10-2005

Net Impact Estimates for Services Provided through the Workforce Investment Act

Kevin M. Hollenbeck
W.E. Upjohn Institute for Employment Research, hollenbeck@upjohn.org

Christopher T. King
The University of Texas at Austin

Wei-Jang Huang
W.E. Upjohn Institute for Employment Research

Daniel Schroeder
The University of Texas at Austin

Net Impact Estimates for Services Provided through the Workforce Investment Act

by

Kevin Hollenbeck
Daniel Schroeder
Christopher T. King
Wei-Jang Huang

Prepared for:

Division of Research and Demonstration
Office of Policy and Research
Employment and Training Administration
U.S. Department of Labor

October 2005

This project has been funded, either wholly or in part, with Federal funds from the U.S. Department of Labor, Employment and Training Administration under Grant Number K-6558-8-00-80-60.

Employment and Training Administration Occasional Paper 2005-06

Table of Contents

List of Tables

List of Figures

Executive Summary

Researchers working under the auspices of the Administrative Data Research and Evaluation (ADARE) Project are conducting a series of research and evaluation projects examining participation in, services provided through, and outcomes from the Workforce Investment Act (WIA) and related federal programs in nine states. This paper reports on the net impacts, or "value-added," of WIA services on employment, earnings, and other outcomes of interest using administrative data from seven of the nine ADARE states: Florida, Georgia, Illinois, Maryland, Missouri, Texas and Washington State. The estimates produced in this assessment are arguably the most rigorous estimates of the net impact of WIA services to date.

We have relied on a *quasi-experimental approach* to estimate the net impacts of participation in WIA. The treatment group is comprised of individuals who received core, intensive, or training services through WIA and exited from the program during a particular period of analysis. The counterfactual is represented by a group of individuals that is carefully constructed from a population of nonparticipants. Because its members were not randomly chosen, the latter group is referred to as a *comparison group* rather than a control group.

We used three different statistical matching methods for each treatment for each state: weighted multivariate matching, nearest-neighbor propensity score matching, and propensity score "blocking." All three techniques rely on estimating each observation's propensity score, a statistic that estimates the probability that the observation will be in the treatment group. Weighted multivariate matching considers each treatment observation and attempts to find the "closest" comparison group pool observation by using an entire set of match variables. Nearest-neighbor matching is done by considering each treatment observation and finding the observation in the comparison group pool that has the closest propensity score. Propensity score "blocking" uses all of the observations in the comparison group pool, sorting the treatment group and comparison group pool by their propensity scores and partitioning each into subgroups (or "blocks") in such a way that each subgroup of the treatment group has a matching subgroup in the comparison group pool that is statistically indistinguishable from it. A wide range of variables was used in the propensity score regression and in the matching. Variables measured at registration included age at registration; disability status; race/ethnicity; veteran status; limited English proficiency status; education; employed at registration; and workforce area/region. Pre-registration variables measured included: employment; industry of employment; average earnings; earnings trend; earnings variance; percent of quarters with multiple employers; earnings 'dip' of 20% or more; quarters in which earnings 'dip' occurred; and percent of earnings that 'dip' represents.

The data used for estimating program impacts are administrative program records drawn from official state WIASRD files and ES records for Program Years 2000 and 2001, linked to Unemployment Insurance (UI) wage records for several years prior to entry into WIA or registration with ES and up to eight quarters after program exit and to TANF records. Such administrative data are generally more accurate and less expensive to obtain than primary data obtained by directly surveying participants and nonparticipants. We restricted the samples to those aged 22 through 64, and also trimmed (or deleted) the top 0.5% of earnings observations in all quarters, removing about 1.1% of the total participant sample.

Based on the net impact estimates presented in this paper, we conclude that WIA services as currently provided in these states are effective and appear to be doing a good job of addressing WIA's stated objectives. Moreover, the approach we have used to generate these estimates is likely to produce impact figures that are inherently conservative.

On average, we estimate that *receiving any WIA services* increases employment rates by about 10 percentage points and average quarterly earnings by about $800 (in 2000$). Furthermore, such services reduce participation in public assistance somewhat as well. All of these measured impacts are statistically significant.

The impacts of *receiving WIA training services* as compared to individuals who were served by WIA or the ES, but did not receive training services were also positive, but generally smaller in magnitude than for the receipt of any WIA services. Adult participants receiving training or referrals to training experienced statistically significant increases in employment of about 4.4 percentage points and in average post-exit earnings among employed adults of more than $660 per quarter and for employed dislocated workers of more than $380 per quarter. Again the range of impacts across states was wide; at least one state showed significant negative impacts on earnings.

The magnitudes of the treatment effects varied somewhat, but their significance and sign were largely consistent across states and population subgroups. The observed impact variation in part may reflect differing "bundles" of services offered by states. Some states allow local workforce boards, One-stop Centers, and service providers considerable leeway in bundling training with intensive services, whereas others do not.

While variation in the size of the impacts was apparent, the impacts for dislocated workers seemed to be consistently larger than those for adults. And, for both adults and dislocated workers, impacts for women were greater than for men, a finding that is largely consistent with the literature on training impacts. An examination of the time trend in outcomes suggests that the positive impacts persist over the first two post-exit years.

I. INTRODUCTION

Researchers working under the auspices of the Administrative Data Research and Evaluation (ADARE) Project are conducting a series of research and evaluation projects examining participation in, services provided through, and outcomes from the Workforce Investment Act (WIA) and related federal programs in nine states.[1] One of these efforts is an analysis of the net impacts of WIA services on employment, earnings, and other key outcomes of interest. Using data from seven of the nine participating ADARE states, the estimates produced in this assessment are arguably the most rigorous estimates of the net impact of WIA services that have been produced to date.[2]

The purpose of a net impact evaluation is to estimate the "value added" for program participants from receiving WIA services. Net impacts are defined as outcomes *net of* what would have occurred absent the intervention. For example, suppose an individual visits a One-stop Center under WIA, receives core and intensive services, and finds employment in a job that averages $5,200 each quarter. Further suppose that if this individual had not accessed WIA services, he or she would have found intermittent employment that averages $3,700 per quarter. The net impact of the WIA services on average quarterly earnings for this individual would be $1,500, the observed quarterly earnings of $5,200 less the $3,700 that he or she would have earned absent the receipt of WIA services.[3]

---

[1] For more information on the ADARE project, its activities and publications, visit the following website: www.ubalt.edu/jfi/adare/.

[2] States included in the net impact estimation are: Florida, Georgia, Illinois, Maryland, Missouri, Texas and Washington State. Data from the other two states currently participating in the ADARE consortium (i.e., California and Ohio) will be incorporated into future net impact estimates.

[3] This report, in general, and this example do not attempt to compute the benefit-to-cost ratio or return on investment from WIA services. Net impact estimates yield a major component of the benefits, but measuring and assigning costs of services to individual observations in the data was not within the scope of the effort. The positive net impacts that are reported in this paper may represent a substantial or meager (or even negative) return on investment; we simply don't know. But even if the net impacts were not statistically significant, we would not know whether the return on investment was large or small.

Net impacts must be *estimated* because individuals cannot be in two places at once—receiving WIA services and not receiving WIA services. This approach compares outcomes for individuals who have received WIA services and have exited from the program to individuals in a carefully constructed *comparison group* of similar individuals who did not participate. The comparisons are done separately by state using each state's administrative program data, which have been linked to Unemployment Insurance (UI) wage records. Then, meta-analytic techniques are used to transform the state-specific estimates into ones representative of the states as a whole.

Both the receipt of *any WIA services* and the *receipt of WIA training services* result in statistically significant, positive net impacts on the rate of employment and on average quarterly earnings after program exit for adults as well as dislocated workers. Furthermore, these interventions tend to reduce reliance on public assistance. These results hold for males and females, although impacts tend to be larger for females, consistent with findings from the training evaluation literature (e.g., Friedlander et al. 1997; King 2004; LaLonde 1995). Impact results are robust as well, varying little by estimation technique or by state.

The next section of this report presents a detailed discussion of our estimation methodology and data. That section is followed by a presentation of the results, which are presented separately for individuals who received any WIA services and for individuals who received WIA training services or were referred to training by ES staff (in two states). Results are shown for males and females and for dislocated workers and adults who are not dislocated workers. The results are followed by a brief section that presents conclusions and implications for workforce development programs, including WIA. A technical appendix and an appendix with supplemental results complete the report.

II. METHODOLOGY AND DATA

Determining the net impacts of a program's intervention requires answering two basic questions:

- What happened to those who participated or received program services?
- What would have happened to them had they not participated or received services?

The first question addresses outcomes for the "treatment group," members of whom have actually participated or received WIA services. The art to an evaluation is to devise a way to answer the second, much more difficult question, called the "counterfactual," because it is not observable. The most rigorous way to estimate net impacts is through a random assignment experiment. When this approach is feasible, individuals are randomly assigned to participate or not to participate. Net impacts are then estimated by comparing outcomes for the treatment group to outcomes for the individuals who were randomly excluded from treatment (called the control group.) The latter individuals' outcomes provide a reasonable counterfactual because there are no systematic (nonrandom) differences between the treatment and the control group that might explain differences in outcomes. Any net impacts that are observed may then be attributed to participation in the program.[4] An experimental approach was followed in the National JTPA Evaluation Study from 1986 to 1993 (see e.g., Bloom et al. 1997; Orr et al. 1995).

However, random assignment is not always feasible or desirable. For example, in WIA, core services are considered to be "universal," offered to anyone who might be

[4] In random assignment experiments, comparisons of outcome differences often use statistical adjustments to control for any differences in the treatment and control groups that might have occurred by chance.

interested in receiving them.[5] Random assignment would require withholding services from some individuals; services to which those individuals were entitled. In these cases, using random assignment raises ethical concerns, and researchers need to use a different methodology for constructing the counterfactual. ADARE researchers have relied on a *quasi-experimental approach* to estimate the net impacts of participation in WIA.[6] The treatment group is comprised of individuals who received core, intensive, or training services through WIA and exited from the program during a particular period. The counterfactual is represented by a group of individuals that is carefully constructed from a population of similar nonparticipants. Because its members were not randomly chosen, the latter group is referred to as a *comparison group* rather than a control group.

An important distinguishing characteristic of the ADARE study is that it relies solely on administrative data, including states' program records for individuals participating in and receiving workforce services, their associated employment and earnings information drawn from UI wage records and their associated welfare receipt data from state Temporary Assistance for Needy Families (TANF) program records. Such administrative data are arguably more accurate and less expensive to obtain than primary data obtained by directly surveying participants and nonparticipants.[7] Program administrators maintain management

[5] In fact, funding for WIA, including core services, is far below the level that would be required to actually *provide* universal service. Even when employment and training services were funded at much higher levels and made available mainly to economically disadvantaged individuals, only about 5 percent of eligibles could be served.

