89	0.50	0.26
	(0.22)	(0.32)	(0.49)	(0.40)	(0.38)	(0.22)	(0.32)	(0.50)	(0.42)	(0.43)	(0.22)	(0.31)	(0.50)	(0.44)
Age 30-39 years old	0.51	0.35	0.29	0.40	0.45	0.50	0.35	0.31	0.40	0.44	0.50	0.35	0.33	0.40
	(0.50)	(0.48)	(0.46)	(0.49)	(0.50)	(0.50)	(0.48)	(0.46)	(0.49)	(0.50)	(0.50)	(0.48)	(0.47)	(0.49)
Age 40+ years old	0.14	0.11	0.09	0.08	0.13	0.13	0.11	0.09	0.08	0.13	0.13	0.11	0.09	0.08
	(0.34)	(0.32)	(0.28)	(0.27)	(0.34)	(0.34)	(0.31)	(0.28)	(0.28)	(0.34)	(0.34)	(0.32)	(0.29)	(0.27)
Teenage mother (at <=19 years)	0.45	0.45	0.51	0.34	0.35	0.45	0.45	0.52	0.36	0.37	0.45	0.45	0.50	0.35
	(0.50)	(0.50)	(0.50)	(0.47)	(0.48)	(0.50)	(0.50)	(0.50)	(0.48)	(0.48)	(0.50)	(0.50)	(0.50)	(0.48)
Never married	0.62	0.69	0.58	0.49	0.34	0.63	0.69	0.60	0.51	0.38	0.63	0.69	0.60	0.51
	(0.48)	(0.46)	(0.49)	(0.50)	(0.47)	(0.48)	(0.46)	(0.49)	(0.50)	(0.48)	(0.48)	(0.46)	(0.49)	(0.50)
Any child 0-5 years old	0.42	0.65	0.69	0.71	0.58	0.43	0.66	0.67	0.69	0.59	0.44	0.65	0.67	0.69
	(0.49)	(0.48)	(0.46)	(0.46)	(0.49)	(0.50)	(0.47)	(0.47)	(0.46)	(0.49)	(0.50)	(0.48)	(0.47)	(0.46)
Any child 6-12 years old	0.70	0.48	0.43	0.52	0.59	0.69	0.49	0.45	0.52	0.58	0.69	0.48	0.46	0.52
	(0.46)	(0.50)	(0.49)	(0.50)	(0.49)	(0.46)	(0.50)	(0.50)	(0.50)	(0.49)	(0.46)	(0.50)	(0.50)	(0.50)
Two children in household	0.34	0.30	0.36	0.33	0.32	0.34	0.30	0.37	0.33	0.33	0.34	0.30	0.36	0.33
	(0.47)	(0.46)	(0.48)	(0.47)	(0.47)	(0.47)	(0.46)	(0.48)	(0.47)	(0.47)	(0.47)	(0.46)	(0.48)	(0.47)
Three or more children in household	0.31	0.27	0.19	0.30	0.28	0.32	0.28	0.20	0.29	0.28	0.31	0.27	0.21	0.29
	(0.46)	(0.44)	(0.39)	(0.46)	(0.45)	(0.47)	(0.45)	(0.40)	(0.45)	(0.45)	(0.46)	(0.45)	(0.41)	(0.45)
Education Characteristics														
10th grade	0.14	0.15	0.13	0.17	0.11	0.14	0.14	0.14	0.18	0.12	0.14	0.15	0.13	0.18
	(0.35)	(0.35)	(0.34)	(0.38)	(0.31)	(0.34)	(0.35)	(0.34)	(0.38)	(0.33)	(0.35)	(0.35)	(0.34)	(0.38)
11th grade	0.17	0.25	0.20	0.22	0.18	0.18	0.26	0.21	0.20	0.17	0.18	0.25	0.21	0.20
	(0.38)	(0.44)	(0.40)	(0.41)	(0.38)	(0.38)	(0.44)	(0.40)	(0.40)	(0.38)	(0.38)	(0.43)	(0.41)	(0.40)
Grade 12 or higher	0.57	0.50	0.54	0.45	0.57	0.58	0.50	0.52	0.45	0.54	0.57	0.50	0.53	0.46
	(0.49)	(0.50)	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)
Highest degree = High School or GED	0.53	0.48	0.54	0.53	0.59	0.53	0.49	0.53	0.52	0.54	0.53	0.48	0.53	0.52
	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Housing Type & Housing Stability														
Lives in public/subsidized house	0.59	0.07	0.16	0.29	0.09	0.58	0.07	0.16	0.26	0.09	0.57	0.07	0.16	0.22
	(0.49)	(0.26)	(0.37)	(0.46)	(0.29)	(0.49)	(0.26)	(0.37)	(0.44)	(0.29)	(0.49)	(0.26)	(0.36)	(0.42)
One or two moves in past 2 years	0.49	0.48	0.51	0.47	0.54	0.49	0.49	0.53	0.48	0.53	0.50	0.49	0.54	0.48
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
3 or more moves in past 2 years	0.08	0.08	0.25	0.23	0.22	0.08	0.08	0.22	0.20	0.21	0.07	0.07	0.19	0.19
	(0.27)	(0.27)	(0.43)	(0.42)	(0.41)	(0.27)	(0.27)	(0.42)	(0.40)	(0.41)	(0.26)	(0.26)	(0.39)	(0.40)
Welfare Use History														
On welfare for less than 2 years	0.26	0.23	0.38	0.32	0.44	0.25	0.23	0.35	0.32	0.40	0.25	0.23	0.34	0.31
	(0.44)	(0.42)	(0.49)	(0.47)	(0.50)	(0.43)	(0.42)	(0.48)	(0.47)	(0.49)	(0.43)	(0.42)	(0.48)	(0.46)

(continues in next page)

Table 1. Descriptive Statistics NEWWS Data (continuation)

Variables	Before Imposing Overlap Conditon					After Imposing Overlap Conditon					After Imposing Overlap (4 sites)			
	ATL	DET	GRP	POR	RIV	ATL	DET	GRP	POR	RIV	ATL	DET	GRP	POR
On welfare for 2-5 years	0.25	0.25	0.31	0.35	0.28	0.25	0.26	0.32	0.35	0.29	0.25	0.25	0.32	0.35
	(0.43)	(0.44)	(0.46)	(0.48)	(0.45)	(0.43)	(0.44)	(0.47)	(0.48)	(0.46)	(0.43)	(0.44)	(0.47)	(0.48)
On welfare 5-10 years	0.24	0.24	0.17	0.23	0.16	0.25	0.24	0.18	0.23	0.16	0.25	0.24	0.19	0.22
	(0.43)	(0.43)	(0.38)	(0.42)	(0.37)	(0.43)	(0.43)	(0.39)	(0.42)	(0.37)	(0.43)	(0.43)	(0.39)	(0.42)
Received Welfare in Q1 before RA	0.97	0.90	0.77	0.79	0.73	0.97	0.90	0.82	0.85	0.78	0.97	0.91	0.84	0.86
	(0.18)	(0.29)	(0.42)	(0.41)	(0.44)	(0.18)	(0.29)	(0.38)	(0.36)	(0.42)	(0.18)	(0.29)	(0.37)	(0.35)
Received Welfare in Q2 before RA	0.93	0.86	0.70	0.74	0.49	0.93	0.85	0.75	0.80	0.61	0.93	0.86	0.76	0.80
	(0.26)	(0.35)	(0.46)	(0.44)	(0.50)	(0.26)	(0.35)	(0.43)	(0.40)	(0.49)	(0.26)	(0.35)	(0.43)	(0.40)
Received Welfare in Q3 before RA	0.85	0.84	0.68	0.72	0.46	0.85	0.84	0.73	0.77	0.56	0.85	0.84	0.73	0.76
	(0.36)	(0.37)	(0.47)	(0.45)	(0.50)	(0.36)	(0.37)	(0.45)	(0.42)	(0.50)	(0.36)	(0.36)	(0.44)	(0.43)
Received Welfare in Q4 before RA	0.73	0.83	0.67	0.69	0.44	0.76	0.82	0.71	0.73	0.51	0.76	0.83	0.71	0.72
	(0.44)	(0.38)	(0.47)	(0.46)	(0.50)	(0.43)	(0.38)	(0.46)	(0.44)	(0.50)	(0.42)	(0.38)	(0.46)	(0.45)
Received Welfare in Q5 before RA	0.69	0.81	0.64	0.64	0.41	0.71	0.81	0.68	0.69	0.48	0.72	0.81	0.68	0.68
	(0.46)	(0.40)	(0.48)	(0.48)	(0.49)	(0.45)	(0.40)	(0.47)	(0.46)	(0.50)	(0.45)	(0.39)	(0.47)	(0.47)
Received Welfare in Q6 before RA	0.66	0.79	0.61	0.61	0.39	0.68	0.79	0.64	0.65	0.45	0.69	0.79	0.64	0.64
	(0.47)	(0.41)	(0.49)	(0.49)	(0.49)	(0.47)	(0.41)	(0.48)	(0.48)	(0.50)	(0.46)	(0.41)	(0.48)	(0.48)
Received Welfare in Q7 before RA	0.64	0.77	0.56	0.58	0.37	0.66	0.77	0.59	0.61	0.42	0.67	0.77	0.60	0.61
	(0.48)	(0.42)	(0.50)	(0.49)	(0.48)	(0.47)	(0.42)	(0.49)	(0.49)	(0.49)	(0.47)	(0.42)	(0.49)	(0.49)
Food Stamps Use History														
Received FS in Q1 before RA	0.97	0.94	0.85	0.86	0.62	0.97	0.94	0.89	0.91	0.74	0.97	0.94	0.88	0.90
	(0.17)	(0.23)	(0.36)	(0.35)	(0.48)	(0.17)	(0.23)	(0.31)	(0.28)	(0.44)	(0.17)	(0.23)	(0.32)	(0.29)
Received FS in Q2 before RA	0.95	0.89	0.76	0.81	0.42	0.95	0.89	0.81	0.87	0.57	0.95	0.89	0.81	0.85
	(0.22)	(0.31)	(0.43)	(0.39)	(0.49)	(0.22)	(0.32)	(0.39)	(0.34)	(0.49)	(0.23)	(0.31)	(0.39)	(0.36)
Received FS in Q3 before RA	0.90	0.87	0.72	0.79	0.39	0.90	0.87	0.77	0.84	0.53	0.90	0.87	0.77	0.81
	(0.30)	(0.34)	(0.45)	(0.41)	(0.49)	(0.30)	(0.34)	(0.42)	(0.37)	(0.50)	(0.30)	(0.34)	(0.42)	(0.39)
Food Stamps Use History (continued)														
Received FS in Q4 before RA	0.83	0.86	0.72	0.76	0.36	0.83	0.86	0.75	0.81	0.48	0.84	0.86	0.75	0.79
	(0.38)	(0.35)	(0.45)	(0.43)	(0.48)	(0.37)	(0.35)	(0.43)	(0.39)	(0.50)	(0.36)	(0.35)	(0.44)	(0.41)
Received FS in Q5 before RA	0.78	0.83	0.67	0.72	0.33	0.79	0.83	0.71	0.77	0.44	0.80	0.83	0.70	0.75
	(0.42)	(0.38)	(0.47)	(0.45)	(0.47)	(0.41)	(0.38)	(0.46)	(0.42)	(0.50)	(0.40)	(0.37)	(0.46)	(0.43)
Received FS in Q6 before RA	0.75	0.81	0.64	0.70	0.31	0.76	0.80	0.67	0.74	0.40	0.77	0.81	0.67	0.72
	(0.43)	(0.40)	(0.48)	(0.46)	(0.46)	(0.43)	(0.40)	(0.47)	(0.44)	(0.49)	(0.42)	(0.39)	(0.47)	(0.45)
Received FS in Q7 before RA	0.72	0.78	0.60	0.66	0.29	0.73	0.78	0.63	0.69	0.38	0.74	0.78	0.63	0.68
	(0.45)	(0.41)	(0.49)	(0.47)	(0.45)	(0.44)	(0.41)	(0.48)	(0.46)	(0.48)	(0.44)	(0.41)	(0.48)	(0.47)
Employment History														
Employed in Q1 before RA	0.18	0.18	0.29	0.23	0.22	0.18	0.18	0.27	0.20	0.19	0.18	0.18	0.26	0.20
	(0.39)	(0.38)	(0.45)	(0.42)	(0.41)	(0.38)	(0.39)	(0.45)	(0.40)	(0.39)	(0.39)	(0.39)	(0.44)	(0.40)
Employed in Q2 before RA	0.18	0.18	0.30	0.25	0.25	0.17	0.18	0.28	0.22	0.21	0.18	0.18	0.26	0.21
	(0.38)	(0.38)	(0.46)	(0.43)	(0.43)	(0.38)	(0.38)	(0.45)	(0.41)	(0.40)	(0.38)	(0.38)	(0.44)	(0.40)
Employed in Q3 before RA	0.19	0.18	0.29	0.25	0.26	0.18	0.18	0.27	0.22	0.22	0.19	0.18	0.26	0.21
	(0.39)	(0.38)	(0.46)	(0.43)	(0.44)	(0.39)	(0.38)	(0.44)	(0.42)	(0.41)	(0.39)	(0.38)	(0.44)	(0.41)
Employed in Q4 before RA	0.22	0.17	0.30	0.24	0.28	0.21	0.17	0.27	0.21	0.23	0.21	0.17	0.26	0.20
	(0.41)	(0.38)	(0.46)	(0.42)	(0.45)	(0.41)	(0.38)	(0.44)	(0.41)	(0.42)	(0.41)	(0.38)	(0.44)	(0.40)
Employed in Q5 before RA	0.24	0.17	0.31	0.24	0.28	0.23	0.17	0.28	0.22	0.24	0.23	0.17	0.28	0.21
	(0.43)	(0.38)	(0.46)	(0.43)	(0.45)	(0.42)	(0.38)	(0.45)	(0.42)	(0.43)	(0.42)	(0.38)	(0.45)	(0.41)

