Robert M.
La Follette School of Public Affairs
at the University of Wisconsin-Madison

Working Paper Series
La Follette School Working Paper No. 2009-013
http://www.lafollette.wisc.edu/publications/workingpapers

New Estimates of Public Employment and Training Program Net Impacts: A Nonexperimental Evaluation of the Workforce Investment Act Program

Carolyn J. Heinrich
Professor, La Follette School of Public Affairs, and associate director,
Institute for Research on Poverty, at the University of Wisconsin-Madison

cheinrich@lafollette.wisc.edu

Peter R. Mueser
University of Missouri, IMPAQ International LLC and Institute for the Study of Labor (IZA)

Kenneth R. Troske
University of Kentucky and Institute for the Study of Labor (IZA)

Kyung-Seong Jeon
University of Missouri

Daver C. Kahvecioglu
IMPAQ International LLC

THE UNIVERSITY of WISCONSIN MADISON

Robert M. La Follette School of Public Affairs
1225 Observatory Drive, Madison, Wisconsin 53706
Phone: 608.262.3581 / Fax: 608.265-3233
info@lafollette.wisc.edu / http://www.lafollette.wisc.edu

New Estimates of Public Employment and Training Program Net Impacts:

A Nonexperimental Evaluation of the Workforce Investment Act Program

Carolyn J. Heinrich
LaFollette School of Public Affairs, University of Wisconsin

Peter R. Mueser
University of Missouri, IMPAQ International, LLC, and IZA

Kenneth R. Troske
University of Kentucky and IZA

Kyung-Seong Jeon
University of Missouri

Daver C. Kahvecioglu
IMPAQ International, LLC

June 2009

Please address correspondence to Peter Mueser, Department of Economics, University of Missouri, Columbia, MO 65211, mueserp@missouri.edu.

ABSTRACT

This paper presents nonexperimental net impact estimates for the Adult and Dislocated Worker programs under the Workforce Investment Act (WIA), the primary federal job training program in the U.S. The key measure of interest is the difference in average quarterly earnings or employment attributable to WIA program participation for those who participate, estimated for up to four years following entry into the program. The Adult program serves disadvantaged workers, who display relatively poor labor market performance, often over extended periods. The Dislocated Worker program serves individuals who were recently laid off, often as a result of firm downsizing or plant closure.

These estimates of WIA program impact are based on administrative data from 12 states, covering approximately 160,000 WIA participants and nearly 3 million comparison group members. Focusing on participants who entered WIA programs July 2003-June 2005, the study considers the impact for all those in a program, the impact for those receiving counseling and related job search services but no training, and the incremental impact of training. Propensity score matching methods are used to compare WIA program participants with comparison groups of individuals who are observationally equivalent across a range of demographic characteristics, social welfare benefit receipt, geographic area, and prior labor market experiences but who either did not receive WIA services or did not receive WIA training. The comparison group used to identify the overall impact of the WIA program consists of individuals participating in the Unemployment Insurance program (nine states) or individuals receiving job search and related services through Wagner-Peyser legislation (three states). Both comparison groups contain individuals facing employment difficulties and are therefore similar in important respects to WIA participants.

The results for the average participant in the WIA Adult program show that participating is associated with a several-hundred-dollar increase in quarterly earnings. Adult program participants who obtain training have lower earnings in the months during training and the year after exit than those who don't receive training, but they catch up within 10 quarters, ultimately registering large total gains. The marginal benefits of training may exceed $400 in earnings each quarter three years after program entry.

Following entry into WIA, Dislocated Workers experience several quarters for which earnings are depressed relative to comparison group workers with the same characteristics and work histories. As a group, their earnings ultimately overtake the comparison group, although the analyses suggest that the benefits they obtain are smaller than for those in the Adult program.

Although it is not possible to rule out the possibility that some of our estimates may be influenced by systematic selection that has not been controlled by these methods, we undertake a variety of robustness tests suggesting that the general pattern of the results almost surely reflects actual program impacts on individual participants.

I. Introduction

In the midst of one of the most severe economic downturns in history, it is not difficult to grasp the reality and implications of increasing labor market volatility that has affected both low-wage workers and more experienced and skilled workers in recent decades. In fact, these trends of stagnating economic mobility, dislocation and longer-term joblessness have been well-documented (Appelbaum, Bernhardt and Murnane, 2003; Bradbury and Katz, 2002; Holzer, 2004; Osterman, 2007). At the same time that U.S. workers have been facing these growing labor market challenges, however, public expenditures on employment and training services have been declining. For example, in fiscal year (FY) 2007, the total federal government appropriations for Workforce Investment Act (WIA) programs—youth employment, adult job training, dislocated worker assistance, Job Corps and other national activities—was $4.4 billion, down 18 percent from FY 2005. Furthermore, within the WIA program, the number of adults receiving training has declined by 17 percent relative to WIA's predecessor, the Job Training Partnership Act (JTPA) program (Frank and Minoff, 2005).

Enacted in August 1998, the central goal of WIA was to create a new, comprehensive workforce investment system. WIA is distinguished from the JPTA program primarily by the introduction of a One-Stop service delivery system designed to improve coordination and integration of services, its use of Individual Training Accounts in training services, and significant changes in governance structures at the state and local level. In actual implementation, WIA has reduced the share of low-income individuals served by one-third, decreased the length of time spent in training and the expenditures per trainee (in addition to the proportion receiving training), and shifted responsibility for some types of activities believed to contribute little (or negatively) to performance outcomes, such as adult basic education, to other programs (Osterman, 2007). Thus, important changes in both investments in and the implementation of public employment training programs have taken place in the last decade, and yet surprisingly little is known about the impact of WIA and its components on labor market outcomes.

Prior to this study, there has been no formal experimental or nonexperimental evaluation of WIA using the administrative data that states are required to collect for purposes of performance evaluation. In fact, the U.S. Office of Management and Budget assigned the WIA program relatively low marks for its evaluation efforts (using its Program Assessment Rating Tool), suggesting that independent evaluations had not been of sufficient scope and rigor to determine WIA's impact on participants' employment and earnings. Although the U.S. Department of Labor (DOL) recently initiated a project to experimentally evaluate the WIA program, results will not be available for at least seven years. Given the current policy context, in which more than 2.7 million workers were added to the unemployment rolls in the last year and the Obama administration and other policymakers are calling for expanded public investments in employment and training to increase individual skill levels and their success in the labor market,[1] we argue that rigorous evidence on WIA's impact and effectiveness is needed now.