[6] Although some reviewers have suggested otherwise, we intentionally use the term "quasi-experimental" rather than "non-experimental" here. The latter suggests that the resulting estimates are no better than gross outcomes with no attempt to address the counterfactual. The former more accurately reflects the fact that researchers have produced impact estimates through the use of comparison groups meant to approximate control groups generated via random assignment. These estimates are thus better viewed as "quasi-" than "non-experimental."

[7] Hotz and Scholz (2001) review several studies that have attempted to calibrate the accuracy of administrative data, especially vis-à-vis survey data. That study is not directly germane to the present study since its focus was on poverty as measured by family income, and the net impacts estimated in this study are measured by individuals' earnings and employment. However, the study points out that some caution is in order with wage record data because of underreporting for flexible, "contingent" workers, and underreporting of tips. More

information system (MIS) data about individuals in order to operate their programs efficiently, and the evaluation economizes on data collection costs by using these data. (See Hollenbeck 2004 for conditions necessary to use program administrative data for evaluation purposes.)

The following sections document the primary aspects of the net impact estimation:

- definition of the treatment
- identification of the comparison group
- outcomes
- net impact estimation techniques
- data

Each is described in turn.

Definition of the Treatment

This evaluation has actually estimated net impacts for two treatments. The first treatment, which we refer to as *receiving any WIA services,* is operationally defined as being included in the U.S. Department of Labor's required WIA Standardized Record Data (WIASRD) reporting system for program years 2000 or 2001. To have a record in WIASRD, an individual must have data entered in his or her state's WIA administrative data system and must exit from, or be deemed to have exited from, the program. Due to the inherent difficulty of constructing reliable comparison groups and measuring impacts for youth, the analysis is limited to individuals aged 22–64 years who were adults or dislocated workers in WIA. The sample size for the seven states in the analysis was 92,787.

---

relevant are several earlier studies directly comparing UI wage-based with survey-based information on employment and earnings. Using data from the National JTPA Study, Kornfeld and Bloom (1997) found that "UI wage records and individual follow-up surveys produce similar impact estimates of program impacts on the earnings and employment of the same individuals. But, UI wage records are much less expensive…." (pp. 29–30). They are also easier to obtain, not subject to recall or response bias and can be accessed over long periods both pre- and post-program (see Baj et al. 1991);

It should be recognized that not all individuals who access WIA services are captured in the WIA administrative data. WIA is legislatively designed to offer services in a sequenced manner; participants first receive core services, and then, if they have not yet secured employment, they proceed to intensive services, and then possibly to training as well.[8] State variation seems to appear in whether WIA administrative data are maintained for individuals who received core services only. Local programs in some states also may "bundle" some training in with intensive services in order to avoid using newly required procedures for securing training from certified training providers under WIA.[9] WIASRD records from one of the seven states in the study, for example, have no observations for individuals receiving only core services. However, it appears that the typical practice is to enter data for individuals who were WIA participants as long as they received some mediated or staff-assisted (non-self help) services. All of the participating states maintained administrative data on all participants who received intensive services or training.

The second treatment, which we refer to as *receiving WIA training services,* is intended to focus the analysis on training services provided to the hardest to employ. Receiving training services is defined as being in the WIASRD and having received training services. Note that the ES administrative data included detailed service codes in two of the states. Approximately four percent of the ES records (after removing duplicate WIA registrants) in these states had received a referral to training. We decided to include these observations to the treatment sample for these states since the individuals received similar services. With the exception of these two states' ES records, the second treatment is a subset of the first treatment, so we expected the net impact results to be reasonably similar for both

[8] The sequence-of-service provisions are contained in WIA (Public Law 105-220), Section 134(d).
[9] For a discussion of these and related WIA data issues, see Barnow and King (2005).

treatments. The sample size for the second treatment was 53,436. That is, 30 to 50 percent of those contained in the WIASRD records received training services across all states. Again, the time period for receiving this "treatment" is program years 2000 and 2001, i.e., July 2000 through June 2002.

In addition to the two primary treatments discussed here, we also estimated net impacts for another treatment, receipt of WIA intensive *or* training services. This treatment is very similar to the first treatment; the only difference is that individuals who received only WIA core services are excluded from the treatment and included in the comparison pool instead. The impact results for this additional treatment are provided in Appendix B.

Identification of the Comparison Group

Each treatment has a slightly different comparison group. As noted above, the creation of comparison groups requires the identification of a population of individuals who did not receive the treatment, but who are quite similar to the individuals who did. The general methodology that was followed in this study was to identify a population of individuals who did not receive the treatment, which we refer to as the *comparison group pool,* and to select a subset of that population that closely matches the observed characteristics of the set of individuals in the treatment group. That matched set of observations comprises the *comparison group.*

For the treatment group comprised of individuals receiving any WIA services (treatment one), we have used the entire ES administrative dataset as the comparison group pool. That is, individuals who accessed the workforce development system through the ES,[10] but who were not also designated as WIA participants became the comparison group pool.

[10] In most areas, these individuals would likely have done so at a One-stop Career Center at which ES staff were co-located; in others, they might have done so at a stand-alone ES office.

This pool assumes that a reasonable counterfactual, the next best alternative absent the program, for WIA participants is registering with and being served through the ES.[11]

Using ES registrants as a comparison group pool allows us to err on the conservative side with our impact estimation. The matching methodology is intended to find observationally-equivalent individuals in the treatment group and comparison group pool. Other things being equal, individuals in the comparison group pool are likely to have more positive employment outcomes than those receiving services through WIA. The reason for this is local discretion at the One-stops as to whether a customer should be considered a WIA participant or a Wagner-Peyser Act (i.e., ES) registrant. There is anecdotal evidence that local One-stops tend to enroll more disadvantaged clients in WIA.[12] To the extent that this is true, on average, ES registrants would be more job-ready or have better labor force histories than WIA participants, which would bias the estimated net impacts downward to some extent. Thus, we can have more confidence in results that are positive and statistically significant.

We have been similarly conservative with constructing the comparison group pool for the second treatment, i.e., receiving WIA training services, as well. In this case, the comparison group pool consists of all of the WIASRD observations and, in two states, ES observations that did not receive or get referred to training. Because of the sequential nature of WIA services, a major reason for not receiving training is that an individual had a successful outcome as a result of receiving core or intensive services. Thus, we know that a

[11] One reviewer of an earlier version of this paper has suggested that private agency job search or self-initiated job search, rather than the ES, may be the next best alternative for WIA participants. From a practical standpoint, we have the administrative data for publicly funded workforce services. But more importantly, from a behavioral viewpoint, we believe that, since the ES is a public labor exchange, it would be frequented by individuals who would have participated in public job training through ES if WIA had not been accessible.
[12] A program administrator in Michigan indicated that the WIB in his area had identified WorkFirst clients or TANF recipients as highest priority; and he believed this to be a common practice in the state.

substantial share of the observations in the comparison group pool had positive employment outcomes, and it is likely that there is a bias toward better outcomes for the comparison group than for the treatment group. Again, for both treatments, we expect that our impact estimates are likely to err on the conservative side for these reasons.

For both treatments, we had reasonable grounds to believe that the observations in the comparison group pool were systematically different, on average, from those in the treatment group. In order to ameliorate systematic differences between these groups, we used a technique known as *statistical matching* to extract comparison groups from the comparison group pools. The next section briefly describes the matching procedures that were undertaken. The technical appendix provides more detail (see Appendix A).

Matching

The general problem that statistical matching is trying to solve is to find the observations in the comparison group pool that most closely resemble the observations in the treatment group. These observations become the comparison group. We used three different statistical matching methods for each treatment for each state: weighted multivariate matching, nearest-neighbor propensity score matching, and propensity score "blocking."

All three techniques rely on estimating each observation's propensity score. This statistic is essentially an estimate of the probability that the observation will be in the treatment group in a statistical model of being either in the treatment group or in the comparison group pool. The observations in the treatment group and in the comparison group pool are combined into a single sample. A dummy variable is created that takes on the value of one for observations in the treatment group and zero for the observations in the comparison group. A limited dependent variable estimation technique, such as logit regression, is used to try to "explain" the treatment group dummy. Then, the results from the regression are transformed into predicted probabilities, or propensities. The expectation is that observations from the treatment group will have propensity scores that are relatively large, on average, and that the observations from the comparison group will have propensity scores that are relatively low, on average.

Weighted multivariate matching considers each treatment observation and attempts to find the "closest" comparison group pool observation by using the entire set of match variables. "Closest" is defined as minimizing a weighted distance function; the weights are coefficients from the propensity score regression, and the distance is minimized through weighted least squares. It is essentially a nearest-neighbor algorithm, where nearest neighbor is calculated from the actual characteristics of the observations.

Nearest-neighbor propensity score matching is done by considering each treatment observation individually and finding the observation (or observations) in the comparison group pool that has the closest propensity score (see Dehejia and Wahba 1999, 2002).

Both weighted multivariate matching and nearest-neighbor propensity score matching can be done with or without replacement in the comparison group pool, with or without a caliper, and with one or more nearest-neighbors. Multivariate matching was done on a one-to-one basis, without replacement, and with a caliper that dropped ten percent of the matches. The nearest-neighbor propensity score matching was done on a one-to-one basis, with replacement, but without a caliper. (Details are provided in Appendix A.)

The *propensity score "blocking" technique* uses all of the observations in the comparison group pool.[13] That is, the comparison group and the comparison group pool are

[13] A few observations will, in general, be deleted. The procedure requires the minimum and maximum values of the propensity score to be identical in the treatment group and the comparison group pool. So observations in

identical. This technique sorts the treatment group and comparison group pool by their propensity scores. It then partitions each into subgroups (or "blocks") in such a way that each subgroup of the treatment group has a matching subgroup in the comparison group pool that is statistically indistinguishable from it based on observables.

The variables used in the propensity score regression and in the matching are as follows:[14]

Characteristics at Registration:

- Age at registration
- Disability status
- Race/ethnicity
- Veteran status
- Limited English proficiency status
- Education, Employed at registration
- Workforce area/region

Pre-registration:

- Employment
- Industry of employment
- Average earnings
- Earnings trend
- Earnings variance
- Percent of quarters with multiple employers
- Earnings 'dip' of 20% or more
- Quarter in which earnings 'dip' occurred
- Percent of earnings that 'dip' represents

The matched comparison group techniques build on earlier work estimating WIA impacts under the ADARE project (see Hollenbeck et al. 2003) and have taken advantage of methodological improvements suggested in the recent literature on estimation with statistical

---

the comparison group pool with propensity scores lower than the minimum in the treatment group are deleted from consideration. Concomitantly, observations in the treatment group with propensity scores greater that the maximum in the comparison group pool are deleted.

[14] Gender is not included in this list because matching was done separately by gender.

matching (e.g., Imbens 2004; Heckman and Navarro-Lozano 2004; Angrist and Hahn 2004; Michalopoulos et al. 2004; and Mueser et al. 2003).