(continues in next page)

Table 1. Descriptive Statistics NEWWS Data (continuation)

Variables	Before Imposing Overlap Conditon					After Imposing Overlap Conditon					After Imposing Overlap (4 sites)			
	ATL	DET	GRP	POR	RIV	ATL	DET	GRP	POR	RIV	ATL	DET	GRP	POR
Employed in Q6 before RA	0.27	0.18	0.34	0.25	0.29	0.26	0.18	0.32	0.23	0.25	0.25	0.18	0.30	0.22
	(0.44)	(0.38)	(0.47)	(0.43)	(0.45)	(0.44)	(0.39)	(0.47)	(0.42)	(0.44)	(0.44)	(0.38)	(0.46)	(0.41)
Employed in Q7 before RA	0.29	0.18	0.36	0.26	0.29	0.28	0.18	0.34	0.25	0.26	0.28	0.18	0.33	0.24
	(0.45)	(0.39)	(0.48)	(0.44)	(0.45)	(0.45)	(0.39)	(0.47)	(0.43)	(0.44)	(0.45)	(0.38)	(0.47)	(0.43)
Employed in Q8 before RA	0.30	0.18	0.39	0.27	0.30	0.28	0.18	0.36	0.26	0.27	0.28	0.18	0.34	0.25
	(0.46)	(0.38)	(0.49)	(0.45)	(0.46)	(0.45)	(0.38)	(0.48)	(0.44)	(0.45)	(0.45)	(0.38)	(0.47)	(0.43)
Employed at RA (self reported)	0.07	0.07	0.13	0.09	0.13	0.08	0.07	0.12	0.08	0.10	0.08	0.07	0.12	0.08
	(0.26)	(0.25)	(0.34)	(0.28)	(0.33)	(0.26)	(0.26)	(0.33)	(0.27)	(0.31)	(0.27)	(0.25)	(0.32)	(0.27)
Ever worked FT 6+ months at same job	0.72	0.46	0.64	0.77	0.71	0.71	0.48	0.63	0.74	0.69	0.71	0.47	0.64	0.74
	(0.45)	(0.50)	(0.48)	(0.42)	(0.45)	(0.45)	(0.50)	(0.48)	(0.44)	(0.46)	(0.45)	(0.50)	(0.48)	(0.44)
Earnings History (real $ /1,000)														
Earnings Q1 before RA	0.23	0.21	0.36	0.33	0.43	0.23	0.21	0.30	0.26	0.30	0.23	0.21	0.32	0.24
	(0.82)	(0.68)	(1.06)	(0.89)	(1.23)	(0.83)	(0.68)	(0.92)	(0.76)	(0.94)	(0.82)	(0.68)	(1.05)	(0.72)
Earnings Q2 before RA	0.26	0.25	0.52	0.41	0.63	0.26	0.25	0.44	0.32	0.44	0.26	0.25	0.42	0.28
	(0.85)	(0.82)	(1.29)	(1.04)	(1.55)	(0.85)	(0.83)	(1.14)	(0.89)	(1.19)	(0.86)	(0.82)	(1.19)	(0.87)
Earnings Q3 before RA	0.29	0.26	0.55	0.41	0.72	0.27	0.26	0.45	0.32	0.49	0.27	0.26	0.43	0.30
	(0.92)	(0.89)	(1.33)	(1.07)	(1.73)	(0.88)	(0.90)	(1.15)	(0.93)	(1.31)	(0.90)	(0.89)	(1.20)	(0.91)
Earnings Q4 before RA	0.41	0.25	0.53	0.43	0.74	0.38	0.25	0.45	0.35	0.53	0.39	0.25	0.45	0.34
	(1.22)	(0.82)	(1.29)	(1.14)	(1.68)	(1.16)	(0.83)	(1.18)	(1.03)	(1.34)	(1.17)	(0.82)	(1.25)	(1.04)
Earnings Q5 before RA	0.51	0.29	0.57	0.46	0.79	0.47	0.29	0.50	0.39	0.58	0.47	0.29	0.51	0.38
	(1.27)	(0.94)	(1.32)	(1.16)	(1.82)	(1.19)	(0.94)	(1.23)	(1.05)	(1.49)	(1.21)	(0.93)	(1.30)	(1.05)
Earnings Q6 before RA	0.62	0.31	0.62	0.51	0.80	0.57	0.31	0.55	0.44	0.64	0.57	0.31	0.55	0.43
	(1.44)	(1.01)	(1.41)	(1.26)	(1.81)	(1.35)	(1.01)	(1.31)	(1.16)	(1.56)	(1.37)	(1.01)	(1.41)	(1.20)
Earnings Q7 before RA	0.72	0.32	0.68	0.54	0.83	0.65	0.33	0.61	0.50	0.68	0.65	0.32	0.63	0.49
	(1.65)	(1.06)	(1.44)	(1.31)	(1.89)	(1.56)	(1.07)	(1.35)	(1.28)	(1.60)	(1.50)	(1.06)	(1.45)	(1.30)
Earnings Q8 before RA	0.74	0.33	0.69	0.57	0.85	0.67	0.34	0.64	0.54	0.70	0.68	0.33	0.65	0.53
	(1.61)	(1.09)	(1.45)	(1.35)	(1.86)	(1.49)	(1.10)	(1.42)	(1.30)	(1.66)	(1.50)	(1.10)	(1.48)	(1.34)
Any earnings year before RA (self-rep)	0.23	0.20	0.46	0.36	0.40	0.23	0.21	0.42	0.33	0.35	0.23	0.20	0.40	0.31
	(0.42)	(0.40)	(0.50)	(0.48)	(0.49)	(0.42)	(0.40)	(0.49)	(0.47)	(0.48)	(0.42)	(0.40)	(0.49)	(0.46)
Local Economic Conditions														
Employment/Population year of RA	0.52	0.46	0.49	0.49	0.29	0.52	0.46	0.49	0.49	0.29	0.52	0.46	0.49	0.49
Average Total earnings year of RA ($1000)	32.39	35.90	29.12	30.00	27.80	32.39	35.90	29.12	29.99	27.81	32.39	35.90	29.12	29.99
Unemployment Rate year of RA	5.93	7.38	7.42	5.36	10.45	5.93	7.38	7.43	5.37	10.47	5.93	7.38	7.43	5.38
Emp/Pop growth rate 2 yrs bef RA (Δ logs)	-0.03	0.00	-0.02	0.00	-0.05	-0.03	0.00	-0.02	0.00	-0.05	-0.03	0.00	-0.02	0.00
Avg Erns grwth rate 2 yrs bef RA (Δ logs)	0.03	0.03	0.01	0.02	-0.01	0.03	0.03	0.01	0.02	-0.01	0.03	0.03	0.01	0.02
Number of observations per site	1,372	2,037	1,374	1,740	2,828	1,301	1,970	1,192	1,419	1,966	1,283	1,994	1,107	1,312
Total number of observations			9,351					7,848				5,696		

Table 2. Summary Results from Balancing of Covariates Analysis

A. Joint tests of equality of means of covariates across all sites

Method	**Number of covariates for which p-value ≤ 0.05**			
	5 sites	**5 sites ever emp**	**5 sites never emp**	**4 sites**
Raw Means Before Overlap	52	52	32	51
Raw Means After Overlap	52	46	30	52
GPS-based Inverse Probability Weighting	10	3	4	1
Total Number of Covariates	52	52	33	52

B. Tests of differences of means of covariates in one site vs all other sites pooled together

Method	**Number of covariates for which p-value ≤ 0.05**			
	5 sites	**5 sites ever emp**	**5 sites never emp**	**4 sites**
Raw Means Before Overlap				
Atlanta vs others	42	40	30	35
Detroit vs others	49	47	28	46
Grand Rapids vs others	34	32	13	48
Portland vs others	36	30	23	33
Riverside vs others	48	50	28	-
Raw Means After Overlap				
Atlanta vs others	40	34	27	30
Detroit vs others	47	36	25	45
Grand Rapids vs others	28	23	10	45
Portland vs others	27	22	21	26
Riverside vs others	43	42	23	-
Blocking on GPS				
Atlanta vs others	15	3	12	1
Detroit vs others	27	2	11	2
Grand Rapids vs others	6	0	8	0
Portland vs others	8	2	0	1
Riverside vs others	7	0	5	-
Total Number of Covariates	52	52	33	52

Table 3. Estimated Average Employment Rate in Two Years after Random Assignment - 5 sites

Estimator	ATL	DET	GRP	POR	RIV	Joint Equality Test (p-value)	Root Mean Sq. Distance	Mean Abs. Distance
A. Outcome in Levels								
RAW_NO_OV	0.04	0.05	0.27	0.05	-0.22	0.000	0.157	0.124
	[-0.01,0.09]	[0.01,0.08]	[0.22,0.31]	[0.01,0.10]	[-0.25,-0.19]			
RAW_OV	0.04	0.06	0.25	0.02	-0.26	0.000	0.166	0.128
	[0.00,0.11]	[0.03,0.10]	[0.18,0.30]	[-0.03,0.08]	[-0.32,-0.22]			
Covariates-Based								
PM_X_NO_OV	0.06	0.11	0.16	0.08	-0.23	0.000	0.142	0.127
	[0.00,0.11]	[0.06,0.16]	[0.11,0.20]	[0.04,0.12]	[-0.27,-0.19]			
PM_X_OV	0.04	0.09	0.14	0.06	-0.25	0.000	0.137	0.115
	[-0.01,0.11]	[0.04,0.14]	[0.08,0.19]	[0.00,0.11]	[-0.30,-0.20]			
PM_X_FLEX_NO_OV	0.07	0.11	0.15	0.08	-0.24	0.000	0.145	0.131
	[0.01,0.13]	[0.07,0.16]	[0.10,0.20]	[0.03,0.13]	[-0.28,-0.20]			
PM_X_FLEX_OV	0.06	0.09	0.13	0.05	-0.25	0.000	0.138	0.118
	[0.00,0.12]	[0.04,0.14]	[0.08,0.19]	[0.00,0.11]	[-0.30,-0.20]			
GPS-Based								
PM_GPS_PAR_OV	0.05	0.01	0.13	0.07	-0.25	0.000	0.131	0.102
	[-0.04,0.15]	[-0.06,0.09]	[0.05,0.18]	[0.00,0.15]	[-0.33,-0.18]			
PM_GPS_NPR_OV	0.05	0.16	0.09	0.09	-0.19	0.169	0.127	0.116
	[-0.12,0.20]	[0.03,0.27]	[-0.03,0.16]	[-0.14,0.24]	[-0.49,-0.10]			
IPW_OV	-0.12	0.07	0.10	0.10	-0.27	0.000	0.149	0.131
	[-0.46,0.21]	[-0.10,0.23]	[-0.01,0.16]	[0.01,0.19]	[-0.35,-0.20]			
IPW_X_OV	0.05	0.08	0.23	0.07	-0.34	0.000	0.189	0.152
	[-0.09,0.19]	[-0.01,0.16]	[0.13,0.28]	[-0.01,0.15]	[-0.40,-0.25]			
B. Outcome in Differences (with respecto to years 1 and 2 before RA)								
RAW_NO_OV	0.06	0.20	-0.02	0.02	-0.18	0.000	0.122	0.094
	[0.01,0.11]	[0.16,0.24]	[-0.06,0.04]	[-0.02,0.06]	[-0.21,-0.15]			
RAW_OV	0.04	0.18	-0.02	0.00	-0.19	0.000	0.117	0.086
	[-0.03,0.07]	[0.13,0.20]	[-0.08,0.04]	[-0.04,0.06]	[-0.24,-0.14]			
Covariates-Based								
PM_X_NO_OV	0.08	0.10	0.08	0.05	-0.18	0.000	0.109	0.099
	[0.02,0.14]	[0.06,0.15]	[0.04,0.13]	[0.01,0.10]	[-0.22,-0.14]			
PM_X_OV	0.05	0.09	0.06	0.03	-0.19	0.000	0.101	0.085
	[-0.01,0.12]	[0.04,0.13]	[0.01,0.12]	[-0.02,0.08]	[-0.24,-0.14]			
PM_X_FLEX_NO_OV	0.06	0.11	0.14	0.06	-0.21	0.000	0.129	0.115
	[0.01,0.11]	[0.07,0.16]	[0.10,0.18]	[0.02,0.10]	[-0.24,-0.18]			
PM_X_FLEX_OV	0.04	0.09	0.12	0.04	-0.22	0.000	0.122	0.102
	[-0.01,0.10]	[0.05,0.13]	[0.07,0.17]	[-0.01,0.08]	[-0.27,-0.18]			
GPS-Based								
PM_GPS_PAR_OV	0.06	0.09	0.11	0.09	-0.24	0.000	0.135	0.119
	[-0.05,0.13]	[0.02,0.16]	[0.03,0.15]	[0.04,0.18]	[-0.32,-0.17]			
PM_GPS_NPR_OV	0.05	0.12	0.11	0.11	-0.25	0.006	0.145	0.128
	[-0.10,0.18]	[-0.05,0.24]	[0.00,0.17]	[-0.06,0.23]	[-0.43,-0.02]			
IPW_OV	0.08	0.13	0.11	0.11	-0.24	0.000	0.146	0.135
	[-0.05,0.27]	[-0.01,0.24]	[0.02,0.16]	[0.04,0.19]	[-0.31,-0.18]			
IPW_X_OV	0.06	0.21	0.01	0.06	-0.26	0.000	0.153	0.119
	[-0.08,0.16]	[0.10,0.27]	[-0.07,0.08]	[-0.01,0.15]	[-0.31,-0.17]			
C. Outcome in Levels adjusted by Local Economic Conditions								
RAW_NO_OV	0.03	-0.08	0.27	-0.04	-0.06	0.000	0.131	0.097
	[-0.02,0.08]	[-0.11,-0.04]	[0.22,0.32]	[-0.08,0.00]	[-0.10,-0.03]			
RAW_OV	0.05	-0.05	0.27	-0.06	-0.10	0.000	0.134	0.106
	[0.01,0.12]	[-0.08,-0.01]	[0.20,0.32]	[-0.11,0.00]	[-0.15,-0.05]			
Covariates-Based								
PM_X_NO_OV	0.04	-0.01	0.16	-0.02	-0.08	0.000	0.082	0.062
	[-0.02,0.10]	[-0.06,0.04]	[0.11,0.21]	[-0.06,0.03]	[-0.12,-0.04]			
PM_X_OV	0.04	-0.02	0.15	-0.03	-0.08	0.000	0.082	0.066
	[-0.01,0.11]	[-0.07,0.02]	[0.10,0.21]	[-0.08,0.03]	[-0.13,-0.04]			
PM_X_FLEX_NO_OV	0.06	-0.01	0.15	-0.02	-0.09	0.000	0.083	0.065
	[0.00,0.12]	[-0.06,0.04]	[0.11,0.20]	[-0.06,0.03]	[-0.12,-0.05]			
PM_X_FLEX_OV	0.06	-0.02	0.15	-0.03	-0.09	0.000	0.083	0.069
	[0.00,0.12]	[-0.07,0.02]	[0.09,0.20]	[-0.08,0.03]	[-0.14,-0.04]			
GPS-Based								
PM_GPS_PAR_OV	0.05	-0.09	0.14	-0.02	-0.09	0.000	0.087	0.076
	[-0.05,0.15]	[-0.16,-0.01]	[0.06,0.19]	[-0.08,0.07]	[-0.16,-0.01]			
PM_GPS_NPR_OV	0.03	0.06	0.10	0.01	-0.03	0.849	0.056	0.047
	[-0.13,0.19]	[-0.08,0.17]	[-0.02,0.18]	[-0.22,0.16]	[-0.33,0.06]			
IPW_OV	-0.14	-0.03	0.10	0.02	-0.11	0.010	0.096	0.082
	[-0.52,0.19]	[-0.23,0.13]	[0.01,0.17]	[-0.06,0.12]	[-0.18,-0.03]			
IPW_X_OV	0.03	-0.02	0.24	-0.02	-0.18	0.000	0.136	0.098
	[-0.10,0.18]	[-0.12,0.07]	[0.14,0.30]	[-0.09,0.07]	[-0.25,-0.09]			