[1] Source: http://origin.barackobama.com/issues/urban_policy/, accessed January 28, 2009.

This study employs nonexperimental methods to evaluate the WIA Adult and Dislocated Worker programs using data from 12 states that cover approximately 160,000 WIA participants and nearly 3 million comparison group members. Within each state, we compare WIA program participants with a matched comparison population of individuals who have not participated in the WIA program but who are observationally equivalent across a range of demographic characteristics, social welfare benefit receipt and labor market experiences. Comparison group members are drawn from those who receive employment services under Wagner-Peyser legislation or who receive Unemployment Insurance benefits. Participants and comparison group members are compared within state and state-established workforce investment areas to assure that they are facing similar local labor markets, and measures of employment are fully comparable for program participants and the comparison group. Research on matching methods, discussed further in the next section, suggests that the research design and data for this evaluation satisfy basic criteria essential for substantially reducing bias in the nonexperimental identification of program impacts.

At the same time, we recognized that, in the absence of data drawn from a representative sample of the population of WIA participants, this study cannot claim to estimate a "national" average impact of WIA. In fact, no experimental or nonexperimental employment and training program evaluation has done this for WIA or any of its predecessor programs. Nonetheless, the sample of WIA participants considered here suitably reflects the diversity of local Workforce Investment Areas, in terms of both geography and environment, including states from each major region in the U.S. and coverage of urban and rural areas; and in terms of operations, with programs that train varying proportions of their participants and manage delivery of services through a variety of organizational configurations in One-Stop centers. The states in this study account for about a fifth of the Workforce Investment Areas in the US.

This paper is organized as follows. Section II describes the methodology employed in the evaluation as well as the study's plan of analysis. Section III introduces the data, providing some basic tabulations, and Section IV provides technical details of the matching methods employed. Sections V and VI present and discuss the results of the analysis, providing impact estimates for the Adult and Dislocated Worker programs, respectively. Section VII provides a brief summary and conclusion.

II. Methodology

This section discusses how the evaluation problem is conceptualized in statistical terms and provides a general review of matching methods used in the analysis reported here. The second subsection presents an overview of the design we employ to implement these methods.

1. The Evaluation Problem

In evaluating the impact of an intervention on its participants, it is necessary to identify both the outcomes for the participants and the outcomes that would have occurred had they not participated. An extensive literature addresses the general issues and challenges of identifying conditions under which such inferences are valid.

Let Y_1 be earnings (or some outcome) for an individual following participation in the program and Y_0 be earnings for that individual over the same period in the absence of participation. It is impossible to observe both measures for a single individual. Specifying D=1 for those who participate and D=0 for those who do not participate, the outcome observed for an individual is:

$$Y = (1 - D)Y_0 + DY_1.$$

Assuming that, given measured characteristics X, participation is independent of the outcome that would occur in the absence of participation,

$$Y_0 \perp\!\!\!\perp D \mid X, \tag{1}$$

the effect of the program on participants conditional on X can be written as

$$E(Y_1 - Y_0 \mid D = 1, X) = E(\Delta Y \mid D = 1, X) = E(Y_1 \mid D = 1, X) - E(Y_0 \mid D = 0, X) \tag{2}$$

where $Y_1 - Y_0 = \Delta Y$ is understood to be the program effect for a given individual and the expectation is across all participants with given characteristics.

Although (2) may be estimated in a variety of ways, in recent years, attention has focused on matching methods, which are designed to ensure that estimates of program impacts are based on outcome differences between comparable individuals. (See Rosenbaum, 2002, Imbens, 2004 and Rubin, 2006 for general discussions of matching methods.) The analysis in this study employs propensity score matching, in which participants are matched with individuals in a comparison group based on an estimate of the probability that the individual receives treatment (the propensity score).

Propensity score matching. In the combined sample of participants and comparison group members, let *P*(*X*) be the probability that an individual with characteristics *X* is a participant. Rosenbaum and Rubin (1983) show that

$$Y_0 \perp\!\!\!\perp D \,\big|\, \mathrm{X} \;\Rightarrow\; Y_0 \perp\!\!\!\perp D \,\big|\, P(X).$$

This means that if participant and comparison group members have the same *P*(*X*), the distribution of *X* across these groups will be the same. Based on this propensity score, the matching problem is reduced to a single dimension. The impact of the treatment on those who are treated can be estimated as

$$\begin{aligned} E\left(\Delta Y \middle| D=1\right) &= E_{X|D=1}\left(\Delta Y \middle| P\left(X\right)\right) \\ &= E_{X|D=1}\left(E(Y_1 \mid P(X), D=1) - E(Y_0 \mid P(X), D=0)\right), \end{aligned}$$

where $E_{X|D=1}$ is the expectation across all values of *X* for participants. The propensity score is thus a balancing score for *X,* assuring that for a given value of the propensity score, the distribution of *X* will be the same for participants and comparison cases.

In practice, the propensity score must be estimated. Normally, a logit or probit function is used for this purpose, and it is critical that the functional form be flexible. The discussion of the details of execution of the matching analysis (Section IV) further explicates these issues.

Difference-in-difference matching. In some cases, the condition in (1) is clearly not met because individuals are selected into the program on the basis of unmeasured personal characteristics that are expected to influence ultimate outcomes. Individual fixed effects estimators provide an alternative approach to controlling for differences across individuals who participate in WIA. This approach, in essence, produces estimates of the impact of participation by comparing a recipient's experience prior to participation with his or her subsequent experience, and then comparing this with the same measure for nonparticipants. Smith and Todd (2005a) spell out the basic approach, which they describe as "difference-in-difference" matching. For treated cases, the dependent variable is the difference between earnings following participation and earnings prior to program participation, and for comparison cases the earnings difference is calculated over the same period. Even if individuals who participate in WIA differ in important ways from those in the comparison group, so long as such differences have a stable impact on earnings, this specification can eliminate bias resulting from differences between participants and others. In other words, even if (1) does not hold, it may be the case that

$$Y_0 - Y_0^{-1} \perp D \mid X \,, \tag{3}$$

where Y_0^{-1} is prior earnings. The program impact can then be estimated as

$$E(\Delta Y \big| D=1, X) = E(Y_1 - Y_0 \big| D=1, X) = E(\,Y_1 - Y_0^{-1} \big| D=1, X) - E(\,Y_0 - Y_0^{-1} \big| D=0, X)\,.$$

In this case, the differenced earnings for those not receiving treatment can be used to estimate what the earnings difference would have been in the absence of services for those who received services, that is, in the counterfactual state.[2]

Despite the benefits of difference-in-difference estimates, depending on the processes underlying earnings dynamics and program participation, estimates may have biases that are not present in cross-sectional matching. The difference-in-difference estimates need to be understood as one of several estimates that make different assumptions.