Outcomes

Broadly stated, the purpose of WIA services in general and of training in particular is to increase the likelihood of being employed in a job with self-sustaining earnings. Implicit in this purpose is the objective of serving the employer community by increasing the availability of a productive workforce. These objectives are consistent with WIA's "dual-customer" focus: WIA views both job seekers and employers as "customers" to be served (e.g., see Barnow and King 2005). These broad outcomes encompass several concepts—becoming employed, wage rates, hours worked, retention in a job, growth of earnings over time, and reduced reliance on income maintenance support through public assistance.

Our methodological approach splits each observation's experiences into three general time periods: pre-registration, program participation, and post-exit. Registration is the point in time when an individual applies for ES or WIA services.[15] Estimating propensity score imputation equations and matching are done based on pre-registration or time-invariant demographic variables. After registration, individuals receive services (i.e., participate in various activities) offered by the ES or by the WIA program. After receiving services, the individuals exit from the ES or WIA.[16] The methodology attributes outcomes that occur after the exit date to program services.

In this paper, we measure impacts for three main outcomes: the employment rate, average earnings, and receipt of TANF benefits, all measured after exit from the program. In

[15] If an individual is referred to WIA by the ES, then registration occurs on the date that the individual is entered in the WIA system.

[16] The exact exit date is often difficult to determine because individuals may simply stop participating or showing up to receive services. A common practice is to assign a "soft" exit. If individuals have not been in contact with the agency (ES or WIA) for six months, then their exit date is set at the date of the last contact.

addition, Appendix B defines and presents net impacts for four of the "common measures" that were proposed as performance standards by OMB and have been adopted by USDOL's Employment and Training Administration for use in program year 2006 (see USDOL/ETA, 2005).[17] The three main outcomes are defined as follows:

> Employment Rate, defined as the percent of full post-exit quarters that an individual had UI wage-based earnings in excess of $100 per quarter (measured in 2000 dollars).
>
> Average Quarterly Earnings, defined as mean quarterly earnings for full post-exit quarters for those with earnings in excess of $100 per quarter (measured in 2000 dollars).
>
> Percent of Individual Months Receiving TANF, defined as the percent of full post-exit months that individuals were reported as receiving TANF benefits.

For employment and average earnings, the period of observation starts in the second full quarter after exit and proceeds up until the most recent quarter for which UI wage record earnings data are available (typically the fourth quarter of 2003, although it varies somewhat by state). Months receiving TANF begins in the first full month after exit and proceeds until the most recent month of available TANF data.

Net Impact Estimation Techniques

In the "gold standard" methodology of random assignment experimentation, the net impact can be estimated by subtracting the mean of the outcome variable(s) for the control group from the mean for the treatment group. Sometimes analysts will regression-adjust the difference in means in order to increase the precision of the net impact estimators. The regression model controls for any systematic differences between the treatment and control

[17] Using these outcomes is not intended as an endorsement of these measures for performance management or evaluation purposes. The purpose of this paper is not to weigh in one way or the other in the debate about appropriate performance measures.

groups that might have occurred simply by chance. The net impact estimation conducted in the quasi-experiments undertaken in this study is quite analogous.

For both the multivariate matching and nearest-neighbor propensity score matching, net impact estimates were calculated by computing regression-adjusted differences in the treatment and comparison group means. In addition, for purposes of testing robustness, the unadjusted differences in means were computed for the nearest-neighbor propensity score matched groups. The blocking technique produces a weighted difference in means; that is, the differences in means for each of the "blocks" are weighted by the proportion of the treatment group in the block and then summed. The standard errors for the two estimates using differences in means were computed using a bootstrapping technique with 100 replications.[18] Standard errors for the regression-adjusted estimates are traditional OLS standard errors.

As described in Appendix A, meta-analysis was used to combine the net impact estimates calculated for each state into a single combined estimate for the seven states. In the calculation of this single impact estimate, state-level results were weighted according to the sizes of their respective treatment populations, so that the single result is representative of the entire treated population across these states. In addition to this single impact estimate, the range of state-level impacts is presented in order to provide a sense of variation across the states.[19]

[18] As explained in Stine (1990), "bootstrapping" is a technique for estimating standard errors using large numbers of replications from a sample in situations involving unorthodox statistics or when the usual formulae for computing standard errors do not apply.

[19] In addition to greatly simplifying the presentation of results, this meta-analytic summary allowed us to comply with our existing state data sharing agreements, which preclude the release of state-specific results without the permission of official's for the particular state.

Data

The data used for estimating program impacts are administrative program records drawn from official state WIASRD files and ES records for Program Years 2000 and 2001, linked to Unemployment Insurance (UI) wage records for several years prior to entry into WIA or registration with ES and eight or more quarters after program exit. In addition, individual WIA records were also linked to TANF records maintained by the states.

The treatment and comparison datasets contained observations with an exit date sometime between July 1, 2000 and June 30, 2002. In the overwhelming number of cases, program services were received for less than a year. Registration dates preceded the exit dates by approximately a year or less.[20] However, there was no attempt to match observations based on registration dates, so it is entirely possible that matching observations had different pre-registration, program services, and outcomes over different periods of time. It was possible in the extreme, for example, for a treatment observation with a WIA registration date of January 2000 (who exited from WIA in the third calendar quarter of 2000) to be "matched" to a comparison observation with an ES registration date of January 2002 (who exited from the ES in the second calendar quarter of 2002). However, there are two methodological factors to keep in mind. First, the outcomes were defined relative to the exit quarter (and in constant dollars), so earnings records immediately after exit would have been used in both these cases. Second, no pairwise comparisons were calculated. The matching was used simply to identify sets of similar individuals.

We restricted the samples to those aged 22 through 64, and also trimmed (or deleted) the top 0.5% of earnings observations in all quarters. The result of the trimming was to

[20] WIA replaced JTPA in mid-1999 (for early implementation states) or 2000, and generally, the states officially "registered" clients at that time.

remove only about 1.1% of the total participant sample even though we had over a dozen quarters of earnings data because of the overlap in outlying earnings records.[21]

[21] Recent work (Bollinger and Chandra 2005) suggests that there may be situations when the common practice of data trimming may induce bias. Nevertheless, our perusal of the data spotted obvious errors and inconsistencies, especially extraneously large earnings values.

III. RESULTS

In this section, we first present the impact results for the key outcomes of interest for the receipt of any WIA service. We then proceed to examining the impact results for the receipt of training under WIA. In each case, we provide impacts for adults and dislocated workers for each of the three main outcomes of interest.

Impacts from Receiving Any WIA Service

The first treatment examines the impact of receiving *any WIA service*—whether core, intensive, or training services—on post-exit employment, earnings, and months on TANF. The hypotheses being tested are that individuals who receive WIA services will have higher rates of post-exit employment and higher average quarterly earnings, as well as fewer months on TANF due to better job search, increased employability, or even improved skills. The net impact analyses results are consistent with these hypotheses.

Post-exit Employment Rate. Table 1 shows that receiving any WIA services results in a statistically significant increase in employment rates of 8.7 percentage points for adults and 13.5 percentage points for dislocated workers.[22] Recall that the employment rate as defined here measures the share of post-exit quarters that individual's had UI wages above the minimum criterion of $100 per quarter. These impacts are large, and though they have a wide range, even the smallest state-level impact is positive and significant. For adults, the range of impacts is from almost six percentage points to over 12.5 percentage points, and for dislocated workers the range is from 10.5 to almost 18 percentage points.

[22] An individual was counted as employed in a quarter if he or she had at least $100 (2000$) in earnings in the quarter as reported in UI wage records data. The employment rate, on an individual basis, is the percentage of quarters after exiting the program (not counting the first post-exit quarter) in which the individual meets the minimum earnings criterion. The rate will range from 0.0 to 100.0.

In general, the estimated impacts are larger for dislocated workers than for adults. This probably stems from the fact that dislocated workers tend to have stronger, more stable work histories, and higher skills. Consequently, they are, on average, more employable. The estimated net employment impacts for dislocated workers are about 50 percent greater than for adults.

For both adults and dislocated workers, the employment impacts for women are much larger than for men. For adults, the difference is 4.4 percentage points, whereas it is 3.4 percentage points for dislocated workers.

Table 1. Employment Impacts of Any WIA Services
(Table entries are percentage points)

	Adults	Dislocated Workers
Overall Impact	8.7%** (0.45%, n = 100,764)	13.5%** (0.59%, n = 91,776)
Impact Range among States	5.7%** — 12.6%** (0.77% — 1.17%)	10.5%** — 17.9%** (2.01% — 1.61%)
Impact for Men	6.2%** (0.53%, n = 43,244)	11.8%** (0.71%, n = 46,310)
Impact for Women	10.6%** (0.70%, n = 57,520)	15.2%** (0.96%, n = 45,466)

Note: ** = p<0.01, * = p<0.05. Standard errors in parentheses, followed by total sample sizes. To protect state confidentiality, sample sizes not shown for range.

It is difficult to compare the employments impacts reported in Table 1 with those in the existing literature for two reasons. First, the definitions used for the employment outcome vary. The post-exit employment rate definition used in this analysis is closer to an employment "intensity" measure than an employment rate as such. It reflects the share of available post-exit quarters that individuals were working subject to a minimum earnings criterion. Second, most of the recent employment and training evaluations (e.g., Orr et al.

1995) and literature surveys (e.g., Friedlander et al. 1997, King 2004) stress the impacts on earnings not employment.

Post-exit Average Quarterly Earnings. As shown in Table 2, receiving any WIA services results in significant increases in average quarterly earnings of those employed (in constant 2000 dollars) as well: adults experienced a $743 quarterly post-exit earnings boost on average, while dislocated workers had an increase of $951, much larger than impacts for dislocated workers that have been reported in the literature to date.[23] Earnings impacts vary widely across the states—the range of estimates is over $1,000 for adults and $1,400 for dislocated workers—but, as for adults, all impacts are positive and statistically significant.

Table 2. Earnings Impacts of Any WIA Services

	Adults	Dislocated Workers
Overall Impact	$743** ($38, n = 98,074)	$951** ($47, n = 88,838)
Impact Range among States	$182** — $1,230** ($39 — $111)	$221 — $1,674** ($120 — $151)
Impact for Men	$685** ($51, n = 41,974)	$895** ($65, n = 44,648)
Impact for Women	$786** ($54, n = 56,100)	$1,008** ($68, n = 44,190)

Note: ** = p<0.01, * = p<0.05. Standard errors in parentheses, followed by total sample sizes.

As they did with the employment rates, dislocated workers exhibited more positive quarterly earnings gains than did adults. Their average quarterly earnings were on the order of 20 to 30 percent greater. While women experience larger earnings impacts than men, the differences are modest, whether for adults or dislocated workers.