Notes: Bootstrap Confidence Intervals between brackets (based on 500 replications).

Table 4. Estimated Average Employment Rate in Two Years after Random Assignment - 5 sites (ever employed in 2 yrs before RA)

Estimator	ATL	DET	GRP	POR	RIV	Joint Equality Test (p-value)	Root Mean Sq. Distance	Mean Abs. Distance
A. Outcome in Levels								
RAW_NO_OV	0.07	0.13	0.16	0.04	-0.24	0.000	0.145	0.127
	[0.00,0.13]	[0.07,0.17]	[0.11,0.21]	[-0.02,0.10]	[-0.28,-0.19]			
RAW_OV	0.04	0.11	0.15	0.00	-0.30	0.000	0.161	0.123
	[-0.03,0.12]	[0.05,0.16]	[0.06,0.20]	[-0.12,0.06]	[-0.35,-0.19]			
Covariates-Based								
PM_X_NO_OV	0.06	0.11	0.15	0.10	-0.25	0.000	0.148	0.134
	[-0.01,0.13]	[0.04,0.17]	[0.09,0.20]	[0.04,0.15]	[-0.30,-0.20]			
PM_X_OV	0.04	0.06	0.14	0.06	-0.27	0.000	0.140	0.112
	[-0.03,0.12]	[0.00,0.12]	[0.05,0.19]	[-0.07,0.13]	[-0.33,-0.16]			
PM_X_FLEX_NO_OV	0.07	0.10	0.15	0.09	-0.25	0.000	0.149	0.134
	[0.00,0.14]	[0.03,0.17]	[0.10,0.21]	[0.03,0.16]	[-0.30,-0.20]			
PM_X_FLEX_OV	0.05	0.05	0.13	0.05	-0.26	0.000	0.135	0.107
	[-0.02,0.13]	[-0.02,0.11]	[0.04,0.19]	[-0.08,0.12]	[-0.33,-0.15]			
GPS-Based								
PM_GPS_PAR_OV	0.12	0.04	0.13	0.12	-0.29	0.000	0.163	0.141
	[-0.03,0.25]	[-0.05,0.15]	[0.02,0.20]	[-0.03,0.21]	[-0.40,-0.10]			
PM_GPS_NPR_OV	0.16	0.02	0.11	0.14	-0.38	0.431	0.200	0.161
	[-0.06,0.40]	[-0.17,0.23]	[-0.06,0.25]	[-0.08,0.34]	[-0.71,0.25]			
IPW_OV	0.22	0.07	0.11	0.15	-0.27	0.000	0.182	0.166
	[0.06,0.35]	[-0.11,0.23]	[-0.02,0.19]	[-0.01,0.24]	[-0.38,-0.11]			
IPW_X_OV	0.16	0.12	0.14	0.08	-0.31	0.000	0.179	0.161
	[-0.01,0.27]	[0.01,0.23]	[0.03,0.22]	[-0.08,0.16]	[-0.40,-0.15]			
B. Outcome in Differences (with respecto to years 1 and 2 before RA)								
RAW_NO_OV	0.07	0.13	0.16	0.04	-0.24	0.000	0.145	0.127
	[0.00,0.13]	[0.07,0.17]	[0.11,0.21]	[-0.02,0.10]	[-0.28,-0.19]			
RAW_OV	0.04	0.11	0.15	0.00	-0.30	0.000	0.161	0.123
	[-0.03,0.12]	[0.05,0.16]	[0.06,0.20]	[-0.12,0.06]	[-0.35,-0.19]			
Covariates-Based								
PM_X_NO_OV	0.05	0.11	0.15	0.12	-0.26	0.000	0.156	0.140
	[-0.02,0.13]	[0.04,0.18]	[0.10,0.21]	[0.06,0.18]	[-0.31,-0.21]			
PM_X_OV	0.03	0.06	0.15	0.08	-0.29	0.000	0.154	0.122
	[-0.04,0.11]	[0.00,0.12]	[0.06,0.20]	[-0.04,0.15]	[-0.35,-0.19]			
PM_X_FLEX_NO_OV	0.06	0.11	0.16	0.12	-0.27	0.000	0.160	0.144
	[-0.02,0.14]	[0.04,0.17]	[0.11,0.22]	[0.05,0.18]	[-0.32,-0.22]			
PM_X_FLEX_OV	0.04	0.06	0.15	0.07	-0.29	0.000	0.152	0.121
	[-0.03,0.13]	[-0.01,0.12]	[0.05,0.20]	[-0.06,0.14]	[-0.36,-0.19]			
GPS-Based								
PM_GPS_PAR_OV	0.12	0.04	0.13	0.12	-0.29	0.000	0.163	0.141
	[-0.03,0.25]	[-0.05,0.15]	[0.02,0.20]	[-0.03,0.21]	[-0.40,-0.10]			
PM_GPS_NPR_OV	0.16	0.02	0.11	0.14	-0.38	0.431	0.200	0.161
	[-0.06,0.40]	[-0.17,0.23]	[-0.06,0.25]	[-0.08,0.34]	[-0.71,0.25]			
IPW_OV	0.22	0.07	0.11	0.15	-0.27	0.000	0.182	0.166
	[0.06,0.35]	[-0.11,0.23]	[-0.02,0.19]	[-0.01,0.24]	[-0.38,-0.11]			
IPW_X_OV	0.15	0.11	0.14	0.05	-0.32	0.000	0.177	0.152
	[0.01,0.27]	[-0.02,0.21]	[0.02,0.20]	[-0.11,0.13]	[-0.40,-0.15]			
C. Outcome in Levels adjusted by Local Economic Conditions								
RAW_NO_OV	0.02	-0.04	0.16	-0.08	-0.03	0.000	0.085	0.068
	[-0.04,0.09]	[-0.10,0.00]	[0.11,0.21]	[-0.14,-0.02]	[-0.07,0.01]			
RAW_OV	0.02	-0.03	0.19	-0.10	-0.07	0.000	0.101	0.083
	[-0.04,0.10]	[-0.09,0.02]	[0.11,0.23]	[-0.21,-0.04]	[-0.12,0.06]			
Covariates-Based								
PM_X_NO_OV	0.01	-0.06	0.15	-0.02	-0.04	0.000	0.076	0.058
	[-0.06,0.08]	[-0.13,0.00]	[0.09,0.20]	[-0.08,0.03]	[-0.09,0.01]			
PM_X_OV	0.01	-0.09	0.17	-0.04	-0.03	0.000	0.087	0.067
	[-0.05,0.11]	[-0.15,-0.02]	[0.09,0.22]	[-0.15,0.03]	[-0.09,0.09]			
PM_X_FLEX_NO_OV	0.02	-0.07	0.15	-0.03	-0.05	0.000	0.080	0.063
	[-0.06,0.09]	[-0.14,-0.01]	[0.09,0.21]	[-0.09,0.04]	[-0.10,0.01]			
PM_X_FLEX_OV	0.02	-0.10	0.16	-0.05	-0.02	0.000	0.088	0.071
	[-0.05,0.12]	[-0.16,-0.03]	[0.08,0.22]	[-0.16,0.02]	[-0.09,0.09]			
GPS-Based								
PM_GPS_PAR_OV	0.10	-0.10	0.15	0.02	-0.06	0.009	0.096	0.086
	[-0.05,0.23]	[-0.19,0.02]	[0.05,0.22]	[-0.12,0.11]	[-0.16,0.14]			
PM_GPS_NPR_OV	0.13	-0.11	0.12	0.04	-0.14	0.440	0.115	0.109
	[-0.08,0.38]	[-0.29,0.10]	[-0.04,0.26]	[-0.16,0.24]	[-0.49,0.50]			
IPW_OV	0.19	-0.07	0.13	0.06	-0.05	0.065	0.114	0.100
	[0.04,0.32]	[-0.22,0.08]	[0.01,0.22]	[-0.09,0.15]	[-0.14,0.13]			
IPW_X_OV	0.13	-0.01	0.16	-0.02	-0.08	0.008	0.100	0.080
	[-0.03,0.24]	[-0.13,0.09]	[0.06,0.24]	[-0.17,0.06]	[-0.17,0.09]			

Notes: Bootstrap Confidence Intervals between brackets (based on 500 replications).

Table 5. Estimated Average Employment Rate in Two Years after Random Assignment - 5 sites (never employed 2 yrs before RA)