Mechanics of matching approaches. Early matching estimators paired individuals in the treated group with those in the comparison group on a one-to-one basis. Although pairwise matching is most intuitive, in recent years alternative approaches have been recognized as superior. It has been found that estimates are more stable (and make better use of all available data) if they consider all comparison cases that are sufficiently close to a given treated case. It is also important to include in the comparison only those cases that are sufficiently "close" to a given treated case, which can be achieved with what has been termed the "caliper" matching approach. Finally, although allowing a given case to be used in many comparisons may inflate sampling error, it is now generally accepted that the benefits of close matches outweigh these costs.

One approach that captures many of the benefits is many-to-one caliper matching with replacement. This estimator of program impact may be written as

$$E(\Delta Y) = \frac{1}{N}\sum_{i=1}^{N}[Y_{1i} - \overline{Y}_{0j(i)}], \tag{4}$$

where $\overline{Y}_{0j(i)}$ is the average outcome for all comparison individuals who are matched with case i, Y_{1i} is the outcome for case i, and N is the number of treated cases. Sometimes referred to as "radius matching," this approach does not limit the number of cases that are matched with a given participant, as long as those cases are "close" enough.

This method is closely related to matching by propensity score category, a method advocated in Dehejia and Wahba (2002). In a comparison of matching methods, Mueser, et al. (2007) found that methods like this one, which use all the available data, produced more precise program impact estimates.

Validity of impact estimates. Dehejia and Wahba (1999, 2002), applying matching methods to data from the National Supported Work demonstration project (originally analyzed by LaLonde, 1986), present a strong case in support of these methods for evaluating job training programs.

[2] It is worth noting that this approach differs from a simple first difference approach as it is often implemented in that personal characteristics X may include measures that do not change over time. Controlling for such measures may be necessary in some cases. For example, if those with higher levels of education experience greater growth over time in earnings, it may be necessary to match individuals with the same levels of education—in effect using education to identify what income growth would be in the absence of program participation.

Their claim that matching can produce the same estimates as random-assignment methodologies remains controversial (see Smith and Todd, 2005a, 2005b, and Zhao, 2003), although the recent work of Mueser et al. (2007) similarly concludes that matching methods may be effective in evaluating job training programs.

Heckman, LaLonde and Smith (1999) provide the most extensive discussion of evaluation of job training programs. Their conclusions regarding matching methods are nuanced, and they argue that methods are likely to be successful under specified circumstances. Their conclusions are largely echoed in Bloom, Michalopolos and Hill (2005), who summarize studies that use experimental evaluations of program impacts and compare these with nonexperimental studies. They conclude that "The most successful methods have the following ingredients: Local comparison groups from the same economic or institutional settings, comparable outcome measures from a common outcome source, longitudinal data on baseline outcome measures, and a nonparametric way to choose comparison group members who are observationally similar to program-group members and eliminate those who are not." The analyses undertaken here satisfy these conditions. WIA and comparison program participants are both in the same state, and local labor markets within the state are identified for each individual. Both WIA participants and the comparison group are at junctures in their careers when they are either facing employment crises or are at least considering alternative vocational options. In all analyses, state UI wage record data are the source of outcome earnings.

Glazerman, Levy and Meyers (2003) undertake a quantitative analysis based on results of studies comparing nonexperimental and experimental methods. Their work is particularly relevant, since it focuses exclusively on evaluations that, like this study, consider earnings outcomes. Their results are largely consistent with the conclusions listed above. They find that both regression and matching techniques are useful in improving estimates, and that both used together show greater benefits than either alone. Again, their work underscores the value of prior information on earnings and on geographic match. Finally, they confirm the value of large sample sizes for the comparison group, suggesting that the administrative datasets used in the current study confer important benefits.

Two recent studies consider the success of nonexperimental studies in reproducing experimental results. Cook, Shadish and Wong (2008) compare nonexperimental and experimental results in a wide range of studies where outcomes include test scores and school attendance as well as labor market success. Their general conclusions are positive, suggesting that nonexperimental analyses may be successful in many cases, but they argue that selection into job training programs is more complex than for other programs, and that this limits the efficacy of nonexperimental methods as compared with studies of other kinds of programs. However, the studies they cite do not consider nonexperimental methods that satisfy the above requirements. Piekes, Moreno and Orzol (2008) are more pessimistic about nonexperimental methods in evaluating job training programs, but the comparisons they cite involve very small sample sizes, and it is doubtful that meaningful inferences can be based on them.

Imbens and Wooldridge (2008), using data from the National Supported Work demonstration project, show the importance of having sufficient sample overlap in the covariate distributions of the treatment and comparison groups to precisely estimate average treatment effects.

As is the case with any nonexperimental evaluation method, the richness and relevance of the data available for an evaluation have important implications for the performance of estimators. A key insight of these studies is that if appropriate observed characteristics are measured consistently and balanced well across the treatment and comparison groups, concerns about whether treatment and comparison group members have similar distributions of unobserved characteristics may be of relatively minor import.

2. Overview of Methods of Analysis

The primary focus of the analyses in the current study is on individuals entering WIA in the period July 2003-June 2005 (program years 2003 and 2004). This allows sufficient time after the program's initial startup phase (2000 in most states), yet it provides an extended follow-up period.