On an annualized basis, assuming full-year employment, the estimated earnings impacts presented in the table translate into roughly a $2,800 annual earnings increase for

[23] See Friedlander et al. (1997), King (2004) and Lalonde (1995) for recent reviews of the literature on training impacts on earnings.

adult men and a $3,000 annual increase for adult women. For female dislocated workers, the estimated quarterly earnings impacts translate into gains of more than $4,000 per year.

Understanding the context for these estimated earnings impacts for adults and dislocated workers is important. Most experimental evaluations and literature surveys present both *per-assignee* and *per-enrollee* impact results. The former are smaller since some portion of those assigned to a given intervention never actually receive it. In the National JTPA Study, applicants were recommended for one of three "service strategies": classroom training; on-the-job training (OJT)/job search assistance (JSA); or other services. However, for the full treatment group sample, the shares actually enrolling in those strategies were just 60.4 percent in classroom training, 50.4 percent in OJT/JSA, and 61.2 percent in other services (see Orr et al. 1997, Exhibit 3.17, pp. 80-81). Some 22.9 percent of those assigned to classroom training and 41.8 percent of those assigned to OJT/JSA never enrolled in any JTPA services.[24] Quasi-experimental earnings impact estimates are more analogous to per-enrollee impacts, in that they represent comparisons between the earnings of those who were reported as receiving a particular service and those who were similar but did not receive such a service. Per-enrollee impacts can be derived by dividing the per-assignee impacts by the percentage of assignees that actually enrolled in the particular service.

With this distinction in mind, how do these earnings impact results for any WIA services compare for adults and dislocated workers? Converting the overall JTPA estimates across the three service strategies to 2000 dollars for comparability (King 2004, pp. 69-70), the per-enrollee impact of JTPA services on quarterly earnings was just over $300 for men and women alike, less than half as large as our estimated earnings impacts for receiving any WIA services. The same pattern holds for dislocated workers, though there are very few

[24] The comparable figure for the inclusive other services strategy is simply 38.8 percent, 100 percent minus the 61.2 percent enrolled in any service.

experimental evaluations to use for comparisons for this group. Men participating in the Texas Worker Adjustment Demonstration—which featured an experimental evaluation (see Bloom 1990)—experienced per-enrollee earnings gains only about $270 per quarter, while women gained about $460 per quarter per enrollee. Our impact estimates for WIA dislocated workers were more than three times as large for men and more than twice as large for women.

Post-exit Months on TANF. Receiving WIA services also resulted in a reduction in the number of post-exit months receiving TANF, as shown in Table 3. Adults exhibited a reduction of 2.6 percentage points, whereas the proportion of months on the rolls declined by 1.9 percentage points on average for dislocated workers. For males, the percentage point decline was smaller than for females, but since most adults on the rolls are females, the impacts in percentage terms are not that different. State variation for adults is not wide; however for dislocated workers, the states varied from a maximum impact of a reduction of 4.3 percentage points to a minimum of 0.0 percentage points (not statistically significant).

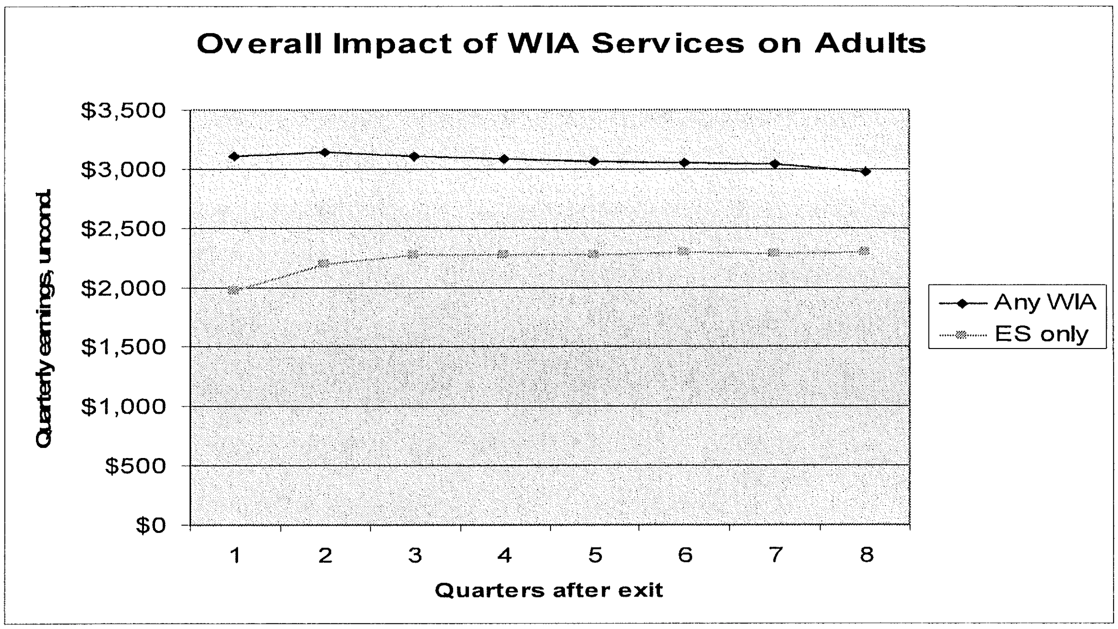
Table 3. Reduction in TANF Impacts of Any WIA Services
(Table entries are percentage points)

	Adults	Dislocated Workers
Overall Impact	−2.6%** (0.18%, n = 99,424)	−1.9%** (0.13%, n = 90,436)
Impact Range among States	−3.6%** — 2.3%** (0.33% — 0.28%)	−4.3%** — 0.0% (0.46% — 0.08%)
Impact for Men	−1.4%** (0.18%, n = 42,964)	−0.9%** (0.12%, n = 46,030)
Impact for Women	−3.5%** (0.30%, n = 56,460)	−2.9%** (0.24%, n = 44,406)

Note: ** = p<0.01, * = p<0.05. Standard errors in parentheses, followed by total sample sizes.

The impacts presented in the tables are averages for the particular period or point in time. But, what about the dynamics of these impacts, their pattern over time? Do they start out large and decay over time, or do they start small and grow larger? Figures 1 and 2 portray quarterly earnings impacts over eight (8) post-exit quarters for adults and dislocated workers receiving any WIA services.

Figure 1. Quarterly Earnings Impact of Any WIA Services on Adults

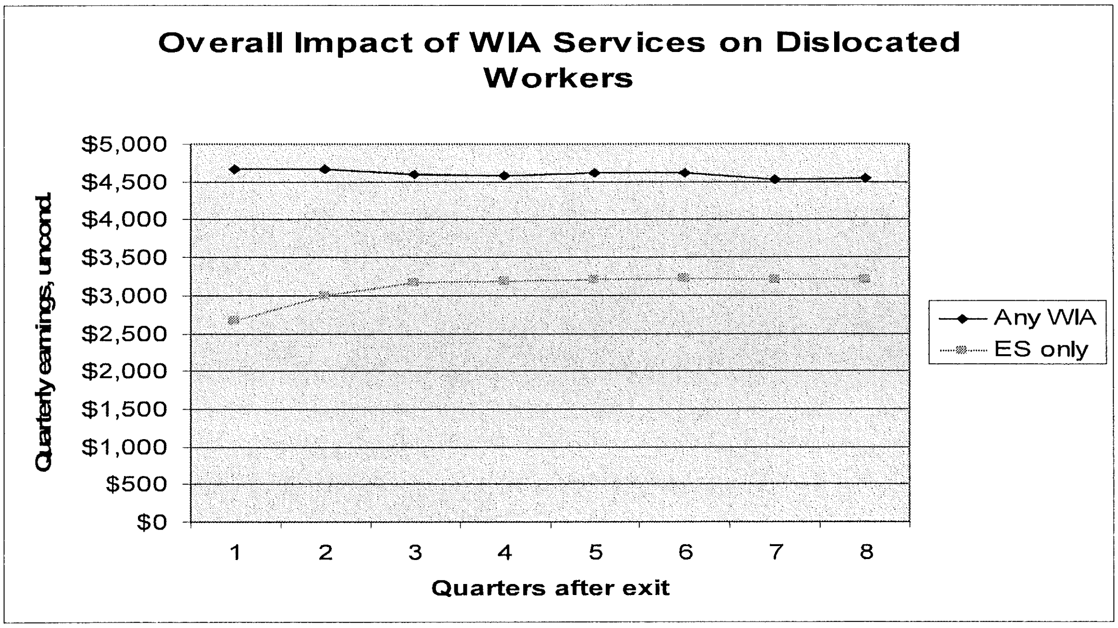


NOTE: Charted values are average unconditional earnings (which includes those with no earnings) of treatment group and matched control group, collapsed across gender.

The figure shows unconditional earnings, or average earnings whether or not employed, and as such reveals a combination of employment and earnings impacts. The pattern suggests that the impacts of receiving any WIA services were relatively stable over time for both adults and dislocated workers, decaying little over the two years after program exit.

Per-enrollee earnings impacts for JTPA adults are available for up to seven years following program exit (see King 2004, pp. 72–74). The pattern for women is similar to that

Figure 2. Quarterly Earnings Impact of Any WIA Services on Dislocated Workers



NOTE: Charted values are average unconditional earnings (which includes those with no earnings) of treatment group and matched control group, collapsed across gender.

described above: statistically significant impacts of around five percent that do not decay in the post-exit years; for men, the pattern is for much smaller (1 percent) earnings impacts per-enrollee that decay over time. There are no comparable figures for dislocated workers.

Impacts from Receiving Training under WIA

We next discuss the results for the second treatment—the impact of receiving training services under WIA . This treatment also yielded positive results for individuals relative to the counterfactual, although they are smaller in magnitude.

Post-exit Employment Rate. As Table 4 shows, adult participants receiving training or referrals to training experienced statistically significant increases in employment of about 4.4 percentage points. The range of impacts across states was relatively wide and includes zero (nonsignificant) impact in at least one state.

Table 4. Employment Impacts of Training
(Table entries are percentage points)

	Adults	Dislocated Workers
Overall Impact	4.4%** (0.40%, n = 54,754)	5.9%** (0.42%, n = 52,692)
Impact Range among States	−1.3% — 11.0%** (0.85% — 2.12%)	−1.3% — 11.0%** (4.32% — 2.12%)
Impact for Men	2.1%** (0.58%, n = 26,050)	5.0%** (0.51%, n = 29,188)
Impact for Women	6.5%** (0.57%, n = 28,704)	7.1%** (0.69%, n = 23,504)

Note: ** = $p<0.01$, * = $p<0.05$. Standard errors in parentheses, followed by total sample sizes.

The pattern of the results is very similar to the estimates for the first treatment—receipt of any WIA services—as would be expected from the overlap in treatment populations. The impacts on employment rates were larger for women (6.5 percentage points) than for men (2.1 percentage points). The impacts for dislocated workers were larger than for adults. For dislocated workers, the impact was 5.9 percentage points overall; 5 percentage points for men and 7.1 percentage points for women.