Estimator	ATL	DET	GRP	POR	RIV	Joint Equality Test (p-value)	Root Mean Sq. Distance	Mean Abs. Distance
A. Outcome in Levels								
RAW_NO_OV	0.04	0.10	0.23	0.05	-0.21	0.000	0.151	0.127
	[-0.02,0.11]	[0.05,0.15]	[0.14,0.33]	[-0.02,0.11]	[-0.26,-0.16]			
RAW_OV	0.04	0.11	0.20	0.04	-0.25	0.000	0.152	0.125
	[-0.03,0.14]	[0.05,0.16]	[0.07,0.28]	[-0.06,0.11]	[-0.36,-0.20]			
Covariates-Based								
PM_X_NO_OV	0.07	0.14	0.18	0.06	-0.25	0.000	0.158	0.141
	[-0.01,0.16]	[0.08,0.20]	[0.09,0.27]	[-0.02,0.12]	[-0.31,-0.19]			
PM_X_OV	0.04	0.14	0.15	0.05	-0.27	0.000	0.155	0.131
	[-0.04,0.15]	[0.04,0.19]	[0.03,0.22]	[-0.05,0.13]	[-0.37,-0.22]			
PM_X_FLEX_NO_OV	0.07	0.15	0.18	0.06	-0.26	0.000	0.160	0.143
	[-0.02,0.15]	[0.08,0.21]	[0.09,0.27]	[-0.01,0.12]	[-0.31,-0.20]			
PM_X_FLEX_OV	0.04	0.13	0.15	0.05	-0.27	0.000	0.154	0.131
	[-0.04,0.15]	[0.04,0.18]	[0.03,0.23]	[-0.05,0.14]	[-0.37,-0.21]			
GPS-Based								
PM_GPS_PAR_OV	-0.03	0.06	0.13	0.09	-0.27	0.000	0.143	0.116
	[-0.15,0.13]	[-0.04,0.16]	[0.00,0.23]	[-0.05,0.21]	[-0.43,-0.20]			
PM_GPS_NPR_OV	-0.04	0.28	0.11	0.11	-0.37	0.004	0.220	0.182
	[-0.26,0.17]	[0.06,0.38]	[-0.05,0.22]	[-0.16,0.29]	[-0.67,-0.05]			
IPW_OV	-0.23	0.25	0.10	0.10	-0.32	0.000	0.215	0.198
	[-0.55,0.22]	[0.01,0.38]	[-0.06,0.21]	[-0.08,0.24]	[-0.45,-0.25]			
IPW_X_OV	-0.03	0.16	0.14	0.09	-0.33	0.000	0.180	0.150
	[-0.20,0.18]	[-0.01,0.23]	[0.02,0.26]	[-0.07,0.21]	[-0.47,-0.26]			
B. Outcome in Differences (with respecto to years 1 and 2 before RA)								
RAW_NO_OV	0.04	0.10	0.23	0.05	-0.21	0.000	0.151	0.127
	[-0.02,0.11]	[0.05,0.15]	[0.14,0.33]	[-0.02,0.11]	[-0.26,-0.16]			
RAW_OV	0.04	0.11	0.20	0.04	-0.25	0.000	0.152	0.125
	[-0.03,0.14]	[0.05,0.16]	[0.07,0.28]	[-0.06,0.11]	[-0.36,-0.20]			
Covariates-Based								
PM_X_NO_OV	0.07	0.14	0.19	0.06	-0.25	0.000	0.160	0.142
	[-0.01,0.15]	[0.07,0.20]	[0.10,0.27]	[-0.02,0.12]	[-0.31,-0.20]			
PM_X_OV	0.04	0.14	0.15	0.05	-0.27	0.000	0.157	0.132
	[-0.04,0.15]	[0.04,0.18]	[0.03,0.23]	[-0.05,0.14]	[-0.38,-0.22]			
PM_X_FLEX_NO_OV	0.07	0.15	0.19	0.06	-0.26	0.000	0.162	0.144
	[-0.02,0.16]	[0.08,0.21]	[0.09,0.28]	[-0.01,0.13]	[-0.31,-0.20]			
PM_X_FLEX_OV	0.04	0.13	0.16	0.06	-0.27	0.000	0.156	0.132
	[-0.04,0.15]	[0.04,0.18]	[0.03,0.23]	[-0.04,0.14]	[-0.37,-0.21]			
GPS-Based								
PM_GPS_PAR_OV	-0.03	0.06	0.13	0.09	-0.27	0.000	0.143	0.116
	[-0.15,0.13]	[-0.04,0.16]	[0.00,0.23]	[-0.05,0.21]	[-0.43,-0.20]			
PM_GPS_NPR_OV	-0.04	0.28	0.11	0.11	-0.37	0.004	0.220	0.182
	[-0.26,0.17]	[0.06,0.38]	[-0.05,0.22]	[-0.16,0.29]	[-0.67,-0.05]			
IPW_OV	-0.23	0.25	0.10	0.10	-0.32	0.000	0.215	0.198
	[-0.55,0.22]	[0.01,0.38]	[-0.06,0.21]	[-0.08,0.24]	[-0.45,-0.25]			
IPW_X_OV	-0.04	0.16	0.13	0.09	-0.33	0.000	0.178	0.149
	[-0.21,0.17]	[-0.01,0.23]	[0.00,0.25]	[-0.08,0.21]	[-0.47,-0.26]			
C. Outcome in Levels adjusted by Local Economic Conditions								
RAW_NO_OV	0.03	-0.02	0.24	-0.04	-0.05	0.000	0.114	0.077
	[-0.04,0.10]	[-0.07,0.03]	[0.15,0.33]	[-0.11,0.02]	[-0.10,-0.01]			
RAW_OV	0.03	0.00	0.22	-0.04	-0.08	0.000	0.107	0.075
	[-0.03,0.14]	[-0.05,0.05]	[0.10,0.31]	[-0.13,0.04]	[-0.17,-0.03]			
Covariates-Based								
PM_X_NO_OV	0.06	0.02	0.19	-0.03	-0.09	0.000	0.100	0.078
	[-0.03,0.14]	[-0.05,0.08]	[0.10,0.28]	[-0.10,0.04]	[-0.15,-0.03]			
PM_X_OV	0.04	0.03	0.17	-0.03	-0.10	0.002	0.091	0.072
	[-0.04,0.15]	[-0.05,0.08]	[0.05,0.25]	[-0.11,0.06]	[-0.19,-0.04]			
PM_X_FLEX_NO_OV	0.05	0.02	0.19	-0.03	-0.10	0.000	0.100	0.079
	[-0.03,0.14]	[-0.05,0.09]	[0.10,0.28]	[-0.10,0.04]	[-0.15,-0.04]			
PM_X_FLEX_OV	0.03	0.02	0.17	-0.02	-0.10	0.002	0.091	0.070
	[-0.05,0.15]	[-0.06,0.08]	[0.06,0.26]	[-0.11,0.07]	[-0.19,-0.04]			
GPS-Based								
PM_GPS_PAR_OV	-0.03	-0.04	0.15	0.02	-0.10	0.045	0.085	0.069
	[-0.16,0.13]	[-0.14,0.06]	[0.03,0.26]	[-0.11,0.14]	[-0.25,-0.02]			
PM_GPS_NPR_OV	-0.06	0.18	0.12	0.03	-0.20	0.180	0.136	0.119
	[-0.28,0.16]	[-0.05,0.29]	[-0.02,0.25]	[-0.23,0.24]	[-0.48,0.13]			
IPW_OV	-0.27	0.15	0.12	0.02	-0.15	0.010	0.161	0.140
	[-0.60,0.21]	[-0.11,0.31]	[-0.03,0.23]	[-0.15,0.18]	[-0.27,-0.07]			
IPW_X_OV	-0.05	0.06	0.16	0.02	-0.17	0.005	0.108	0.090
	[-0.22,0.17]	[-0.11,0.14]	[0.05,0.29]	[-0.15,0.15]	[-0.30,-0.10]			

Notes: Bootstrap Confidence Intervals between brackets (based on 500 replications).

Table 6. Estimated Average Employment Rate in Two Years after Random Assignment - 4 sites

Estimator	ATL	DET	GRP	POR	Joint Equality Test (p-value)	Root Mean Sq. Distance	Mean Abs. Distance
A. Outcome in Levels							
RAW_NO_OV	-0.05	-0.05	0.18	-0.04	0.000	0.097	0.080
	[-0.09,-0.01]	[-0.08,-0.01]	[0.13,0.22]	[-0.08,0.00]			
RAW_OV	-0.04	-0.03	0.16	-0.06	0.000	0.088	0.071
	[-0.09,0.01]	[-0.06,0.01]	[0.10,0.21]	[-0.11,0.00]			
Covariates-Based							
PM_X_NO_OV	-0.03	0.00	0.05	-0.02	0.143	0.031	0.025
	[-0.08,0.02]	[-0.04,0.04]	[0.00,0.10]	[-0.07,0.04]			
PM_X_OV	-0.03	0.00	0.05	-0.02	0.297	0.030	0.024
	[-0.07,0.04]	[-0.04,0.04]	[-0.01,0.10]	[-0.08,0.04]			
PM_X_FLEX_NO_OV	-0.02	0.00	0.04	-0.02	0.283	0.025	0.020
	[-0.07,0.03]	[-0.04,0.04]	[0.00,0.09]	[-0.07,0.03]			
PM_X_FLEX_OV	-0.01	0.00	0.04	-0.02	0.494	0.025	0.019
	[-0.06,0.05]	[-0.04,0.04]	[-0.02,0.09]	[-0.08,0.03]			
GPS-Based							
PM_GPS_PAR_OV	-0.01	-0.04	0.06	0.01	0.348	0.035	0.029
	[-0.10,0.07]	[-0.10,0.04]	[-0.01,0.13]	[-0.08,0.10]			
PM_GPS_NPR_OV	-0.03	0.08	0.01	0.03	0.600	0.046	0.037
	[-0.16,0.09]	[-0.05,0.18]	[-0.10,0.13]	[-0.18,0.17]			
IPW_OV	-0.13	0.04	0.01	0.05	0.494	0.073	0.059
	[-0.34,0.09]	[-0.06,0.14]	[-0.07,0.10]	[-0.06,0.16]			
IPW_X_OV	-0.02	-0.01	0.13	-0.01	0.049	0.065	0.042
	[-0.14,0.07]	[-0.09,0.06]	[0.06,0.23]	[-0.10,0.08]			
B. Outcome in Differences (with respecto to years 1 and 2 before RA)							
RAW_NO_OV	-0.02	0.12	-0.09	-0.06	0.000	0.083	0.073
	[-0.06,0.03]	[0.09,0.16]	[-0.14,-0.05]	[-0.10,-0.02]			
RAW_OV	-0.05	0.10	-0.08	-0.04	0.000	0.068	0.064
	[-0.10,0.01]	[0.06,0.14]	[-0.14,-0.02]	[-0.10,0.02]			
Covariates-Based							
PM_X_NO_OV	0.01	0.01	0.00	-0.02	0.825	0.014	0.011
	[-0.05,0.06]	[-0.03,0.05]	[-0.05,0.05]	[-0.07,0.03]			
PM_X_OV	0.00	0.01	0.00	-0.02	0.908	0.011	0.009
	[-0.06,0.06]	[-0.03,0.05]	[-0.06,0.05]	[-0.08,0.04]			
PM_X_FLEX_NO_OV	-0.03	0.01	0.04	-0.03	0.107	0.029	0.027
	[-0.07,0.02]	[-0.02,0.05]	[0.00,0.09]	[-0.08,0.01]			
PM_X_FLEX_OV	-0.03	0.01	0.04	-0.03	0.262	0.027	0.026
	[-0.07,0.03]	[-0.02,0.05]	[-0.02,0.08]	[-0.09,0.02]			
GPS-Based							
PM_GPS_PAR_OV	-0.02	0.02	0.02	0.05	0.664	0.031	0.027
	[-0.10,0.06]	[-0.04,0.09]	[-0.05,0.08]	[-0.02,0.13]			
PM_GPS_NPR_OV	-0.01	0.02	0.01	0.06	0.934	0.032	0.025
	[-0.13,0.09]	[-0.10,0.13]	[-0.10,0.13]	[-0.11,0.20]			
IPW_OV	0.01	0.04	0.04	0.05	0.980	0.039	0.036
	[-0.09,0.16]	[-0.07,0.14]	[-0.03,0.13]	[-0.04,0.13]			
IPW_X_OV	-0.02	0.12	-0.05	0.00	0.020	0.065	0.048
	[-0.13,0.07]	[0.04,0.19]	[-0.13,0.02]	[-0.09,0.09]			
C. Outcome in Levels adjusted by Local Economic Conditions							
RAW_NO_OV	0.00	-0.04	0.16	-0.08	0.000	0.089	0.068
	[-0.04,0.04]	[-0.08,0.00]	[0.11,0.20]	[-0.12,-0.03]			
RAW_OV	0.01	-0.02	0.14	-0.09	0.000	0.083	0.065
	[-0.04,0.06]	[-0.06,0.01]	[0.08,0.19]	[-0.14,-0.04]			
Covariates-Based							
PM_X_NO_OV	0.02	0.01	0.03	-0.05	0.178	0.030	0.026
	[-0.03,0.06]	[-0.03,0.05]	[-0.01,0.08]	[-0.10,0.00]			
PM_X_OV	0.02	0.01	0.03	-0.05	0.295	0.032	0.027
	[-0.03,0.08]	[-0.04,0.04]	[-0.03,0.08]	[-0.11,0.01]			
PM_X_FLEX_NO_OV	0.03	0.01	0.02	-0.05	0.128	0.034	0.030
	[-0.02,0.08]	[-0.04,0.05]	[-0.02,0.07]	[-0.10,0.00]			
PM_X_FLEX_OV	0.03	0.00	0.02	-0.06	0.248	0.034	0.028
	[-0.02,0.09]	[-0.04,0.04]	[-0.04,0.07]	[-0.12,0.00]			
GPS-Based							
PM_GPS_PAR_OV	0.03	-0.03	0.03	-0.02	0.525	0.030	0.029
	[-0.06,0.11]	[-0.09,0.04]	[-0.04,0.10]	[-0.11,0.06]			
PM_GPS_NPR_OV	0.00	0.09	-0.02	0.00	0.620	0.045	0.026
	[-0.13,0.13]	[-0.04,0.19]	[-0.12,0.11]	[-0.21,0.14]			
IPW_OV	-0.11	0.05	-0.01	0.02	0.586	0.060	0.047
	[-0.32,0.11]	[-0.06,0.15]	[-0.10,0.08]	[-0.10,0.13]			
IPW_X_OV	0.01	0.00	0.10	-0.04	0.170	0.054	0.038
	[-0.11,0.11]	[-0.09,0.06]	[0.03,0.20]	[-0.13,0.05]			

Notes: Bootstrap Confidence Intervals between brackets (based on 500 replications).