WIA programs and activities. We evaluate two WIA programs: the Adult program, serving largely disadvantaged individuals, and the Dislocated Worker program, serving workers who have lost jobs. Given that the two programs serve very different functions, each will be analyzed separately. For these programs, three levels of service are defined, and the services are sequenced from least to most intensive. All participants who enter WIA are eligible to receive Core services, services that involve staff-assisted job search and counseling, corresponding closely to the staff-assisted services offered by state offices under U.S. Employment Service (ES) Wagner-Peyser legislation. Nationwide about one in five WIA participants receives only these services. Once individuals receive Core services, staff may recommend that they receive Intensive services, which involve more extensive counseling and possibly short courses (generally no more than a few days). Finally, participants in Intensive services may then be recommended to receive Training services.[3] As of 2005, about 43 percent of WIA exiters nationwide are coded as receiving Training services.[4] For most of these, training is provided through an Individual Training Account (ITA), a voucher for training from an outside provider, often based on the recommendation of WIA program staff.

These analyses consider how outcomes differ by various levels of service. Given that the distinction between Core and Intensive services is relatively small, their combined effect is estimated, comparing recipients of these services with those who are not WIA participants. In contrast, the impact of Training is estimated based on comparisons with WIA participants who do not receive training, obtaining a measure that reflects the value of the additional services training recipients receive. The level of services defined as Core and Intensive may well vary across states. Separate analyses for each state and groups of states allow for the identification of such differences.

[3] The strict sequential structure of services may not be followed in all sites.
[4] These figures are based on analyses of the WIASRD data undertaken by Social Policy Research Associates (2007, pp. 36, 114).

Outcome measures. The outcome measures in this study are individual earnings and employment. Earnings or employment (identified by positive earnings in the wage record data) are compared in a given quarter for those participating in WIA (or a particular WIA service) with the average earnings/employment in the same period for the matched comparison sample. Effects on earnings and employment are estimated by quarter for up to 16 quarters following the quarter of program entry. Early quarters after program entry are expected to show negative "effects" of training on earnings and employment, reflecting participants' involvement in program activities rather than employment. These are often described as "lock-in" effects. Later earnings effects are expected to be positive, as skills obtained during the program interact with job experience.

Comparison pool. Estimates of the WIA program impact (overall impact for participants receiving Core/Intensive services) use a comparison group drawn from either Unemployment Insurance (UI) claimants or from U.S. Employment Service (ES) participants (i.e., individuals who register with the state's job exchange service and receive some services under Wagner-Peyser legislation). Of the 12 states for which data are available, nine have UI claimant data while three have data for ES participants.

There is substantial but not complete overlap between the UI claimant population and those receiving ES services. In most states, the majority of UI recipients are required to register for ES services, but some claimants do not face this requirement. Conversely, although a majority of ES recipients are or have been receiving UI benefits, anyone seeking services to aid in job search is eligible to receive ES services. Generally, the level of services received for participants in both programs is minimal, and one may view such individuals as representing a "no treatment" control. Alternatively, given that ES and related services are widely available, even if they are believed to provide substantial benefits, they may be viewed as representing a highly relevant "counterfactual" that reflects the program options faced by individuals in the absence of the WIA program.

One important shortcoming of UI recipients as a comparison group is that recipients must have earnings above a minimum (over the past five quarters) in order to receive UI benefits. As a result, it may be difficult to find appropriate UI recipient matches for some WIA participants. For those states where it is available, those who apply for UI benefits but whose claims are rejected are included in the comparison sample, allowing for the possibility that some rejected applicants can serve as matches for WIA participants with weak employment histories. For both comparison groups, any individual who subsequently participates in the WIA program is omitted.

Estimates of the incremental impact of training use a comparison group consisting of WIA participants who did not receive training services, i.e., of those receiving only Core or Intensive services.

The estimation approach depends on the assumption that the no-treatment outcome is independent of whether an individual receives the treatment once measured characteristics are controlled, as specified in equation (1) (or equation (3) for the difference-in-difference specification). If this assumption is violated, that is, if a matched treated case and comparison individual would have had different earnings in the absence of the treatment, the impact estimate will contain bias.[5] The plausibility of the assumption depends on the particular characteristics available for matching.

Control variables. It has long been recognized that controls for the standard demographic characteristics such as gender, age, education and race are important. Such information is available in the current study. Local labor market is captured in the current study using aggregates of county of residence or service, or, where county is not available, the local Workforce Investment Area. It is also widely recognized that the details of the labor market experiences of individuals in the period immediately prior to program participation are critical.[6] The data here provide information on labor force status at the time of initial program involvement, and wage record data for prior years are used to identify previous employment transitions. Additional relevant variables include controls for veteran status, prior earnings, as well as prior year TANF receipt.

As men and women tend to have very different labor market experiences, analyses are performed separately by gender. Where possible, WIA participants who enter in a given quarter are also matched with individuals in the comparison sample who have contact with their respective programs in the same quarter, providing an exact match on quarter of entry.[7] The exact match on quarter of entry assures that any economic trends that could influence labor market success are fully controlled—since they affect both participants and comparison group members.

Hence, control variables include: calendar quarter of program entry (exact match), gender (exact match), age, education attained, race/ethnicity (separate categories for nonwhites and Hispanics), disability status, veteran status, local labor market (local WIA area or other county-based measure), employment information based on wage record data over the two years prior to program entry, including employment transitions and earnings, industry of most recent employment, program participation history (WIA; UI or ES), current and prior TANF receipt, and time since layoff.

Treatment and comparison samples. Table II.1 shows treatment samples and the groups used in each comparison. Columns (a) and (b) indicate for which programs the comparison is undertaken, whereas (c) and (d) identify the treatment and comparison groups. In each of the comparisons, identified by rows, matching is performed on the basis of the individual characteristics.

---

[5] Note selection into the program is normally expected to be correlated with unobserved variables. Such selection only causes bias if it is associated with outcome measures once the observed variables are controlled.

[6] In particular, movements into and out of the labor force and between employment and unemployment in the 18 months prior to program participation are strongly associated with both program participation and expected labor market outcomes (Heckman, LaLonde and Smith, 1999; Heckman and Smith, 1999).

[7] Comparison group individuals may contribute more than one unit as potential matches if they had contact with the program in multiple quarters. In such cases, when a later quarter for a comparison case is chosen to match with a WIA participant, prior quarters of participation in the comparison program must correspond for these cases. Further detail is provided in Section IV.