Post-exit Average Quarterly Earnings. As shown in Table 5, training (or training referrals) resulted in statistically significant increases in average post-exit earnings among employed adults of more than $660 per quarter and for employed dislocated workers of more than $380 per quarter. Again, the range of impacts across states was wide; at least one state showed significant negative impacts on earnings.

Interestingly, the earnings impacts for dislocated workers are smaller than for adults. This pattern is in contrast to the employment and earnings impacts for the first treatment (i.e., any WIA services) and for the employment impacts presented in Table 4. The implication is that dislocated workers who receive training or an ES training referral do find employment,

Table 5. Earnings Impacts of Training

	Adults	Dislocated Workers
Overall Impact	\$669** (\$50, n = 53,582)	\$386** (\$43, n = 51,078)
Impact Range among States	−\$260** — \$1,182** (\$85 — \$106)	−\$248** — \$1,245** (\$66 — \$112)
Impact for Men	\$552** (\$72, n = 25,478)	\$357** (\$78, n = 28,260)
Impact for Women	\$775** (\$70, n = 28,104)	\$422** (\$50, n = 22,818)

Note: ** = p<0.01, * = p<0.05. Standard errors in parentheses, followed by total sample sizes.

but their earnings lag somewhat. The pattern of larger estimated impacts for women than for men continues to hold for dislocated workers; average post-exit quarterly earnings are 40 to 50 percent greater for women than for men who are dislocated workers.

The estimated earnings impacts of receiving WIA training for men are comparable to the per-enrollee impacts estimated for the National JTPA Study when converted to quarterly 2000 dollars: while we estimate an earnings impact from WIA training of \$552 across the participating states, the JTPA impact range for men was about \$330–\$400 per quarter. Our impact estimates for WIA training tend to be considerably higher than those from the JTPA study: our estimated earnings impact for women receiving training is \$775 per quarter, compared to \$130–\$364 for women in JTPA. Note that the lower end of the JTPA range for men and women is provided by the estimated impact for classroom training, while the upper end of the range is the estimated impact for the OJT/JSA service strategy. Unfortunately, there are no reliable experimental estimates of the impact of training for dislocated workers with which to compare the current ones for WIA.

Post-exit Months on TANF. The training intervention also resulted in a reduction in the number of months receiving TANF as shown in Table 6. The impacts were smaller in magnitude than the reductions that resulted from any WIA services. Adults exhibited a reduction of 1.5 percentage points, whereas the percentage of months on the rolls declined by 1.0 percentage points on average for dislocated workers. For males, the percentage point decline (0.5 percentage points for adults and dislocated workers) was smaller than for females (2.4 percentage points and 1.6 percentage points for adults and dislocated workers, respectively), but again the impacts in percentage terms are not that different.

Table 6. Reduction in TANF Impacts of Training

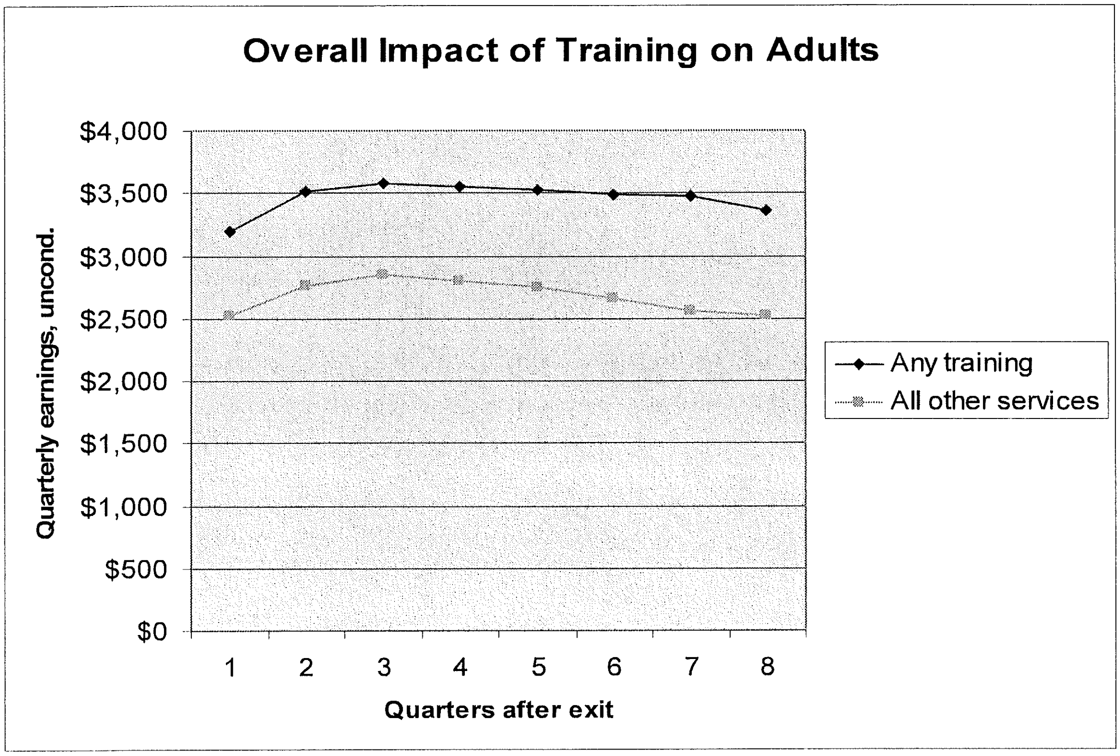
(Table entries are percentage points)

	Adults	Dislocated Workers
Overall Impact	−1.5%** (0.17%, n = 53,874)	−1.0%** (0.10%, n = 51,812)
Impact Range among States	−3.0%** — 0.5%** (0.27% — 0.10%)	−2.7%** — 0.2% (0.24% — 0.13%)
Impact for Men	−0.5%** (0.16%, n = 25,884)	−0.5%** (0.09%, n = 29,022)
Impact for Women	−2.4%** (0.26%, n = 27,990)	−1.6%** (0.20%, n = 22,790)

Note: ** = $p<0.01$, * = $p<0.05$. Standard errors in parentheses, followed by total sample sizes.

Figures 3 and 4 portray quarterly earnings impacts over eight (8) post-exit quarters for adults and dislocated workers. The figures suggest that while both participants and comparison group members had slightly improved earnings subsequent to the program, the positive impact of training persisted throughout the two-year period. The longer-term pattern of earnings impacts for adults receiving training under JTPA is similar for women but not for men (see King 2004, pp. 72–74). Most of the per-enrollee earnings impacts for women in JTPA were for those in the other services group, which actually posted growing impacts over

Figure 3. Quarterly Earnings Impact of Training on Adults

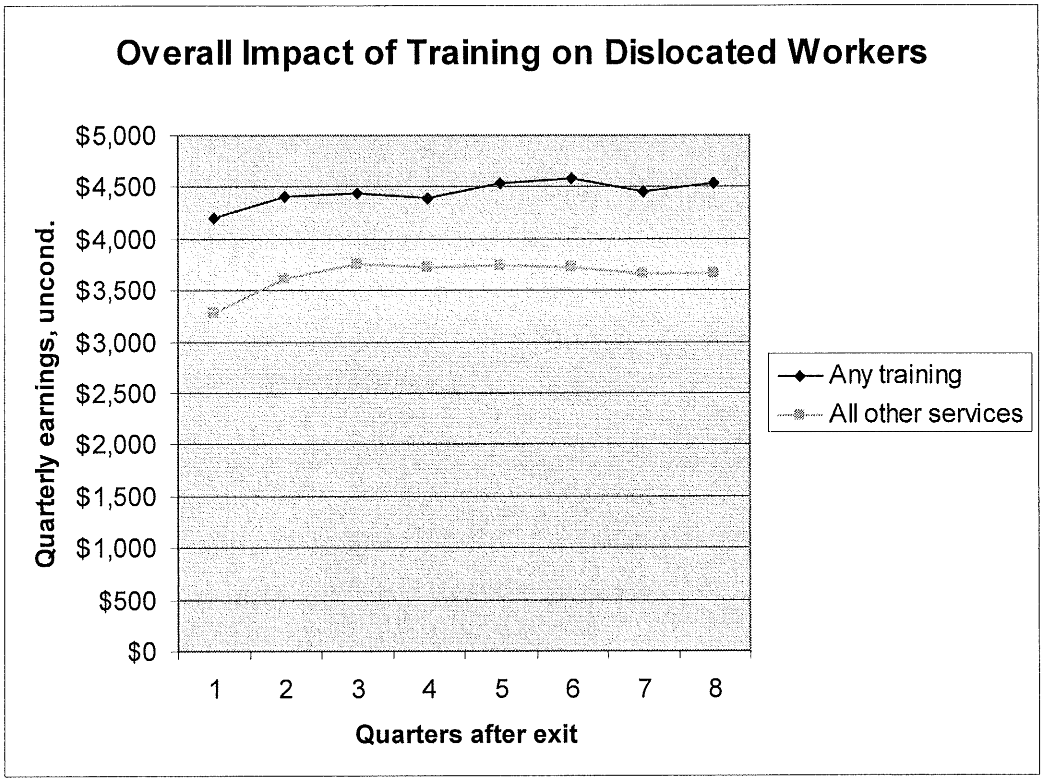


NOTE: Charted values are average unconditional earnings (which includes those with no earnings) of treatment group and matched control group, collapsed across gender.

the seven post-exit years. Adult women in the OJT/JSA and classroom training strategies experienced increasing gains for the first three post-exit years, but these diminished thereafter: impacts remained positive for those in OJT/JSA but disappeared altogether for those in classroom training.

Additional impact results are presented in Appendix B (Supplemental Results). In addition to the outcomes discussed here, we also estimated the impacts on the four common performance measures that have been proposed by OMB. Furthermore, we computed net impact estimates for another treatment, similar to the first.

Figure 4. Quarterly Earnings Impact of Training on Dislocated Workers



NOTE: Charted values are average unconditional earnings (which includes those with no earnings) of treatment group and matched control group, collapsed across gender.

IV. CONCLUSIONS AND IMPLICATIONS

Do WIA services add value for participants? Based on the net impact estimates presented in this paper, which are based on a rigorous, quasi-experimental methodology using multi-state data, we conclude that WIA services as currently provided in these states are effective and appear to be doing a good job of addressing WIA's stated goals and objectives. Moreover, as explained in the paper, the approach we have used to generate these estimates is likely to produce impact figures that are inherently conservative.

On average, we estimate that receiving any WIA services increases employment rates by about 10 percentage points and average quarterly earnings by about $800 (in 2000$). Furthermore, such services reduce participation in public assistance somewhat as well. All of these measured impacts are statistically significant. These net impacts were derived by comparing individuals who had received any WIA services to highly similar individuals who had received services provided by the ES.