Table 7. Summary Estimators Performance - 4 sites
Outcome: Employment Rate in Two Years after Random Assignment

	P-Value from Joint Equality Test (Wald Test)					Root Mean Square Distance (RMSD)								Mean Absolute Distance (MAD)							
	Estimated	Bootstrap Statistics				Estimated	Relative	Bootstrap Statistics						Estimated	Relative	Bootstrap Statistics					
Estimator	Value	Prop. H_0 Rejected at 5% level	5th Pctile	Median	95th Pctile	Value	Value to Raw_no_ov	Bias	S.E.	RMSE	5th Pctile	Median	95th Pctile	Value	Value to Raw_no_ov	Bias	S.E.	RMSE	5th Pctile	Median	95th Pctile
A. Outcome in Levels																					
RAW_NO_OV	0.000	1.000	0.000	0.000	0.000	0.097	1.00	0.099	0.012	0.099	0.079	0.099	0.118	0.080	1.00	0.080	0.010	0.080	0.063	0.080	0.096
RAW_OV	0.000	1.000	0.000	0.000	0.001	0.088	0.91	0.087	0.015	0.089	0.062	0.088	0.112	0.071	0.89	0.069	0.012	0.070	0.048	0.069	0.089
Covariates-Based																					
PM_X_NO_OV	0.143	0.488	0.000	0.052	0.633	0.031	0.32	0.037	0.012	0.039	0.017	0.037	0.057	0.025	0.32	0.031	0.010	0.033	0.015	0.032	0.049
PM_X_OV	0.297	0.246	0.005	0.171	0.821	0.030	0.31	0.035	0.013	0.038	0.015	0.035	0.057	0.024	0.31	0.030	0.011	0.032	0.012	0.029	0.047
PM_X_FLEX_NO_OV	0.283	0.326	0.002	0.118	0.755	0.025	0.26	0.033	0.011	0.035	0.015	0.032	0.052	0.020	0.26	0.028	0.010	0.030	0.013	0.028	0.045
PM_X_FLEX_OV	0.494	0.166	0.007	0.260	0.862	0.025	0.25	0.033	0.013	0.035	0.014	0.032	0.056	0.019	0.24	0.028	0.011	0.030	0.012	0.027	0.046
GPS-Based																					
PM_GPS_PAR_OV	0.348	0.268	0.003	0.176	0.818	0.035	0.36	0.049	0.016	0.052	0.024	0.049	0.079	0.029	0.36	0.042	0.015	0.045	0.019	0.042	0.069
PM_GPS_NPR_OV	0.600	0.150	0.006	0.393	0.947	0.046	0.47	0.078	0.047	0.091	0.032	0.070	0.151	0.037	0.47	0.064	0.033	0.072	0.026	0.058	0.124
IPW_OV	0.494	0.222	0.004	0.216	0.874	0.073	0.75	0.090	0.040	0.099	0.036	0.084	0.165	0.059	0.74	0.072	0.029	0.077	0.028	0.070	0.121
IPW_X_OV	0.049	0.746	0.000	0.009	0.342	0.065	0.67	0.081	0.020	0.083	0.047	0.080	0.115	0.042	0.53	0.064	0.017	0.066	0.037	0.063	0.093
B. Outcome in Differences (with respecto to years 1 and 2 before RA)																					
RAW_NO_OV	0.000	1.000	0.000	0.000	0.000	0.083	0.85	0.085	0.012	0.086	0.067	0.085	0.106	0.073	0.92	0.075	0.011	0.076	0.059	0.074	0.092
RAW_OV	0.000	0.996	0.000	0.000	0.002	0.068	0.70	0.076	0.014	0.077	0.054	0.076	0.099	0.064	0.81	0.069	0.013	0.070	0.049	0.069	0.090
Covariates-Based																					
PM_X_NO_OV	0.825	0.104	0.015	0.353	0.920	0.014	0.14	0.026	0.011	0.028	0.010	0.025	0.046	0.011	0.14	0.023	0.010	0.025	0.008	0.021	0.040
PM_X_OV	0.908	0.076	0.027	0.363	0.935	0.011	0.12	0.028	0.011	0.030	0.010	0.028	0.049	0.009	0.11	0.024	0.010	0.026	0.008	0.023	0.043
PM_X_FLEX_NO_OV	0.107	0.520	0.000	0.043	0.494	0.029	0.30	0.034	0.011	0.036	0.018	0.033	0.053	0.027	0.34	0.030	0.010	0.032	0.015	0.029	0.047
PM_X_FLEX_OV	0.262	0.324	0.002	0.134	0.748	0.027	0.28	0.033	0.012	0.035	0.016	0.032	0.054	0.026	0.33	0.029	0.011	0.031	0.013	0.028	0.048
GPS-Based																					
PM_GPS_PAR_OV	0.664	0.174	0.010	0.256	0.879	0.031	0.32	0.046	0.015	0.049	0.022	0.045	0.074	0.027	0.34	0.039	0.013	0.041	0.017	0.039	0.064
PM_GPS_NPR_OV	0.934	0.060	0.033	0.615	0.956	0.032	0.33	0.066	0.043	0.079	0.027	0.058	0.135	0.025	0.31	0.054	0.027	0.060	0.022	0.049	0.101
IPW_OV	0.980	0.056	0.042	0.464	0.933	0.039	0.40	0.059	0.021	0.063	0.027	0.057	0.097	0.036	0.46	0.050	0.018	0.053	0.023	0.048	0.081
IPW_X_OV	0.020	0.752	0.000	0.007	0.289	0.065	0.67	0.076	0.019	0.078	0.046	0.076	0.108	0.048	0.61	0.064	0.017	0.066	0.037	0.064	0.093
C. Outcome in Levels adjusted by Local Economic Conditions																					
RAW_NO_OV	0.000	1.000	0.000	0.000	0.000	0.089	0.92	0.091	0.013	0.092	0.069	0.091	0.110	0.068	0.86	0.073	0.010	0.073	0.055	0.072	0.090
RAW_OV	0.000	1.000	0.000	0.000	0.005	0.083	0.85	0.083	0.016	0.084	0.056	0.083	0.107	0.065	0.82	0.067	0.013	0.068	0.045	0.066	0.089
Covariates-Based																					
PM_X_NO_OV	0.178	0.438	0.000	0.069	0.639	0.030	0.31	0.037	0.012	0.039	0.017	0.036	0.058	0.026	0.33	0.031	0.011	0.033	0.015	0.031	0.051
PM_X_OV	0.295	0.318	0.001	0.133	0.762	0.032	0.33	0.039	0.014	0.042	0.017	0.038	0.065	0.027	0.34	0.033	0.012	0.035	0.014	0.032	0.054
PM_X_FLEX_NO_OV	0.128	0.484	0.000	0.057	0.678	0.034	0.35	0.040	0.013	0.042	0.017	0.039	0.063	0.030	0.37	0.034	0.012	0.036	0.015	0.034	0.055
PM_X_FLEX_OV	0.248	0.372	0.001	0.100	0.712	0.034	0.35	0.042	0.015	0.045	0.019	0.041	0.069	0.028	0.36	0.035	0.013	0.037	0.016	0.035	0.058
GPS-Based																					
PM_GPS_PAR_OV	0.525	0.160	0.008	0.295	0.911	0.030	0.31	0.044	0.016	0.047	0.019	0.043	0.073	0.029	0.37	0.037	0.014	0.040	0.015	0.036	0.064
PM_GPS_NPR_OV	0.620	0.124	0.010	0.426	0.965	0.045	0.46	0.076	0.049	0.091	0.031	0.068	0.143	0.026	0.33	0.062	0.034	0.071	0.024	0.056	0.118
IPW_OV	0.586	0.192	0.008	0.271	0.904	0.060	0.61	0.084	0.039	0.092	0.031	0.078	0.156	0.047	0.59	0.067	0.028	0.072	0.025	0.064	0.115
IPW_X_OV	0.170	0.488	0.000	0.052	0.665	0.054	0.56	0.070	0.021	0.073	0.038	0.070	0.106	0.038	0.48	0.058	0.018	0.060	0.031	0.057	0.089

Notes: Results based on 500 Bootstrap replications.

Figure 1. Local Economic Conditions

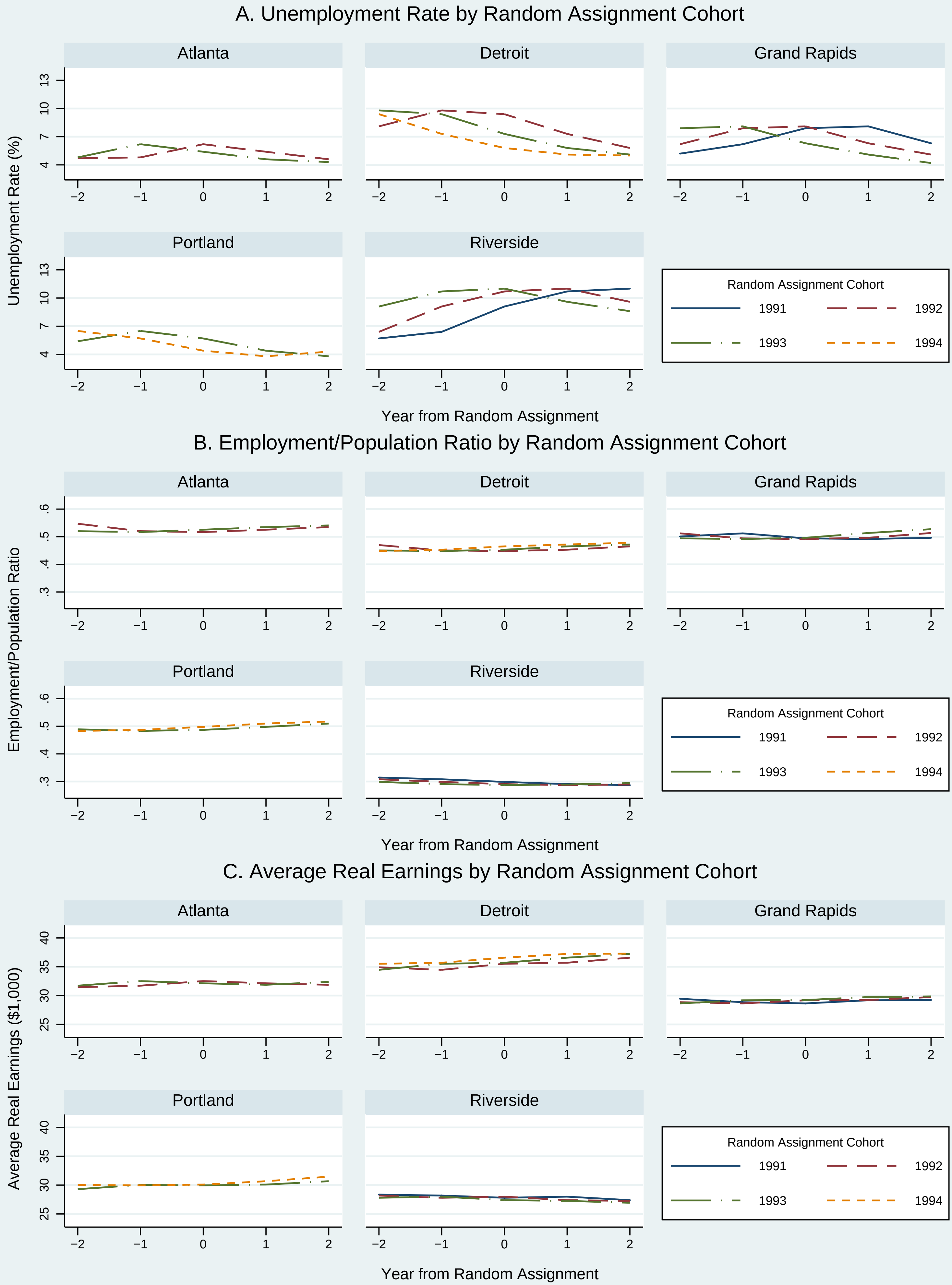


Figure 2. Outcome in Levels (5 Sites)

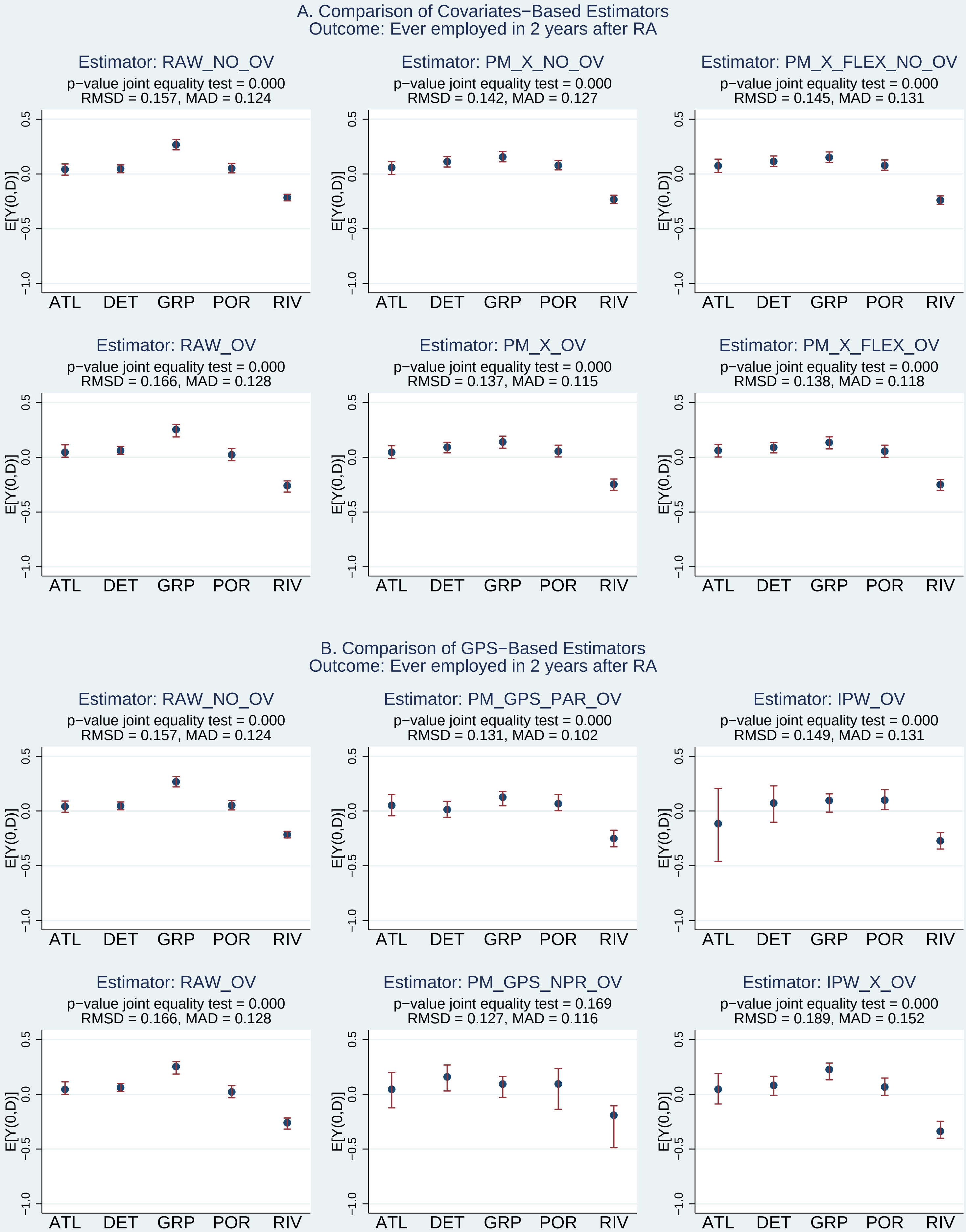


Figure 3. Outcome in DID (5 Sites)