Line 1 lists comparisons of WIA participants—regardless of services received—with comparison group individuals who have filed for UI benefits or received ES services. These comparisons provide measures of the impact of the WIA program taken as whole. Line 2 lists comparisons that consider the degree to which participants who do not go through training benefit from participation in WIA. Line 3 identifies the comparison between those individuals who participate in WIA Training services and other WIA participants. This comparison allows one to identify the extent to which training, per se, is associated with employment and earnings outcomes. As noted above, all comparisons are undertaken within a given state.

Table II.1
Treatment and Comparison Samples

	WIA Program Group		Sample Group	
	Adult	DW	Treatment	Comparison[a]
	(a)	(b)	(c)	(d)
1.	X	X	WIA	UI Claim or ES
2.	X	X	WIA Core/Intensive	UI Claim or ES
3.	X	X	WIA Training	WIA Core/Intensive

Matching approach. The propensity score $P(X)$ is estimated using a logit specification with a highly flexible functional form allowing for nonlinear effects and interactions. It is necessary to test to assure that the estimated propensity score is successful in balancing values of matched treatment and comparison cases. Following the matching, tests for statistically significant differences between variable means for the treated cases and the weighted comparison sample are performed to assure that the score in fact balances the independent variables (see Smith and Todd, 2005a).

Radius matching is used in order to identify comparison cases that correspond with treated cases. For each treated case within the subgroup, "matches" are designated as all comparison cases for which the value of the log odds of the propensity score is within a given radius of the treated case. This method not only allows for *more than one* comparison case to be matched with a treated case, but, because the search in the comparison sample is done with replacement, it also allows *a given* comparison case to be matched with more than one treated case. The mean outcome for cases matched with a given treated case is an estimate of the outcome that would occur for the treated case in the absence of the service, so impact estimates follow the form of the expression (4) above.

The vast majority of studies using propensity score matching measure the proximity of cases as the absolute difference in the propensity score. As Smith and Todd (2005a) note, such an approach is not robust to "choice-based sampling," where the treated and comparison pools are obtained from different sources. Matching on the log odds of the propensity score assures that results are invariant to choice-based sampling. In addition, since the logit is used to predict propensity score, the log odds are a linear combination of the independent variables, and a constant radius in the log odds will translate into the same metric at different propensity score levels.

The samples here, like those used in most nonexperimental evaluation research, are choice-based, and matching on the log odds ratio is invariant to the sampling proportions. Comparison samples are also very large, often more than 50 times larger than the sample of treated cases, and, as a result, propensity scores for most cases—both treatment and comparison—are small, generally less than 0.05. As a result, a large share of the comparison cases is compressed into a very small range of propensity scores. Matching on the log odds of the propensity score has the advantage that it "spreads out" the density of very low or very high propensity scores. Preliminary experiments suggested that matching is much more successful with the log odds.

The choice of radius involves a trade-off between potential bias and statistical stability. When the radius is too small, although any comparison case matched to a given treated case may be almost identical in terms of measured variables, other comparison cases that may be quite similar to a given treated case are lost. As a result, there may be no or few comparison cases available for some treated cases, so effect estimates may be unstable, and the analysis may omit some treated cases that cannot be matched. Conversely, where the radius is too large, comparison cases will not be sufficiently similar to treated cases. The weighted cross-validation method outlined in Galdo, Smith and Black (2008), which is designed to minimize the mean squared error of the matching estimator, was used as an aid to choosing the optimal radius.

Although propensity score matching assures that differences in characteristics between treated and matched comparison cases are small, for certain variables that are particularly important in determining the outcome, it may be prudent to assure that matching is exact. As noted above, analyses are undertaken separately by gender, assuring that a male is never matched with a female. Labor market opportunities and other experiences may also be influenced in a direct way by seasonal and other time factors, and so most analyses employ exact matching by calendar quarter, assuring that impact estimates are based on a comparison of individuals during the same time period.

Generalizing results. The analyses in this study are designed to provide estimates of average impact for participants in WIA Adult and Dislocated Worker programs in 12 states that provided data. To what degree can these results be generalized to the remainder of the states? Sampled states come from all the main geographic regions in the country, and they include five of the 30 largest U.S. cities. Although the sampling frame does not have any of the 10 largest cities, several older eastern cities are included, as are several slow-growing Midwestern cities. Given the decentralized structure of the WIA program, differences between Workforce Investment Boards within a state are often very large, and differences within a state between areas due to demographic and economic environments may dwarf between-state differences. Thus, the sampling frame in this study is less restricted than might initially be assumed. These considerations suggest that estimates of program impact applying to the 12 states in the current study may well be valid as well for many of the remaining states.

The clearest threat to generalization would be if states were selected (or had selected themselves) on the basis of actual program performance. In this case, the 12 states might display impacts that were wholly unrepresentative of the remaining states. Although this possibility cannot be rejected, previous work suggests that neither local administrators nor state agencies are able to judge the efficacy of programs, particularly when considering program impacts over time. State administrative and data handling idiosyncrasies may have played a dominant role in determining willingness to provide data for the study.

In addition to presenting estimates of average program impact, we also discuss effects for individual states. Many of the patterns are common—or dominant—across the states in the sample. In large part, it is these patterns that provide the most useful results. Even in the absence of a formal statistical test, it is reasonable to assume such results would be observed in the remaining states.

III. The Data

This study uses administrative data from 12 states, dividing the data for each state into three classes: base data, comprising WIA program participants; comparison data, providing information on individuals in other programs who are matched to treated cases; and outcome data, merged by individual identifier to the base and comparison data. This section introduces these data and provides basic descriptive statistics.

1. Data Sources

Twelve states provided usable data within the time frame necessary for the analysis. These states were Connecticut, Indiana, Kentucky, Maryland, Missouri, Minnesota, Mississippi, Montana, New Mexico, Tennessee, Utah, and Wisconsin. The agreement for use of these data required that results for these states not be separately reported.