Results are computed and presented here for a second comparison—receiving WIA training as compared to similar individuals who were served by WIA or the ES, but did not receive training services or referrals to training. These training impacts were also positive, but generally smaller in magnitude than for the receipt of any WIA services.

The magnitudes of the effects varied considerably, but the significance and sign of the effects were largely consistent across states and major population subgroups. The observed impact variation in part may reflect differing "bundles" of services offered by states. Some states allow local workforce boards, One-Stop Centers, and service providers considerable leeway in "bundling" training with intensive services, whereas others do not.

While variation in the size of the impacts was apparent, the impacts for dislocated workers seemed to be consistently larger than for adults. And, for both adults and dislocated workers, impacts for women were greater than for men, a finding that is largely consistent with the literature on training effects. Furthermore, an examination of the time trend in outcomes suggests that the positive impacts persist over the first two post-exit years.

From a methodological point of view, we observed that the estimate impacts were also quite consistent across methods of estimation. This robustness strengthens the confidence that we have in these results. That is, even though we used three different impact estimation methods—multivariate matching, propensity score matching, and blocking—the results were not materially different. Of course, should the service delivery context for WIA change in the future—in particular, moving away from the universal availability of unassisted core WIA services, applying experimental estimation techniques would certainly be desirable to validate and reinforce our quasi-experimental estimates. For now, these are the only impact estimates available for WIA.

We conclude that WIA services, including training, are effective interventions for adults and dislocated workers, when measured in terms of net impacts on employment, earnings, and receipt of TANF for participants. They yield positive results for men and women and for both the adult and dislocated worker populations. The impacts tend to be greater for dislocated workers, and results tend to be generally better for women than for men.

We further conclude that the quasi-experimental evaluation approach using linked administrative records offer policymakers attractive opportunities for estimating the impact of WIA and related services at relatively low cost. Maintaining these opportunities is going to require additional effort and resources. Administrators at all levels should strive to improve the quality and accessibility of these data, while ensuring the appropriate privacy and confidentiality protections. Resources should to be allocated to accomplish these tasks in the future, as well as to perform benefit/cost analyses of WIA services.

Appendix A

Technical Details

The Net Impact Evaluation Problem

The net impact evaluation problem may be stated as follows: Individual i, who has characteristics X_{it} at time t will be observed to have outcome(s) $Y_{it}(1)$ if he or she receives a "treatment," such as receiving WIA services or being referred to training, and will be observed to have outcome(s) $Y_{it}(0)$ if he or she doesn't participate. The net impact of the treatment for individual i is $Y_{it}(1) - Y_{it}(0)$. Of course this difference is never observed because an individual cannot simultaneously receive and not receive the treatment.

To simplify the notation without loss of generality, the time subscript is omitted in the following discussion. Let $W_i = 1$ if individual i receives the treatment, and $W_i = 0$ if i does not receive the treatment. Let T represent a data set with observations about individuals who receive the treatment and let n_T represent the number of individuals with data in T. This data set is referred to as the Treatment group. Let U represent a data set with observations about individuals who may be similar to individuals who received the treatment and let n_U be its sample size. This data set is referred to as the Comparison Group pool. In the matching techniques described below, a subset of U is identified that contains observations that "match" those in T. This subset is C, and n_C is its sample size. It is referred to as the Comparison sample.

Being in the treatment group is assumed to be a random event—individuals happened to be in the right place at the right time to learn about the program, or the individuals may have experienced randomly the eligibility criteria for the program—so W_i is a stochastic outcome that can be represented as follows:

(1) $W_i = g(X_i, e_i)$, where

e_i is a random variable that includes unobserved or unobservable characteristics about individual i as well as a purely random component.

An assumption that is made about $g(.)$ is that $0 < \text{prob}(W_i = 1|X_i) < 1$. This is referred to as the "support" or "overlap" condition that is necessary so that the outcome functions described below are defined for all X.[25]

In general, outcomes are also assumed to be stochastically generated. As individuals in the treatment group encounter the treatment, they gain certain skills and knowledge and encounter certain networks of individuals. We characterize their outcomes with the following mapping:

(2) $Y_i(1) = f_1(X_i) + e_{1i}$

Individuals not in the treatment group progress through time and also achieve certain outcomes according to another stochastic process, as follows:

(3) $Y_i(0) = f_0(X_i) + e_{0i}$

Let $f_k(X_i) = E(Y_i(k)|X_i)$, so e_{ki} are deviations from expected values that reflect unobserved or unobservable characteristics, for $k = 0,1$.

As mentioned, the problem is that $Y_i(1)$ and $Y_i(0)$ are never observed simultaneously. What is observed is the following:

(4) $Y_i = (1 - W_i)Y_i(0) + W_iY_i(1)$

The expected value for the net impact of the treatment on the sample of individuals treated:

(5)
$$\begin{aligned} E[Y_i(1) - Y_i(0)|X, W_i = 1] &= E(\Delta Y \mid X, W = 1) \\ &= E[Y(1)|X, W = 1] - E[Y(0)|X, W = 0] \\ &\quad + E[Y(0)|X, W = 0] - E[Y(0)|X, W = 1] \\ &= \hat{f}_1(X) - \hat{f}_0(X) + \text{BIAS}, \end{aligned}$$
where

$\hat{f}_k(X)$, $k = 1, 0$, are the outcome means for the treatment and comparison group samples, respectively, and

[25] Note that Imbens (2004) citing Heckman, Ichimura, and Todd (1997) shows that this condition can be slightly weakened to $\Pr(W_i = 1|X_i) < 1$ if the outcome of interest is average treatment effect on the treated.

BIAS represents the expected difference in the $Y(0)$ outcome between the comparison group (actually observed) and the treatment group (the counterfactual.)

A key assumption that allows estimation of equation (5) is that $Y(0) \perp W|X$. This orthogonality assumption states that given X, the outcome (absent the treatment), $Y(0)$, is random whether or not the individual is a participant. This is equivalent to the assumption that participation in the treatment can be explained by X up to a random error term. The assumption is called "unconfoundedness," "conditional independence," or "selection on observables." If the assumption holds, then the net impact is identified because BIAS goes to 0, or

(6) $$E\left[\Delta Y \middle| X, W = 1\right] = \hat{f}_1(X) - \hat{f}_0(X)$$

In random assignment, the X and W are uncorrelated through experimental control, so the conditional independence assumption holds by design. In any other design, the conditional independence is an empirical question. Note that because the orthogonality assumption holds asymptotically (or for very large samples), in practice, it may make sense to regression adjust equation (6) whether or not the data come from a random assignment experiment.

In our case, experimental data are unavailable. Instead, we have the WIASRD data set that contains information about individuals who have encountered WIA (the T data set) and we have ES data that contains information about individuals who may comprise a comparison group for the treatment cases (the U data set).

Various estimation techniques have been suggested in the literature, but they may be boiled down to two possibilities: (1) use all of the U set or (2) try to find observations in U that closely match observations in T. The "blocking" technique is one of the former techniques and multivariate matching and propensity score matching are among the latter.

Participation Model and Propensity Scores Imputation

All three techniques rely on the estimation of a participation model (equation 1) and two of the techniques require observation-by-observation imputation of propensity scores. We have used a logit to estimate the equation as in (7).

(7) $\text{prob}\,(W_i = 1) = \Lambda\,(B\,X_i),\ i$ where

X_i = i-the observation's values for the vector of common variables in T and U
Λ = logistic cumulative distribution function
B = parameters to be estimated.

The imputed propensity score is the predicted value for each observation. That is,

(8) $p_i^* = \Lambda\left(\hat{B}\,X_i\right),$ where

$\hat{B}$ = estimated logit parameters.

With a few exceptions, noted here, the dimensions listed in Table A-1 were used in the logistic regression predicting treatment group membership. In Georgia, the veteran, limited-English proficiency (LEP), disability status indicators, and several categories of industry codes were omitted from the matching dimensions due to unavailable data elements and very small numbers in some of these categories. In Texas, an 'education unknown' category was added to account for missing data on the Educational attainment indicators. In Maryland and Illinois, the LEP indicator was omitted. And, in Maryland only, another indicator for unemployment claimant or exhaustee was added to the basic set.

Table A-1: Baseline Dimensions for Matching, and Regressors for Statistical Adjustment of Impacts

Dimension	Description and rationale
Local board code	As reported in WIASRD item 301 or in ES records
Age at registration, years	Based on birth date and WIA or ES registration date (WIASRD items 102 and 302 or ES records).
Gender	Binary based on WIASRD item 103 or ES records: 1=female, 0=male.
Disability	Binary based on WIASRD item 104 or ES records: 1=yes (any), 0=no.
Eth. White	Binary based on WIASRD item 110 or ES records: 1=yes, 0=no. (omitted level: non-white)
Veteran	Binary based on WIASRD item 111 or ES records: 1=yes (any), 0=no.
Employed at registration	Binary based on WIASRD item 115 or ES records: 1=yes, 0=no.
Limited English	Binary based on WIASRD item 116 or ES records: 1=yes, 0=no.
Unemployment compensation claimant or exhaustee	Binary based on WIASRD item 118 or ES records: 1=yes, 0=no.
Education less than high school	Education binary based on WIASRD item 123 or ES records, highest grade completed. 1=less than high school graduate or GED, 0=greater.
Education high school graduate	(omitted category). Education binary based on WIASRD item 123 or ES records. 1= high school graduate or GED, 0=lesser or greater
Education beyond high school	Education binary based on WIASRD item 123 or ES records. 1=some college or greater, 0=less.
Pre-registration employment measures:	
Employment rate	Percent of quarters employed in pre-registration quarters 3-8, beginning with first employment in pre-registration interval. Employment was defined as receipt of at least $100 in a quarter.
Conditional earnings	Average earnings in pre-registration quarters 3-8, of those quarters in which employed.
Earnings trend	Linear trend in earnings in pre-registration quarters 3-8.
Earnings variation	Coefficient of variation of earnings in pre-registration quarters 3-8.
Turnover	Average number of employers per quarter in pre-registration quarters 3-8.
Earnings dip measures:	
Had an Earnings dip	Binary=1 if, across possible pre-post comparisons in the 8 quarters prior to WIA or ES entry, the largest pre-post average difference is greater than 20% of the pre-dip average. Otherwise=0.
Quarters before registration in which dip occurred	If had an earnings dip, number of quarters prior to registration in which the dip occurred. Otherwise=0.
Percent of earnings the dip represents	If had an earnings dip, percent of pre-dip earnings the dip represents (minimum 20%). Otherwise=0.
Industry of Employment	For last job prior to WIA or ES registration, according to UI records, the 1-digit SIC code of employer industry. Zero if missing or unknown.

Matching Techniques

In the multivariate matching and in the propensity score nearest-neighbor matching, U denotes the set of observations from which is chosen a subset C (for matched comparison group) that will be used in the net impact analyses. The idea is to have C be comprised of the observations where individuals are most 'like' the individuals comprising T. Matching adds a whole new layer of complexity to the net impact estimation problem. The estimator becomes a function of how the match is done in addition to the characteristics of the sample. Since the matching process is a structured algorithm specified by the analyst, the statistical error associated with the net impact estimator now includes a component that may be identified as matching error in addition to the sampling error and model specification error.