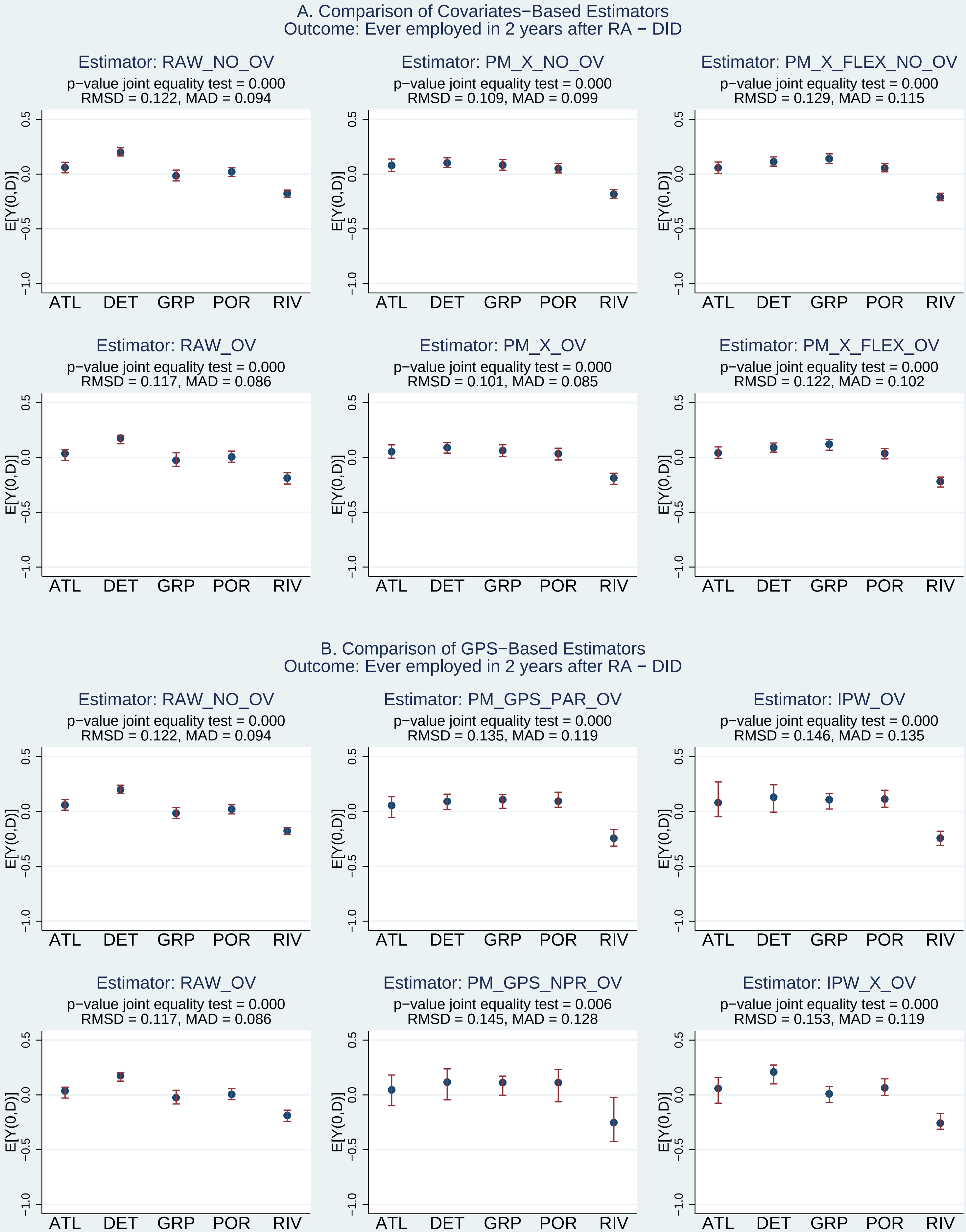


Figure 4. Outcome Adjusted by LEC (5 Sites)

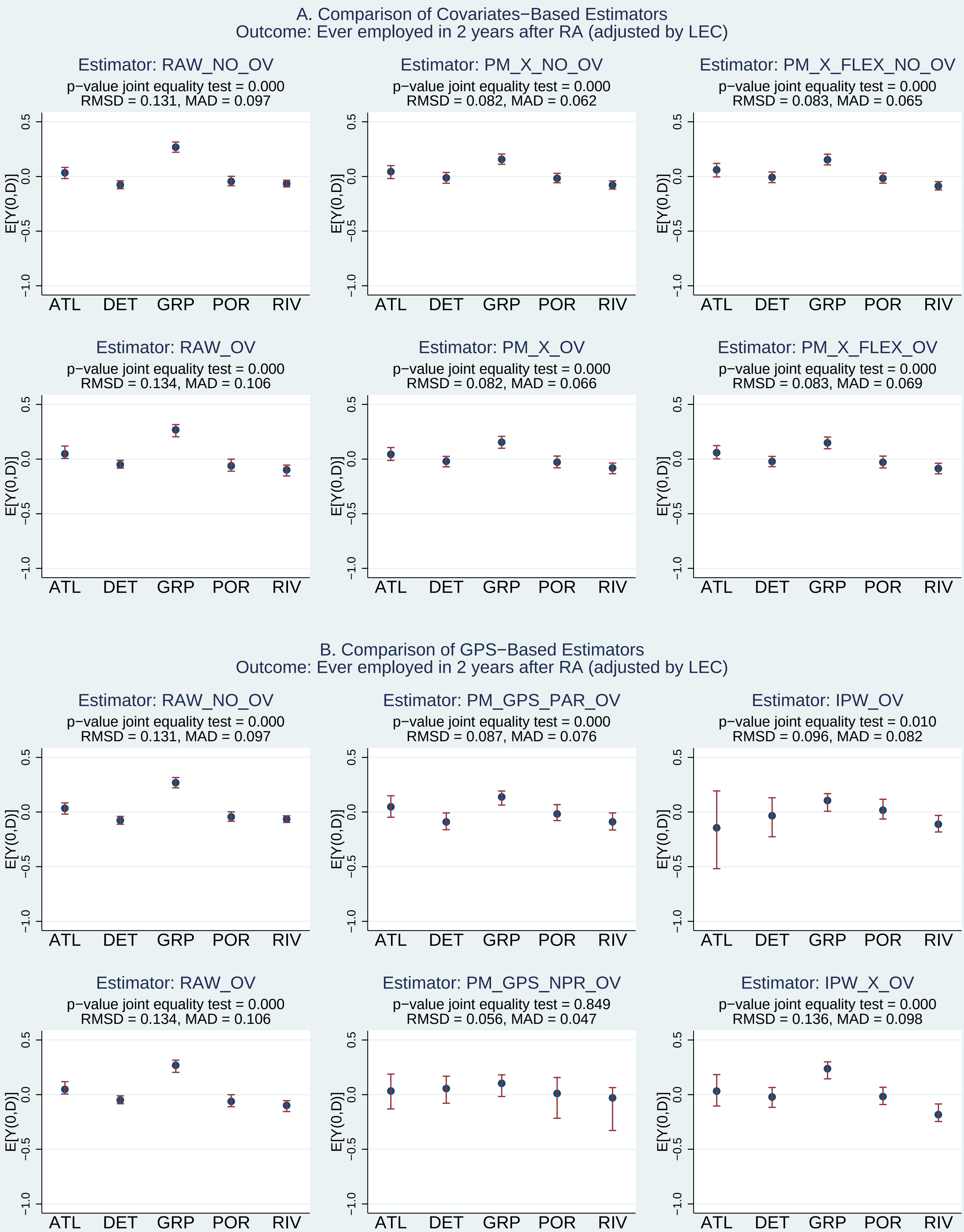


Figure 5. Outcome in Levels (4 Sites)

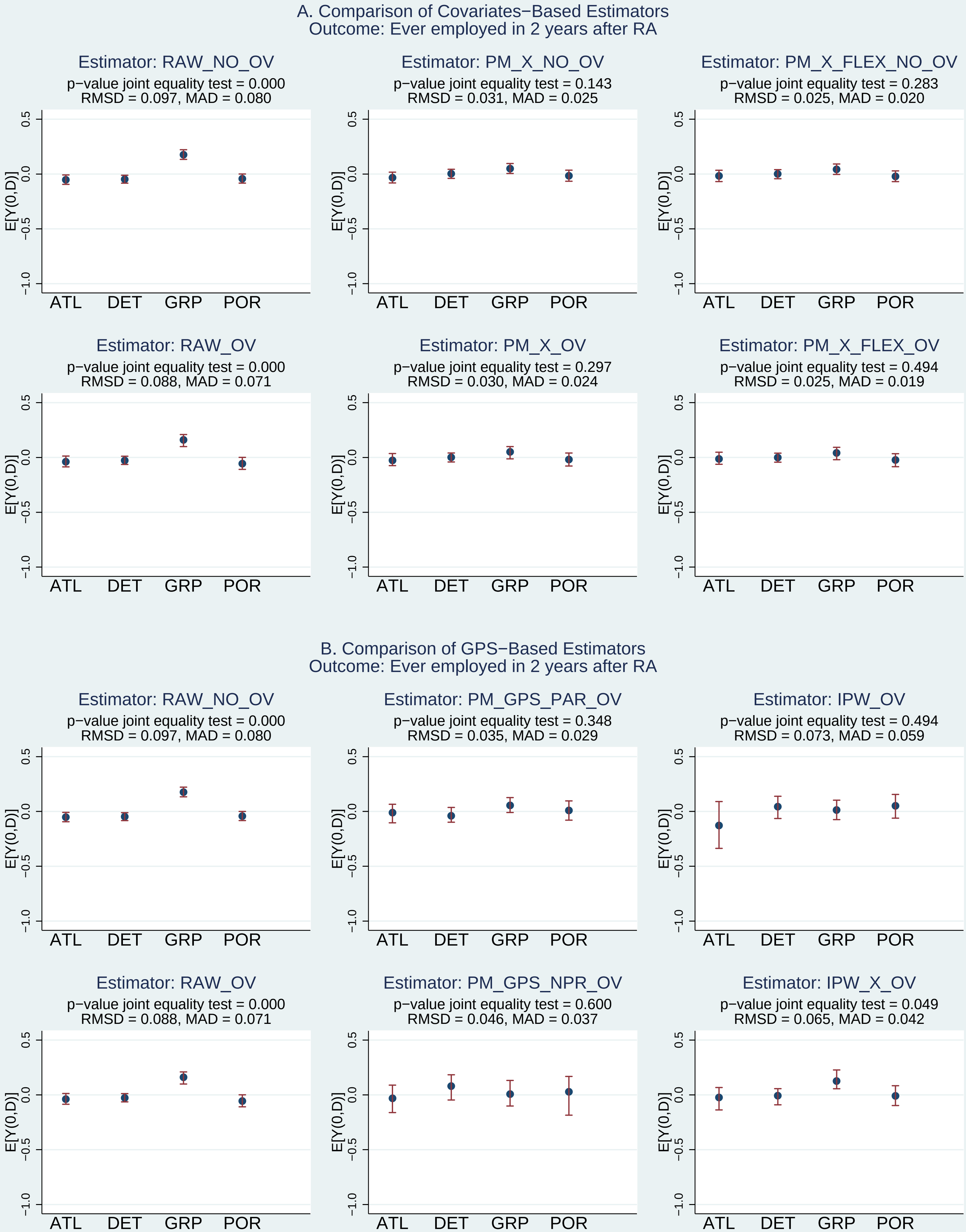


Figure 6. Outcome in DID (4 Sites)