The base data include annual Workforce Investment Act Standardized Record Data (WIASRD) or closely related files obtained from each state, providing information on all participants exiting the WIA program within a program year (July-June). For most states, the data files extend through June 2007 (Program Year 2006). These data also include an individual identifier to allow a match with other state data. The focus of the current analysis is on WIA participants who entered the WIA program in the period July 2003-June 2005. In most cases no information is available on individuals who did not exit the program by June 2007.[8]

In nine of the states, the comparison group is constructed from Unemployment Insurance (UI) claim data, and in the other three states from U.S. Employment Service (ES) data (individuals seeking job counseling or job search assistance under Wagner-Peyser legislation). In addition to

[8] Two of the twelve states provided WIA exit data extending through only June 2006. Since WIA participants who did not exit the program by this date are omitted, a larger share of individuals are omitted in these states. Because of data problems, in two states, the study examined program entries for periods other than July 2003-June 2005, one for calendar year 2003 and the other for January 2004-June 2005.

facilitating the construction of comparison groups, these data were used to control program participation prior to the quarter of program entry for both participants and comparison group individuals. In all but three states, at least six quarters of such information are available prior to the first quarter of program participation.

As noted above, the UI and ES samples are expected to include many of the same individuals. The ES sample has the advantage that it includes any individual who chooses to obtain services without regard to prior employment history, whereas UI provides benefits only to those who have sufficient prior work experience. Although, where possible, UI applicants who were refused benefits have been retained, the UI sample may not provide matches for WIA participants with very limited work history. As a practical matter, negotiating use of the ES data was more complex, and it was not possible to arrange for use of ES data in most states.

Unemployment Insurance (UI) Wage Record data provide quarterly earnings for all employees in UI-covered firms within a state. Data extend through calendar year 2007, which, when matched with WIASRD information and information for individuals in the comparison groups, generate the study's primary outcomes measures. These include earnings and employment for participants for up to 16 quarters following participation and for comparison group members in the same periods. These data also include quarters prior to WIA participation, facilitating the construction of employment histories of participants and comparison group members.[9]

2. Descriptive Statistics

Table III.1 provides sample sizes and means for WIA participants and the comparison group in the 12 states. A total of 95,580 unique individuals entered the WIA Adult program during the observation window. Since about 2 percent entered the program more than once, the total number of entries was 97,552. Similarly, 63,515 individuals entered the Dislocated Worker program, producing a total of 64,089 total program entries.[10] The rightmost column identifies the number of individuals who participate in comparison programs and are available to be matched to program participants. The upper entry indicates that approximately 2.9 million unique individuals are available, contributing nearly 6.2 million quarters of program activity.[11] A very large number of comparison cases is available for matching.

---

[9] In one state, wage record data extend only through June 2007. Wage record data are available for at least four quarters prior to the first quarter of analysis in every state, and in all but three states a full two years of wage record data are available for all WIA entry dates considered.

[10] Where an individual entered the program more than once during a quarter, this was considered to be only a single entry. Data cleaning also eliminated multiple entries when these appeared to be due to data entry errors or when they pertained to the same set of services.

[11] The matching methods employed here consider all quarters of comparison program participation, allowing a given individual to be matched to WIA participants in more than one quarter.

Table III.1
Summary Statistics for WIA Participants and Comparison Group in 12 States

	WIA Adult			WIA Dislocated Worker			
	Overall	No Training	Training	Overall	No Training	Training	Comparison Group
Sample Size							
Unique individuals	95,580	68,255	27,325	63,515	43,513	20,002	2,929,496
Total quarters of participation	97,552	69,712	27,840	64,089	43,894	20,195	6,161,510
Demographic	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Male	0.420	0.445	0.356	0.482	0.494	0.456	0.585
Black	0.445	0.512	0.277	0.330	0.391	0.198	0.171
Hispanic	0.031	0.014	0.072	0.022	0.013	0.043	0.064
Age	32.70	32.91	32.16	40.24	40.14	40.46	39.59
Years of education	12.27	12.21	12.43	12.55	12.52	12.63	12.42
Employment							
Employment-employment	0.297	0.294	0.307	0.462	0.465	0.456	0.476
Employment-not employed	0.208	0.195	0.241	0.281	0.256	0.335	0.279
Not employed-employed	0.325	0.336	0.297	0.183	0.199	0.149	0.225
Not employed-not employed	0.168	0.175	0.151	0.070	0.078	0.053	0.040
Earnings second year prior	8507	8203	9306	19402	17782	23487	20156
Earnings in prior year	8149	8050	8398	20499	19450	22779	21584
Earnings following year	9426	9128	10171	11527	11840	10845	15649
Earnings second year after	10846	9916	13175	14572	14213	15352	17102
Program Experience							
WIA in prior two years	0.052	0.058	0.035	0.041	0.044	0.034	0.020
Comparison program participation in prior two years	0.211	0.178	0.297	0.409	0.353	0.551	0.668

Turning to the next panel, we can see that individuals who participated in the WIA Adult program are more likely to be female and minority than participants in the comparison program; they are also appreciably younger and have slightly less education. These differences likely reflect the fact that participants in the WIA Adult program tend to be economically disadvantaged, whereas participants in the comparison program (UI claimants or ES participants) are individuals who have recently lost jobs. Therefore, individuals in the comparison program have the characteristics of individuals with relatively strong labor market attachments—white, male, older workers with more education. Comparing participants in the WIA Dislocated Worker program with the comparison group, it is clear there are fewer differences—participants in the WIA Dislocated Worker program are more likely to be female and are slightly older, but differences are smaller.

The data on past employment and earnings for these groups provide further evidence that participants in the WIA Adult program have weaker labor market attachments and are more economically disadvantaged than comparison program participants. Participants in the WIA Adult program are less likely to have worked continuously in the six prior quarters (30 percent versus 48 percent) and are much more likely to have not worked in any of the six quarters (17 percent versus 4 percent) prior to entering the program; they also have much lower annual earnings in the two years prior to entering the program and in the two subsequent years. In contrast, participants in the WIA Dislocated Worker program have similar labor market attachment and only slightly lower earnings than those in the comparison program.

The bottom panel of the table shows that 4 to 5 percent of WIA entrants had previously participated in WIA (either the Adult or Dislocated Worker program), and that the number participating in the comparison program was substantial. About a fifth of Adult program participants had prior comparison program experience, compared to over two-fifths of Dislocated Workers. By definition, a comparison case participates in the comparison program in the specified quarter; the table shows that about two-thirds of such individuals had participated in that program in the prior two years.