Nearest-neighbor algorithms minimize a distance metric between observations in T and U. If we let X represent the vector of variables that are common to both T and U, and let X_j, X_k be the values of X taken on by the jth observation in T and kth observation in U, then C will be comprised of the k observations in U that minimize a distance metric for all j. If the matching is done without replacement, then when an observation in U is found to be a match, that observation is deleted from consideration in all subsequent matches.

The literature usually suggests that the distance metric be a weighted least squares distance, $(X_j - X_k)\mathrm{N}\Sigma^{\mathrm{B}1} (X_j - X_k)$, where $\Sigma^{\mathrm{B}1}$ is the inverse of the covariance matrix of X in the comparison sample. This is called the Mahalanobis metric. If we assume that the X_j are uncorrelated, then this metric simply becomes least squared error. Imbens (2004) has a discussion of the effect of using different metrics, although in practice the Mahalanobis metric is used most often. The multivariate matching technique used in this study departs from this standard metric by weighting the distance summation using the absolute values of the coefficients from the logit, i.e., the $\hat{B}$.[26]

The second type of matching done in this study uses the imputed propensity scores, p_i^* (see Dehejia and Wahba 2002). Rosenbaum and Rubin (1983) showed that the conditional independence assumption, $Y(0) \perp W|X$ implies that $Y(0) \perp W|g(X)$, where $g(X)$ is the conditional probability of receiving the treatment = Prob(W = 1|X). This result implies that the observation-by-observation matching can be done, at considerably reduced dimensionality, with the X_i replaced by p_i^*. Treatment observations are matched to observations in the comparison sample with the closest propensity scores. In this case, the distance metric that was used was the absolute value of the differences in propensity scores, and the matching was done with replacement, so that all observations in U were considered as match candidates for all observations in T.

The whole reason for matching is to find similar observations in the comparison group to those in the treatment group when the 'overlap' or statistical support is weak. Consequently, the nearest-neighbor approaches may be adjusted to require that the distance between the observations that are paired be less than some criterion distance. This is called *caliper or radii matching*. This was done for both matching techniques.

Once the Comparison set C has been constructed, the net impact(s) can be estimated by taking simple differences in means such as in equation (9).

(9) $$\tau = \frac{1}{n_T}\sum_{i \in T} Y_1(1) - \frac{1}{n_C}\sum_{i \in C} Y_j(0),$$ where

τ = net impact.

[26] This procedure is also followed in Zhao (2004).

We computed this estimator for the propensity score nearest-neighbor approach. However, we have been concerned about small sample sizes in some of the states, so for both matching approaches, we estimated all the net impacts through regression adjustment, as in (10) using a pooled T and C dataset.

(10) $Y_i = a + B'X_i + \tau W_i + e_i.$

Blocking

The blocking algorithm is a full sample estimator (it uses all of the observations in U and does not rely on a matched subset C) (see Dehejia and Wahba 2002). The intuition here is to partition the treatment and full sample into "blocks" or strata by propensity score, such that there is no statistical difference between the covariates, X, in each block. This essentially achieves the conditional independence assumption locally in each block. Then the average treatment effect is a weighted average of the treatment effects in each block.

The algorithm proceeds by first deleting some observations that have outlying imputed propensity scores. In particular, all observations in the Comparison Group pool, U, with propensity scores less than the minimum propensity score in the Treatment sample T, are deleted. Similarly, all observations in T with propensity scores greater than the maximum in U are deleted. Let N be the number of observations remaining in T.

The remaining observations in U are partitioned into deciles. The observations in T are also divided in 10 groups using the decile values from U. (Of course, the 10 groups are not of equal size). A joint F-test is computed for a key set of characteristics in the "matching" subgroups to determine whether they are (locally) statistically indistinguishable. If the F-test fails for any of the subgroups (blocks), i.e., the two groups are found to be different in at least one dimension, then these blocks are further subdivided and re-tested. The procedure continues until all blocks are independent.

Assume there are K blocks. Let the kth block be defined as all treatment or full comparison sample cases with values of X such that $p(X) \in [p_{1k}, p_{2k}]$. Let NT_k be the number of treatment cases and NU_k be the number of comparison cases from the full sample in the kth block. The treatment effect with each block k is as follows:

$$\text{(11)} \qquad \tau_k = \sum_{\substack{i=1 \\ i \in T}}^{NT_k} \frac{1}{NT_k} Y_i(1) - \sum_{\substack{j=1 \\ j \in U}}^{NU_k} \frac{1}{NU_k} Y_j(0)$$

and the overall estimated average treatment effect is given as follows:

$$\text{(12)} \qquad \tau = \sum_{k=1}^{K} \frac{NT_k}{N} \tau_k$$

Meta-Analysis

Meta-analysis refers to a collection of techniques for the statistical analysis of a large set of analytic results for purposes of integrating and summarizing the findings. WIA net impact estimates were consolidated across matching methods in order to provide a uniform answer to each research question. Similarly, impacts were consolidated across states in order to characterize an effect that generalizes across vastly different policy and demographic environments. Consolidating impacts across states avoids the lengthy, and sometimes difficult, process of obtaining clearance for state-level impact estimates from state officials who voluntarily allowed the use of their data for this study.

Consolidation of impacts was done at several levels. First, the impacts were consolidated across the four estimation methods. Next, using different techniques that are more appropriate for independent tests, impacts were consolidated variously across states, across gender, and across adult/dislocated worker status. These techniques are described below.

Across estimation methods. Unlike the procedures to follow, consolidation of impacts across estimation methods was quite straightforward. Since the use of up to four distinct but related estimation methods on one comparison should be considered to represent four *non-independent* tests, the goal in the first level of consolidation was not to combine the impacts meta-analytically, but to select one impact estimate as representative of the group of estimation methods. For this purpose, the net impact values themselves were averaged across estimation methods. Likewise, the statistical significance of this across-method impact estimate was computed by first converting the p-values associated with the four statistical tests into their corresponding standard normal deviates, or z-scores, selecting the median of these z-scores, and converting back to a single p-value.[27] Finally, because the sample sizes vary among the estimation methods, the sample size for this across-method estimate was computed as the median of the four sample sizes.[28] These three consolidated statistics, the typical impact estimate, statistical significance (or p-value), and sample size, were chosen to represent the impacts for each subgroup. Thus, for each combination of comparison and dependent variable, one consolidated impact estimate was produced for each state (7) by gender (2) by adult/dislocated worker status (2) combination. These were further consolidated as described below.

Across sub-populations. Further consolidation of impact estimates across various sub-populations, including state, gender, and adult/dislocated worker status, differs from the

[27] When an even number of tests are summarized this way, the median is computed as the average of the two values above and below the midpoint of the distribution. This averaging is not appropriate to use on the p-values themselves, thus necessitating the conversion of p-values to z-scores first.
[28] This 'typical' sample size was not used for determination of statistical significance at this step, rather it was used only for purposes of further consolidation of results described in the next section.

consolidation described above primarily due to the fact that these are now *independent* tests of the same hypothesis, differing only in the sub-populations to which they apply. Meta-analytic combination of effects and combined statistical significance of such effects for *independent* tests has a long history in the statistical and social sciences literatures. It consists of two parts: 1) calculation of weighted average impacts, and 2) computation of a combined statistical test.

Meta-analytic calculation of weighted average impacts is greatly simplified in this study due to the fact that the impact estimates being combined are all measured on the same scale (e.g., quarterly earnings, or percent employed). This avoids the complex issue of choosing an effect size estimator; instead, all effects are allowed to remain in their original units. As suggested by Mosteller and Bush (1954), we weight the effect estimates by their sample sizes so that the comparisons based on greater statistical power carry greater weight in the combined estimate. Furthermore, since this study is done with populations rather than samples, this weighting procedure carries the added benefit of making the combined results representative of the combined populations of interest in the ADARE states. For example, if state A has 10 times the dislocated workers as state B, and their impact estimates differ, the final impact estimate for dislocated workers combined across states will much more closely resemble the impact estimate for state A than for state B. Weighted average impacts across subpopulations are computed by multiplying the impact by the sample size for each subpopulation, summing these values across subpopulations, and dividing by the total sample size across subpopulations.

Meta-analytic combination of the results of independent *significance* tests essentially amounts to a test of the significance of the combined effect, as if the independent samples had been collapsed. Although there are numerous methods of computing this joint test, we utilize the method of Stouffer (described in Wolf, 1986) because it takes into account both the significance and *direction* of the tests being summarized. In other words, if several of the tests to be combined are significant but with effects going in opposite directions, this test might appropriately report the combined effect to be non-significant, whereas other methods may not make this distinction. The Stouffer test is based on the fact that the sum of normal deviates is itself a normal deviate, with variance equal to the number of observations summed. The test involves converting the p-values associated with the tests to be summarized into standard normal deviates, or z-scores.[29] These z-scores are then summed, and the total is divided by the square root of the number of tests to yield a combined z score, the statistical significance of which can be obtained from a z-table.[30]

[29] The directionality of the individual findings is preserved by assigning negative z values for those tests whose impacts are in the opposite direction.

[30] Critical values of z, using a two-tailed test, are 1.96 for significance at the 0.05 level, and 2.56 for significance at the 0.01 level

Appendix B

Supplemental Results

This appendix presents two sets of supplemental results. Table B-1 provides net impact estimates for the four common measures proposed by OMB for the two treatments described in the text—receiving any WIA services and receiving training (through WIA or ES referral). Table B-2 provides net impact estimates for all seven outcome variables for a treatment that is of practical and theoretical interest but not presented in the text: receiving WIA intensive *or* training services. This treatment is very similar to the first treatment presented in the text. The only difference is that individuals who received only WIA core services are excluded from the Treatment and included in the Comparison Sample pool instead.