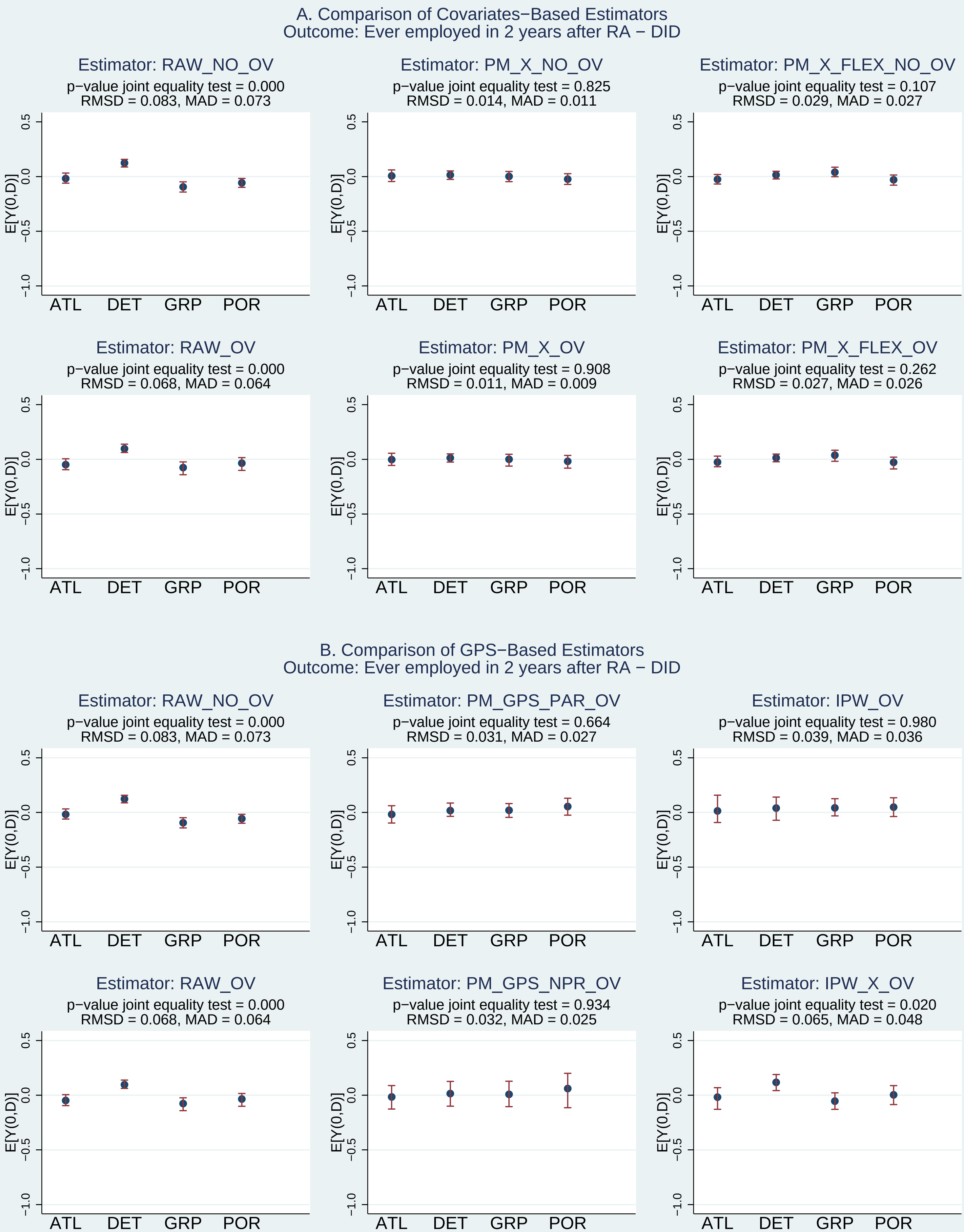
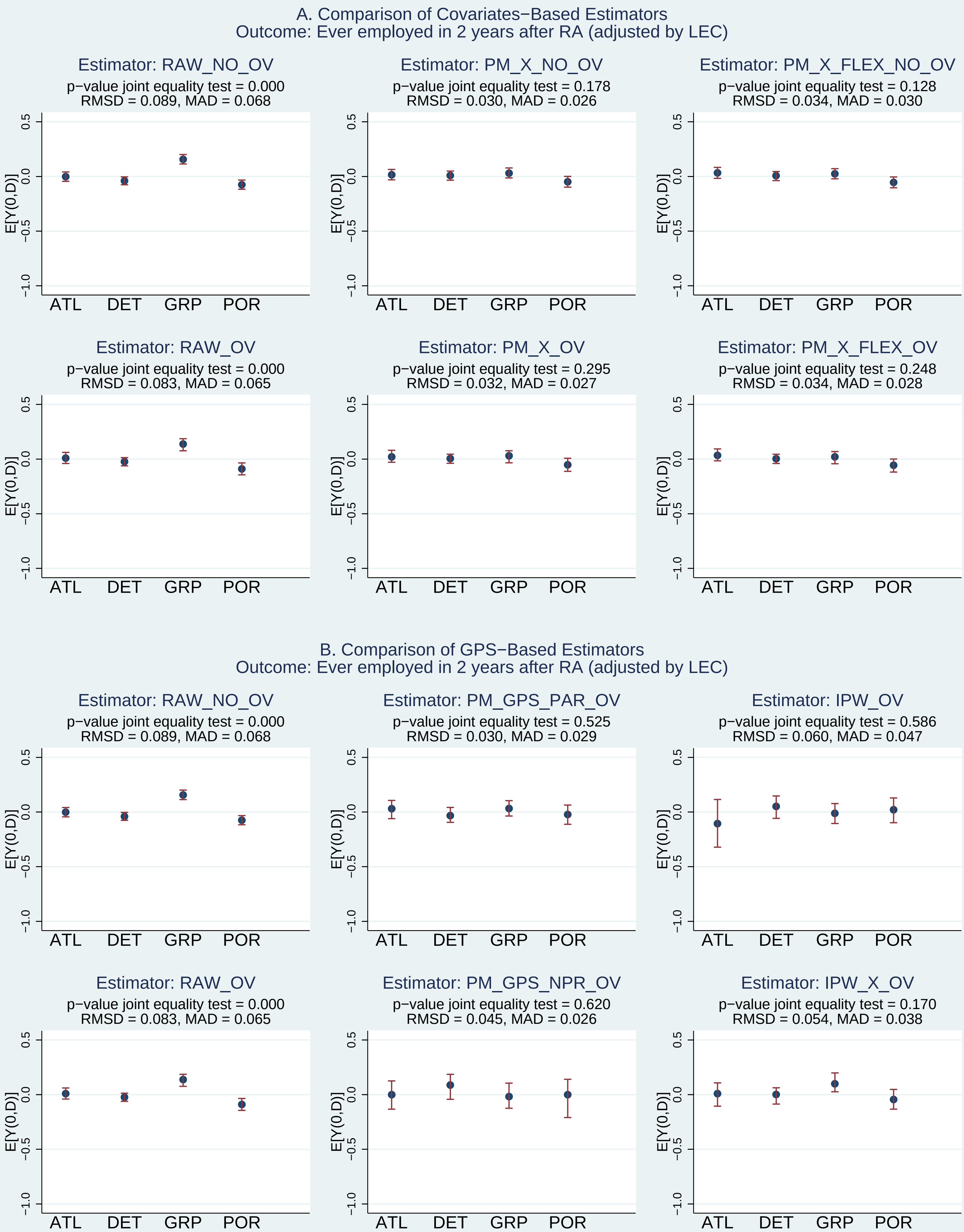


Figure 7. Outcome Adjusted by LEC (4 Sites)



Appendix Table 1. Balancing of covariates analysis based on joint tests of equality of means across all sites - 4 sites

Variable	P-Values Joint tests of equality of means of covariates across all sites			Standardized Means by Site: Means Before Overlap				Standardized Means by Site: Means After Overlap				Standardized Means by Site: Means using GPS IPW			
	Raw	Raw w/Ovlp	GPS IPW	ATL	DET	GRP	POR	ATL	DET	GRP	POR	ATL	DET	GRP	POR
Black	0.000	0.000	0.459	0.69	0.56	-0.42	-0.86	0.69	0.57	-0.24	-0.74	-0.02	0.11	0.17	0.17
Age 30-39 years old	0.000	0.000	0.487	0.25	-0.07	-0.18	0.03	0.24	-0.06	-0.12	0.04	-0.06	0.05	0.08	0.00
Age 40+ years old	0.000	0.000	0.998	0.11	0.03	-0.06	-0.07	0.09	0.03	-0.04	-0.08	-0.02	-0.01	0.00	0.00
Teenage mother	0.000	0.000	0.958	0.04	0.03	0.15	-0.19	0.05	0.04	0.13	-0.17	0.06	0.05	0.02	0.02
Never married	0.000	0.000	0.765	0.04	0.18	-0.04	-0.22	0.05	0.19	-0.01	-0.18	-0.11	-0.01	0.02	-0.01
Any child 0-5 years old	0.000	0.000	0.068	-0.43	0.06	0.14	0.17	-0.39	0.05	0.08	0.14	0.13	-0.06	-0.08	-0.12
Any child 6-12 years old	0.000	0.000	0.384	0.36	-0.10	-0.20	-0.01	0.32	-0.08	-0.13	-0.02	-0.09	-0.03	0.06	0.00
2 children in household	0.002	0.003	0.197	0.02	-0.07	0.06	0.01	0.01	-0.06	0.07	0.00	0.07	0.04	-0.07	0.05
3+ children in household	0.000	0.000	0.141	0.09	0.00	-0.18	0.07	0.10	0.01	-0.13	0.04	-0.06	-0.03	0.10	-0.04
10th grade	0.013	0.023	0.668	-0.03	-0.01	-0.05	0.07	-0.03	-0.01	-0.04	0.07	0.16	0.01	0.01	-0.05
11th grade	0.000	0.000	0.905	-0.10	0.09	-0.04	0.00	-0.09	0.09	-0.01	-0.04	-0.02	0.00	0.03	-0.02
Grade 12 or higher	0.000	0.000	0.314	0.13	-0.02	0.06	-0.12	0.13	-0.02	0.04	-0.11	-0.03	0.06	-0.04	0.08
Highest degree = HS/GED	0.001	0.031	0.278	0.02	-0.08	0.05	0.04	0.02	-0.07	0.02	0.00	-0.09	0.03	-0.09	-0.08
Lives public/subss house	0.000	0.000	0.187	0.77	-0.43	-0.23	0.08	0.72	-0.43	-0.23	-0.09	-0.09	0.07	-0.01	-0.09
1-2 moves in past 2 years	0.161	0.008	0.688	0.00	-0.01	0.05	-0.03	0.02	0.00	0.12	-0.01	0.10	0.03	0.07	0.12
3+ moves in past 2 years	0.000	0.000	0.214	-0.21	-0.21	0.27	0.20	-0.22	-0.22	0.11	0.11	-0.17	-0.04	-0.12	-0.05
On welfare < 2 years	0.000	0.000	0.006	-0.07	-0.14	0.20	0.06	-0.11	-0.14	0.11	0.04	0.10	0.13	-0.09	-0.11
On welfare for 2-5 years	0.000	0.000	0.469	-0.09	-0.08	0.04	0.13	-0.09	-0.08	0.06	0.14	-0.09	-0.02	0.05	-0.01
On welfare 5-10 years	0.000	0.001	0.157	0.04	0.04	-0.12	0.01	0.06	0.04	-0.09	0.00	0.05	-0.09	0.01	0.02
On Welfare Q1 before RA	0.000	0.000	0.462	0.31	0.13	-0.26	-0.20	0.31	0.14	-0.06	0.00	-0.06	0.02	0.11	0.07
On Welfare Q2 before RA	0.000	0.000	0.683	0.31	0.13	-0.28	-0.17	0.30	0.13	-0.11	-0.02	-0.08	0.04	0.09	0.08
On Welfare Q3 before RA	0.000	0.000	0.939	0.17	0.16	-0.22	-0.14	0.18	0.16	-0.10	-0.04	0.01	0.06	0.09	0.09
On Welfare Q4 before RA	0.000	0.000	0.952	-0.01	0.20	-0.15	-0.11	0.06	0.20	-0.07	-0.03	0.02	0.03	0.06	0.04
On Welfare Q5 before RA	0.000	0.000	0.633	-0.03	0.23	-0.13	-0.13	0.04	0.23	-0.06	-0.05	0.03	-0.02	0.07	0.05
On Welfare Q6 before RA	0.000	0.000	0.087	-0.03	0.24	-0.14	-0.15	0.03	0.24	-0.07	-0.08	0.06	-0.05	0.13	0.07
On Welfare Q7 before RA	0.000	0.000	0.142	-0.02	0.25	-0.18	-0.14	0.05	0.25	-0.11	-0.08	0.06	-0.03	0.13	0.07
Rec. FS in Q1 before RA	0.000	0.000	0.939	0.22	0.12	-0.20	-0.16	0.21	0.13	-0.08	-0.01	-0.08	0.06	0.05	0.03
Rec. FS in Q2 before RA	0.000	0.000	0.955	0.27	0.10	-0.27	-0.11	0.26	0.11	-0.12	-0.01	-0.06	0.04	0.05	0.04
Rec. FS in Q3 before RA	0.000	0.000	0.934	0.21	0.12	-0.27	-0.09	0.21	0.12	-0.15	-0.03	-0.04	0.03	0.06	0.05
Rec. FS in Q4 before RA	0.000	0.000	0.929	0.08	0.15	-0.20	-0.08	0.12	0.16	-0.13	-0.02	-0.04	0.01	0.04	0.01
Rec. FS in Q5 before RA	0.000	0.000	0.630	0.05	0.17	-0.20	-0.08	0.10	0.18	-0.14	-0.02	-0.06	-0.04	0.04	0.05
Rec. FS in Q6 before RA	0.000	0.000	0.192	0.04	0.17	-0.21	-0.07	0.09	0.17	-0.14	-0.03	-0.02	-0.06	0.09	0.06
Rec. FS in Q7 before RA	0.000	0.000	0.122	0.04	0.19	-0.22	-0.08	0.08	0.19	-0.15	-0.05	-0.04	-0.08	0.09	0.06
Employed Q1 before RA	0.000	0.000	0.323	-0.08	-0.09	0.17	0.03	-0.08	-0.09	0.11	-0.05	-0.13	-0.11	-0.04	-0.03
Employed Q2 before RA	0.000	0.000	0.729	-0.11	-0.10	0.18	0.06	-0.11	-0.10	0.10	-0.04	-0.14	-0.08	-0.06	-0.06
Employed Q3 before RA	0.000	0.000	0.838	-0.08	-0.11	0.17	0.06	-0.08	-0.12	0.08	-0.03	-0.11	-0.07	-0.09	-0.04
Employed Q4 before RA	0.000	0.000	0.983	-0.02	-0.12	0.17	0.02	-0.03	-0.13	0.08	-0.06	-0.07	-0.08	-0.07	-0.09
Employed Q5 before RA	0.000	0.000	0.653	0.02	-0.14	0.17	0.02	-0.01	-0.15	0.10	-0.05	-0.12	-0.03	-0.07	-0.03
Employed Q6 before RA	0.000	0.000	0.518	0.04	-0.16	0.21	0.00	0.01	-0.16	0.11	-0.07	-0.13	0.00	-0.07	-0.06
Employed Q7 before RA	0.000	0.000	0.734	0.05	-0.19	0.22	0.00	0.03	-0.19	0.15	-0.06	-0.12	-0.02	-0.06	-0.07
Employed Q8 before RA	0.000	0.000	0.824	0.05	-0.21	0.26	0.01	0.03	-0.21	0.16	-0.06	-0.07	-0.03	-0.09	-0.09
Emply at RA (self reported)	0.000	0.000	0.411	-0.05	-0.07	0.15	-0.01	-0.04	-0.07	0.11	-0.02	-0.10	-0.04	-0.02	0.00
Ever wrkd FT 6+ mths sm. job	0.000	0.000	0.955	0.17	-0.36	0.01	0.27	0.15	-0.35	0.01	0.22	0.01	-0.03	-0.04	-0.06
Earnings Q1 before RA	0.000	0.012	0.501	-0.06	-0.08	0.10	0.06	-0.06	-0.08	0.05	-0.04	-0.04	-0.09	0.04	-0.03
Earnings Q2 before RA	0.000	0.000	0.945	-0.09	-0.10	0.17	0.06	-0.09	-0.10	0.07	-0.06	-0.02	-0.05	0.00	-0.05
Earnings Q3 before RA	0.000	0.001	0.928	-0.08	-0.10	0.17	0.04	-0.09	-0.10	0.06	-0.06	-0.04	-0.07	-0.03	-0.04
Earnings Q4 before RA	0.000	0.000	0.984	0.02	-0.13	0.13	0.03	-0.01	-0.13	0.05	-0.05	-0.02	-0.05	-0.03	-0.04
Earnings Q5 before RA	0.000	0.000	0.838	0.06	-0.13	0.11	0.02	0.03	-0.13	0.06	-0.05	-0.02	-0.09	-0.06	-0.05
Earnings Q6 before RA	0.000	0.000	0.988	0.10	-0.14	0.10	0.01	0.06	-0.14	0.05	-0.05	-0.06	-0.03	-0.03	-0.04
Earnings Q7 before RA	0.000	0.000	0.825	0.13	-0.16	0.10	0.00	0.08	-0.16	0.07	-0.04	-0.06	0.01	-0.02	-0.05
Earnings Q8 before RA	0.000	0.000	0.960	0.13	-0.17	0.10	0.01	0.09	-0.17	0.07	-0.02	-0.03	-0.01	-0.04	-0.05
Any earns yr before RA (slf-rep)	0.000	0.000	0.240	-0.17	-0.22	0.33	0.13	-0.15	-0.22	0.21	0.01	-0.21	-0.15	-0.11	-0.06