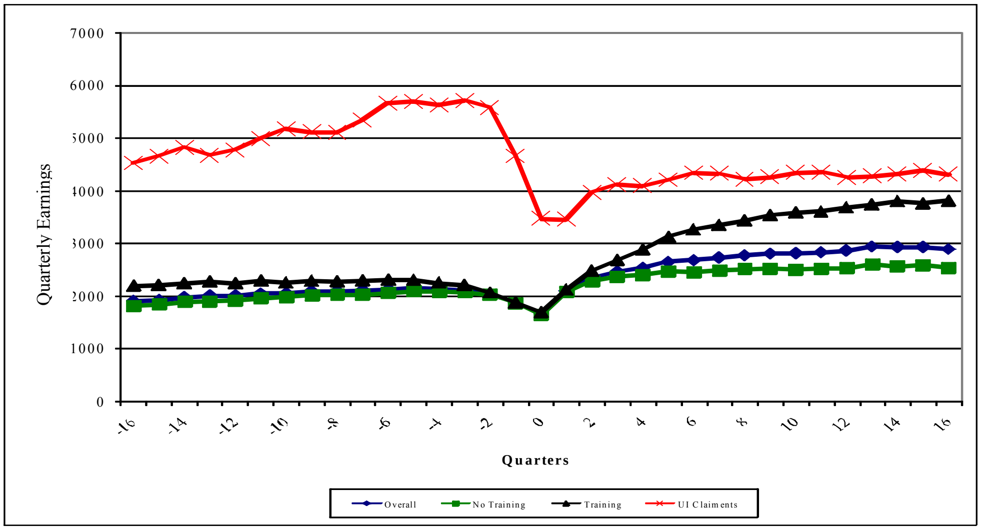
Comparing columns 2 and 3, and columns 5 and 6, it can be seen that participants who receive Training services are more likely to be female and much less likely to be black than participants who do not receive Training services.[12] Differences in education are very small. Based on prior earnings, those receiving Training services appear to have had greater labor market success, but measures of employment imply only small differences in employment activity.

Notwithstanding these differences, there are important similarities in the patterns of earnings for individuals in these states. Figure III.1 graphs quarterly earnings for WIA Adult program participants and the sample of individuals in the comparison group. Figure III.2 provides comparable plots for the Dislocated Worker program. In these figures the negative numbers on the horizontal axis indicate quarters prior to program entry; quarter 0 is the quarter an individual begins participating in a program; and the positive numbers indicate quarters after entry into the program. In each plot, separate lines are provided for all WIA Adult participants, participants who receive Training services, WIA participants who do not receive Training services and comparison cases.[13]

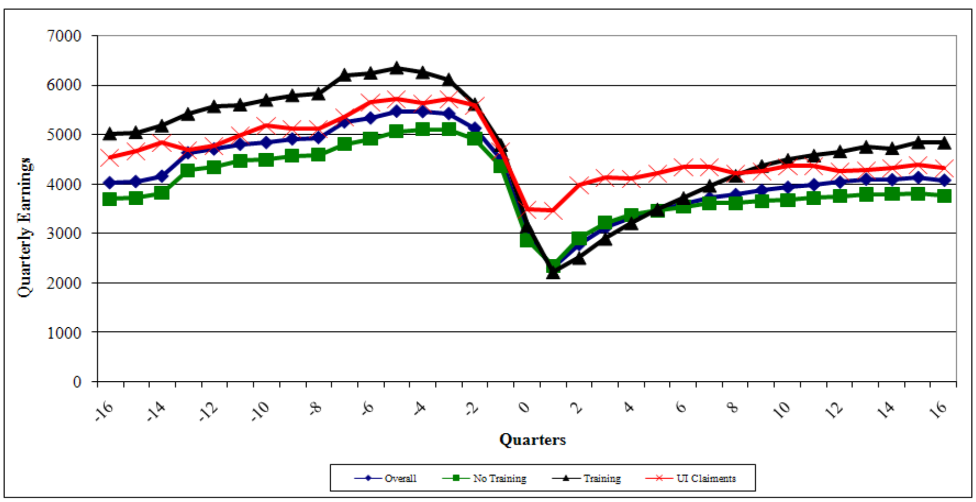
[12] Any WIA participant who does not receive Training services is coded as receiving Core/Intensive services only. In most states, codes separately identify receipt of Intensive or Training services; anyone not so identified is assumed to receive Core services.

[13] Sample sizes are very large, so that even modest differences in quarterly earnings are statistically significant. However, earnings more than eight quarters before or more than ten quarters after entry are based on the subsample of states with that information, and reported means outside this range should be treated with caution, since this subsample is systematically different from the full sample. The presentation of impact estimates incorporates a statistical adjustment to account for such differences.

Figures III.1
Quarterly Earnings for WIA Adult Program and Comparison Program Participants in 12 States, Prior to and Following Participation



Figures III.2
Quarterly Earnings for WIA Dislocated Workers and Comparison Program Participants in 12 States, Prior to and Following Participation



The most notable pattern in both figures is the decline in earnings that occurs in the several quarters prior to program entry, a pattern that has been called the "Ashenfelter dip" (Ashenfelter, 1978; Heckman and Smith, 1999). This reflects the fact that individuals often enter such programs following a period of setbacks in employment. In attempting to find a comparison group, this pattern may set program participants apart from potential comparison individuals. It is therefore significant that there is a decline preceding program participation for the comparison group.

The matching analysis introduced above and described in detail in Section IV is designed to ensure that the analysis only considers comparison group participants who look similar to WIA participants in terms of prior earnings and labor market experience, as well as demographic characteristics. If these variables fully identify those factors that affect outcomes and program participation, program impact estimates based on the differences between the treated and matched comparison cases are valid. That the comparison group displays a similar basic pattern in earnings to the two WIA programs —including a dip at the time of program participation— confirms that there will be sufficient numbers of individuals in the comparison sample to match with WIA participants on the basis of prior employment activities and earnings. It also suggests that there may be similarities in the individual employment environments faced by the comparison and treatment groups, implying that unmeasured factors may be similar as well. Nonetheless, it bears repeating that even with great care in matching, there is no guarantee that all unmeasured differences in factors affecting outcomes between the treated and matched comparison groups will be eliminated. Specification tests are an essential part of the analyses that follow.