Table B-1: Net Impacts on OMB Common Measures of Receiving WIA Services and Receiving Training

	Adults	Dislocated Workers	Adults	Dislocated Workers
PANEL 1: TREATMENT IS RECEIVING WIA SERVICES				
	Rate of Job entry		**Rate of job retention**	
Overall Impact	0.19** (0.009, n = 97,442)	0.24** (0.010, n = 89,782)	0.06** (0.004, n = 93,434)	0.08** (0.004, n = 85,072)
Impact Range among States	0.13** – 0.28** (0.030 – 0.025)	0.13** – 0.42** (0.030 – 0.038)	0.03** – 0.11** (0.011 – 0.014)	0.03 – 0.11** (0.036 – 0.010)
Impact for Men	0.15** (0.009, n = 42,258)	0.21** (0.012, n = 45,414)	0.05** (0.006, n = 40,208)	0.07** (0.006, n = 42,540)
Impact for Women	0.22** (0.014, n = 55,184)	0.27** (0.017, n = 44,368)	0.06** (0.006, n = 53,226)	0.09** (0.007, n = 42,532)
	Pre-to-Post earnings change		**Post-reemployment earnings change**	
Overall Impact	27.0%** (1.30%, n = 94,442)	21.7%** (1.16%, n = 85,306)	−4.5%** (0.65%, n = 94,434)	−2.5% (3.35%, n = 85,296)
Impact Range among States	12.6%** – 37.3%** (1.28% – 3.38%)	8.1%** – 42.2%** (1.85% – 3.80%)	−11.6%** – −1.2% (2.37% – 1.32%)	−25.8%** – 3.7%** (2.91% – 1.37%)
Impact for Men	26.8%** (1.79%, n = 40,494)	23.4%** (1.79%, n = 42,774)	−5.4%** (1.14%, n = 40,486)	−1.9% (46.48%, n = 42,770)
Impact for Women	27.2%** (1.82%, n = 53,948)	20.0%** (1.48%, n = 42,532)	−3.9%** (0.75%, n = 53,948)	−3.0% (2.65%, n = 42,526)
PANEL 2: TREATMENT IS RECEIVING TRAINING				
	Adults		**Rate of job retention**	
Overall Impact	0.09** (0.007, n = 52,944)	0.12** (0.008, n = 51,586)	0.04** (0.004, n = 52,276)	0.04** (0.004, n = 49,142)
Impact Range among States	−0.03** – 0.18** (0.011 – 0.037)	−0.04* – 0.26** (0.018 – 0.023)	−0.02 – 0.07** (0.025 – 0.009)	−0.02* – 0.08** (0.012 – 0.010)
Impact for Men	0.05** (0.010, n = 25,518)	0.10** (0.010, n = 28,680)	0.02** (0.006, n = 24,948)	0.03** (0.005, n = 27,222)
Impact for Women	0.12** (0.011, n = 27,426)	0.14** (0.012, n = 22,906)	0.05** (0.006, n = 27,328)	0.05** (0.006, n = 21,920)
	Pre-to-Post earnings change		**Post-reemployment earnings change**	
Overall Impact	22.2%** (1.33%, n = 52,276)	10.9%** (1.17%, n = 49,142)	−0.4% (0.83%, n = 52,180)	−1.6% (1.30%, n = 49036)
Impact Range among States	4.1%* – 38.1%** (1.65% – 3.42%)	1.0% – 30.8%** (1.15% – 2.85%)	−6.7%* – 1.6% (2.65% – 2.51%)	−12.1%** – 1.6% (3.19% – 2.70%)
Impact for Men	18.5%** (1.74%, n = 24,948)	11.3%** (1.83%, n = 27,222)	−0.6% (3.29%, n = 24,906)	−2.1% (1.40%, n = 27,176)
Impact for Women	25.5%** (1.98%, n = 27,328)	10.4%** (1.47%, n = 21,920)	−0.2% (0.40%, n = 27,274)	−0.9% (13.69%, n = 21,860)

Note: ** = p<0.01, * = p<0.05. Standard errors in parentheses, followed by total sample sizes. To protect state confidentiality, sample sizes not shown for range.

Table B-2: Net Impacts of Receiving WIA Intensive or Training Services

	Adults	Dislocated Workers	Adults	Dislocated Workers
	Employment		**Earnings**	
Overall Impact	6.7%** (0.40%, n = 75,770)	11.6%** (0.51%, n = 76,858)	$764** ($40, n = 73,830)	$868** ($45, n = 74,422)
Impact Range among States	4.0%** – 11.5%** (0.57% – 1.18%)	10.2%** – 16.0%** (0.91% – 1.44%)	$287** – $980** ($67 – $89)	$312** – $1,435** ($110 – $129)
Impact for Men	3.8%** (0.43%, n = 32,230)	10.0%** (0.60%, n = 39,438)	$685** ($52, n = 31,358)	$842** ($64, n = 38,010)
Impact for Women	9.0%** (0.62%, n = 43,540)	13.3%** (0.85%, n = 37,420)	$823** ($60, n = 42,472)	$894** ($64, n = 36,412)
	Months on TANF			
Overall Impact	−2.1%** (0.16%, n = 75,770)	−1.7%** (0.11%, n = 76,856)		
Impact Range among States	−3.3%** – 2.4%** (0.30% – 0.29%)	−3.6%** – −0.1% (0.32% – 9.87%)		
Impact for Men	−0.8%** (0.10%, n = 32,230)	−0.8%** (0.08%, n = 39,438)		
Impact for Women	−3.1%** (0.29%, n = 43,540)	−2.6%** (0.21%, n = 37,418)		
	Rate of job entry		**Rate of job retention**	
Overall Impact	0.16** (0.008, n = 72,896)	0.21** (0.009, n = 75,060)	0.05** (0.005, n = 70,656)	0.07** (0.004, n = 71,739)
Impact Range among States	0.11** – .28** (0.014 – 0.025)	0.15** – 0.36** (0.014 – 0.033)	0.03 – 0.07** (0.016 – 0.016)	0.04** – 0.10** (0.013 – 0.009)
Impact for Men	0.12** (0.008, n = 31,410)	0.19** (0.011, n = 38,634)	0.04** (0.007, n = 30,203)	0.06** (0.006, n = 36,341)
Impact for Women	0.19** (0.013, n = 41,486)	0.24** (0.015, n = 36,426)	0.07** (0.006, n = 40,453)	0.08** (0.007, n = 35,398)
	Pre-to-post earnings change		**Post-reemployment earnings change**	
Overall Impact	26.7%** (1.36%, n = 71,310)	18.8%** (1.23%, n = 71,654)	−3.2%** (0.88%, n = 71,300)	−3.5%* (1.60%, n = 71,644)
Impact Range among States	13.0%** – 35.5%** (1.59% – 3.23%)	6.8%** – 34.1%** (1.55% – 3.38%)	−11.0%** – −0.9% (2.30% – 0.85%)	−23.9%** – 0.1% (3.04% – 0.30%)
Impact for Men	25.9%** (1.83%, n = 30,326)	20.6%** (1.88%, n = 36,504)	−4.7%** (1.38%, n = 30,322)	−5.3%** (1.34%, n = 36,498)
Impact for Women	27.3%** (1.95%, n = 40,984)	16.9%** (1.56%, n = 35,150)	−2.1%* (1.01%, n = 40,978)	−1.6% (2.32%, n = 35,146)

Note: ** = p<0.01, * = p<0.05. Standard errors in parentheses, followed by total sample sizes. To protect state confidentiality, sample sizes not shown for range.

References

Angrist, Joshua and Jinyong Hahn. 2004. "When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects." *The Review of Economic and Statistics* 86(1): 58–72.

Baj, John, Charles E. Trott, and David W. Stevens. 1991. *A Feasibility Study of the Use of Unemployment Insurance Wage-Record Data as an Evaluation Tool for JTPA,* Washington, D.C.: National Commission for Employment Policy, Research Report Number 90-02, January.

Barnow, Burt S. and Christopher T. King. 2005. *The Workforce Investment Act in Eight States: Final Report,* Washington, D.C.: U.S. Department of Labor, Employment and Training Administration, Office of Policy Development and Research, ETA Occasional Paper 2005-01.

Bloom, Howard S. et al. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National JTPA Study." *Journal of Human Resources* 32(3) (Summer): 549–576.

Bollinger, Christopher and Amitabh Chandra. 2005. "Iatrogenic Specification Error: Cleaning Data can Exacerbate Measurement Error Bias." *Journal of Labor Economics* 23(2): 235-57.

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(448): 1053–1062.

———. 2002. "Propensity Score Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84(1): 151–161.

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35 (4 December): 1809–1855.

Heckman, James and Salvador Navarro-Lozano. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *The Review of Economic and Statistics* 86(1): 30–57.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5), September, pp. 1017-1098.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluation a Job Training Programme." Review of Economic Studies, Vol. 64, pp. 605-654.

Hollenbeck, Kevin. 2004. *On the Use of Administrative Data for Workforce Development Program Evaluation.* Paper presented at the National Workforce Investment

Research Colloquium sponsored by the U.S. Department of Labor, Employment and Training Administration held in Arlington, VA, May 24.

Hollenbeck, Kevin, Christopher T. King, and Daniel Schroeder. 2003. *Preliminary WIA Net Impact Estimates: Administrative Records Opportunities and Limitations.* Paper presented at the Bureau of Labor Statistics and the Workforce Information Council's Symposium, "New Tools for a New Era!" Washington, DC, July 23–24.

Hotz, V. Joseph, Robert Goerge, Julie Balzekas, and Francis Margolin, Eds. 1998. Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development, A Report of the Advisory Panel on Research Uses of Administrative Data. Chicago, IL: Northwestern University/University of Chicago Joint Center for Poverty Research.

Hotz, V. Joseph and John K. Scholz. 2001. "Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data." Institute for Research on Poverty Discussion Paper no. 1224-01.

Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *The Review of Economic and Statistics* 86(1): 4–29.

King, Christopher T. 2004. "The Effectiveness of Publicly Financed Training in the United States: Implications for WIA and Related Programs." In Christopher J. O'Leary, Robert A. Straits, and Stephen A. Wandner, Eds., *Job Training Policy in the United States.* Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Kornfeld, Robert and Howard S. Bloom. 1997. "Measuring Program Impacts on Earnings and Employment: Do UI Wage Records from Employers Agree with Surveys of Individuals?" Unpublished paper prepared for the U.S. Department of Labor. July.

LaLonde, Robert J. 1995. "The Promise of Public Sector-sponsored Training Programs." *Journal of Economic Perspectives* 9(2): 149–168.

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?" *The Review of Economic and Statistics* 86(1): 156–179.

Mosteller, F. M. and R. R. Bush. 1954. in G. Lindzey (ed.) *Handbook of Social Psychology*, vol. 1, Cambridge, Mass: Addison-Wesley.

Mueser, Peter, Kenneth R. Troske, and Alexey Gorislavsky. 2003. "Using State Administrative Data to Measure Program Performance." Unpublished manuscript. University of Missouri, Department of Economics, Columbia, MO.

Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin, and George Cave. 1995. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study,* Washington, D.C.: The Urban Institute Press.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41–55.

Stine, Robert A. 1990. "An Introduction to Bootstrap Methods: Examples and Ideas." *Modern Methods of Data Analysis*. J. Fox and J.S. Long, eds. Newbury Park, CA: Sage Publications.

U.S. Department of Labor, Employment and Training Administration. 2005. *Training and Employment Guidance Letter No. 18-04.* Washington, D.C.: USDOL/ETA. February 28.

Wolf, Frederic M. 1986. *Meta-analysis: Quantitative methods for research synthesis*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no 59. Beverly Hills: Sage.

Zhao, Zhong. 2004. "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence." *The Review of Economic and Statistics* 86(1): 91–107.