Notes: Variables have been standardized to mean zero and standard deviation 1 (before imposing overlap)

Appendix Table 2.
Balancing of covariates analysis based on standardized differences of means in one site vs all other sites pooled together - 4 sites

Variable	**Means Before Overlap**				**Means After Overlap**				**Means after Blocking on GPS**			
	ATL	**DET**	**GRP**	**POR**	**ATL**	**DET**	**GRP**	**POR**	**ATL**	**DET**	**GRP**	**POR**
Black	0.87***	0.81***	-0.53	-1.17	0.71***	0.66***	-0.47	-1.13	0.31***	0.09**	-0.04	-0.02
Age 30-39 years old	0.32***	-0.10	-0.23	0.04	0.29***	-0.12	-0.17	0.03	0.04	-0.02	0.00	0.03
Age 40+ years old	0.14***	0.05*	-0.08	-0.10	0.10***	0.04	-0.06	-0.11	0.04	-0.03	-0.03	-0.04
Teenage mother	0.05*	0.04*	0.19***	-0.26	0.05	0.04	0.15***	-0.23	0.04	0.03	0.04	-0.04
Never married	0.05*	0.27***	-0.05	-0.30	0.02	0.24***	-0.06	-0.28	0.04	0.04	0.01	-0.02
Any child 0-5 years old	-0.55	0.08***	0.17***	0.23***	-0.47	0.11***	0.13***	0.21***	-0.05	-0.01	0.03	0.02
Any child 6-12 years old	0.45***	-0.14	-0.25	-0.01	0.40***	-0.15	-0.18	-0.05	0.07	-0.04	-0.02	0.04
2 children in household	0.03	-0.10	0.08**	0.01	0.02	-0.09	0.09***	0.01	0.00	-0.01	-0.02	0.03
3+ children in household	0.12***	0.00	-0.23	0.09***	0.12***	0.00	-0.17	0.04	-0.01	0.01	0.02	0.05
10th grade	-0.03	-0.01	-0.06	0.09***	-0.03	-0.01	-0.05	0.10***	-0.02	0.01	0.01	-0.01
11th grade	-0.13	0.14***	-0.05	0.00	-0.12	0.14***	-0.01	-0.05	-0.04	0.02	0.01	-0.01
Grade 12 or higher	0.16***	-0.03	0.08***	-0.17	0.16***	-0.03	0.04	-0.15	0.09*	0.00	-0.02	0.03
Highest degree = HS/GED	0.02	-0.11	0.06**	0.05*	0.05	-0.08	0.04	0.02	-0.02	-0.05	-0.04	0.00
Lives public/subss house	0.97***	-0.62	-0.29	0.10***	0.99***	-0.57	-0.22	-0.04	-0.05	0.02	0.00	0.05
1-2 moves in past 2 years	0.00	-0.02	0.06**	-0.04	-0.01	-0.04	0.11***	-0.04	0.00	0.05	0.04	0.02
3+ moves in past 2 years	-0.26	-0.31	0.34***	0.27***	-0.18	-0.21	0.23***	0.25***	-0.09	-0.04	0.00	0.07
On welfare < 2 years	-0.09	-0.20	0.25***	0.09***	-0.08	-0.15	0.19***	0.11***	-0.02	0.03	0.01	-0.03
On welfare for 2-5 years	-0.11	-0.12	0.05	0.18***	-0.11	-0.11	0.08**	0.18***	-0.06	-0.02	0.03	0.04
On welfare 5-10 years	0.05*	0.06**	-0.15	0.01	0.06**	0.04	-0.12	-0.01	0.02	-0.02	-0.02	0.00
On Welfare Q1 before RA	0.40***	0.19***	-0.33	-0.27	0.26***	0.05**	-0.20	-0.14	0.06	0.04	-0.05	-0.10
On Welfare Q2 before RA	0.39***	0.18***	-0.36	-0.23	0.28***	0.07***	-0.25	-0.14	0.03	0.03	-0.05	-0.06
On Welfare Q3 before RA	0.21***	0.23***	-0.28	-0.19	0.14***	0.15***	-0.21	-0.14	0.07	0.02	-0.02	-0.04
On Welfare Q4 before RA	-0.01	0.29***	-0.19	-0.15	-0.01	0.21***	-0.17	-0.12	0.05	0.00	-0.03	-0.04
On Welfare Q5 before RA	-0.04	0.33***	-0.17	-0.18	-0.03	0.25***	-0.15	-0.15	0.07	0.00	-0.02	-0.04
On Welfare Q6 before RA	-0.04	0.35***	-0.18	-0.20	-0.04	0.28***	-0.17	-0.18	0.07	0.00	0.01	-0.03
On Welfare Q7 before RA	-0.02	0.37***	-0.23	-0.19	-0.01	0.30***	-0.21	-0.18	0.08	0.03	0.00	-0.01
Rec. FS in Q1 before RA	0.28***	0.18***	-0.26	-0.21	0.18***	0.08***	-0.19	-0.11	0.05	0.03	-0.05	-0.09
Rec. FS in Q2 before RA	0.34***	0.14***	-0.34	-0.15	0.24***	0.06**	-0.23	-0.11	0.05	0.02	-0.04	-0.08
Rec. FS in Q3 before RA	0.26***	0.17***	-0.34	-0.12	0.20***	0.11***	-0.25	-0.11	0.07	0.00	-0.04	-0.06
Rec. FS in Q4 before RA	0.10***	0.22***	-0.25	-0.12	0.08***	0.16***	-0.22	-0.10	0.06	0.00	-0.02	-0.07
Rec. FS in Q5 before RA	0.06**	0.25***	-0.26	-0.11	0.06*	0.19***	-0.24	-0.09	0.05	-0.03	-0.02	-0.04
Rec. FS in Q6 before RA	0.05	0.25***	-0.26	-0.09	0.06*	0.20***	-0.23	-0.10	0.06	-0.03	0.00	-0.02
Rec. FS in Q7 before RA	0.05*	0.27***	-0.28	-0.10	0.05	0.22***	-0.24	-0.12	0.07	-0.01	-0.02	-0.01
Employed Q1 before RA	-0.11	-0.13	0.22***	0.04	-0.06	-0.07	0.19***	-0.01	-0.03	-0.06	0.01	0.03
Employed Q2 before RA	-0.14	-0.15	0.23***	0.09***	-0.07	-0.08	0.19***	0.01	-0.05	-0.06	0.01	0.02
Employed Q3 before RA	-0.10	-0.17	0.21***	0.09***	-0.04	-0.10	0.17***	0.03	-0.04	-0.04	-0.01	0.03
Employed Q4 before RA	-0.02	-0.18	0.21***	0.03	0.02	-0.12	0.16***	-0.01	-0.03	-0.05	0.02	-0.01
Employed Q5 before RA	0.03	-0.21	0.22***	0.02	0.05*	-0.16	0.18***	0.00	-0.03	-0.03	0.02	0.00
Employed Q6 before RA	0.05	-0.24	0.26***	0.00	0.07**	-0.18	0.20***	-0.03	0.00	-0.05	0.01	-0.02
Employed Q7 before RA	0.07**	-0.27	0.28***	0.00	0.10***	-0.23	0.24***	-0.02	-0.01	-0.06	0.02	-0.02
Employed Q8 before RA	0.07**	-0.31	0.32***	0.01	0.10***	-0.25	0.26***	-0.01	0.02	-0.06	0.01	-0.05
Emply at RA (self reported)	-0.06	-0.09	0.19***	-0.01	-0.03	-0.08	0.15***	0.00	-0.06	-0.01	0.03	0.04
Ever wrkd FT 6+ mths sm. job	0.22***	-0.52	0.02	0.37***	0.24***	-0.48	0.06*	0.33***	-0.06	-0.06	0.01	0.06
Earnings Q1 before RA	-0.08	-0.12	0.13***	0.09***	-0.02	-0.06	0.12***	0.00	0.00	-0.04	0.03	0.03
Earnings Q2 before RA	-0.12	-0.14	0.21***	0.08***	-0.04	-0.07	0.15***	-0.01	0.02	-0.04	0.03	0.05
Earnings Q3 before RA	-0.10	-0.15	0.22***	0.06**	-0.04	-0.07	0.14***	0.00	0.01	-0.03	0.00	0.05
Earnings Q4 before RA	0.02	-0.19	0.16***	0.05	0.05*	-0.13	0.13***	0.00	0.01	-0.02	0.02	0.02
Earnings Q5 before RA	0.08**	-0.19	0.14***	0.02	0.09***	-0.14	0.13***	-0.01	0.04	-0.04	0.01	-0.01
Earnings Q6 before RA	0.13***	-0.21	0.12***	0.02	0.13***	-0.16	0.11***	-0.01	0.04	-0.04	0.00	0.00
Earnings Q7 before RA	0.17***	-0.23	0.13***	0.00	0.15***	-0.20	0.13***	-0.01	0.03	-0.04	0.01	-0.02
Earnings Q8 before RA	0.17***	-0.24	0.13***	0.02	0.15***	-0.21	0.12***	0.01	0.03	-0.05	0.01	-0.01
Any earns yr before RA (slf-rep)	-0.21	-0.33	0.42***	0.18***	-0.11	-0.24	0.35***	0.10***	-0.09	-0.12	0.01	0.06

Notes: * Significant at 10%; ** Significant at 5%; **** Significant at 1%
Variables have been standardized to mean zero and standard deviation 1 (before imposing overlap)