The aggregate numbers presented in Table III.1 hide differences across states in programs. The total number of participants entering the Adult and Dislocated Worker programs during the period of the study varies across states from as little as 1,500 to well over 50,000. For that reason, an important element of the analysis that follows is examining whether the patterns are similar in various subsets of the states. Where patterns are similar, this suggests that results are not driven by a small number of large states. One important difference in the character of the programs is reflected in the proportion of individuals who receive training. Seven of the state programs provide training to more than 60 percent of participants, one state provides training to about half of its participants, and the remaining states provide training to less than 40 percent.[14]

Table III.2 presents summary statistics for the seven state programs providing training services to at least 60 percent of participants and Table III.3 provides statistics for the balance of the state programs. The two sets of programs differ in terms of average size. The seven high-training programs average approximately 3,200 Adult participants and 2,400 Dislocated Workers, for a total of approximately 38,000 participants in the seven programs. Approximately two-thirds of all recipients in this group receive training. The comparison group comprises 1.6 million individuals who contribute 3.5 million records available for matching.

Despite differences in WIA program structure, overall patterns for the WIA and the comparison program are similar in Table III.2 and Table III.3. In the set of low-training states, nearly 120,000 WIA participants are identified, and, in aggregate, just 17 percent of WIA participants receive Training services. Over 1.3 million comparison program participants contribute 2.7

[14] In general, those programs that are most likely to provide training to participants in the Adult program are also likely to provide training to participants in the Dislocated Worker program.

Table III.2
Summary Statistics for WIA Participants and Comparison Group in 7 States with High Training Rates

	WIA Adult			WIA Dislocated Worker			
	Overall	No Training	Training	Overall	No Training	Training	Comparison Group
Sample size							
Unique individuals	22,646	7,114	15,532	16,520	5,027	11,493	1,601,399
Total quarters of participation	22,694	7,133	15,561	16,536	5,036	11,500	3,479,550
Demographic	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Male	0.381	0.421	0.363	0.471	0.510	0.453	0.580
Black	0.190	0.256	0.160	0.095	0.124	0.083	0.132
Hispanic	0.107	0.089	0.115	0.068	0.073	0.067	0.090
Age	33.46	35.77	32.40	41.73	42.99	41.17	39.67
Years of education	12.32	12.23	12.36	12.60	12.69	12.57	12.45
Employment							
Employment-employment	0.280	0.269	0.286	0.475	0.494	0.467	0.497
Employment-not employed	0.240	0.214	0.252	0.347	0.300	0.367	0.253
Not employed-employed	0.309	0.361	0.285	0.121	0.150	0.108	0.215
Not employed-not employed	0.161	0.140	0.170	0.042	0.037	0.044	0.031
Earnings second year prior	9526	10303	9218	26156	27732	25462	20701
Earnings in prior year	8352	8626	8226	24618	26062	23985	21435
Earnings following year	10579	11603	10109	11906	16321	9973	15693
Earnings second year after	12903	12452	13109	16428	19742	14976	17092
Program Experience							
WIA in prior two years	0.016	0.017	0.016	0.007	0.010	0.006	0.014
Comparison program participation in prior two years	0.325	0.307	0.332	0.667	0.643	0.676	0.706

million matching units. Demographic differences correspond closely, although levels are different, reflecting differences in state populations. For example, the proportion that is black is approximately half as great in the first set of states, but this difference exists for both WIA and comparison group participants.

One interesting difference is in the proportion of WIA entrants who had been in the program at some point in the prior two years. In the high-training states, only 1.6 percent of Adult program participants and 0.7 percent of Dislocated Workers had prior WIA experience, whereas comparable figures for the low-training states were 6.1 and 4.9 percent. Prior participation in the comparison program, however, is higher in the high-training states, with nearly a third of WIA Adult program participants registering activities in the comparison program in the prior two years, as compared with 18 percent in low-trainings states. A similar differential exists for Dislocated Workers.

Table III.3
Summary Statistics for WIA and Comparison Program Participants in 5 States with Low Training Rates

	WIA Adult			WIA Dislocated Worker			
	Overall	No Training	Training	Overall	No Training	Training	Comparison Group
Sample size							
Unique individuals	72,934	61,141	11,793	46,995	38,486	8,509	1,328,097
Total quarters of participation	74,858	62,579	12,279	47,553	38,858	8,695	2,681,960
Demographic	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Male	0.431	0.448	0.346	0.486	0.492	0.460	0.591
Black	0.522	0.541	0.426	0.412	0.426	0.350	0.221
Hispanic	0.008	0.006	0.018	0.006	0.005	0.013	0.030
Age	32.46	32.58	31.87	39.73	39.77	39.52	39.48
Years of education	12.26	12.21	12.50	12.53	12.50	12.71	12.38
Employment							
Employment-employment	0.302	0.296	0.334	0.458	0.462	0.442	0.449
Employment-not employed	0.198	0.192	0.226	0.258	0.250	0.291	0.313
Not employed-employed	0.329	0.333	0.313	0.205	0.205	0.202	0.238
Not employed-not employed	0.170	0.179	0.127	0.080	0.083	0.064	0.051
Earnings second year prior	8232	8011	9414	17783	16945	21609	19269
Earnings in prior year	8088	7985	8614	19066	18593	21183	21777
Earnings following year	9076	8846	10248	11395	11260	11998	15592
Earnings second year after	10223	9627	13258	13926	13496	15848	17114
Program Experience							
WIA in prior two years	0.061	0.062	0.057	0.049	0.047	0.058	0.027
Comparison program participation in prior two years	0.181	0.166	0.258	0.351	0.330	0.441	0.628

As noted above, in three of the states the comparison program is U.S. Employment Service (ES) participants rather than Unemployment Insurance (UI) claimants. Table III.4 presents tabulations for states where the ES program is the comparison group. Perhaps the most important difference is in the economic position of ES participants relative to the WIA treatment groups. Whereas in Table III.1 prior earnings for the comparison group (three-quarters of them UI claimants) were two to three times the level of earnings for Adult program participants, ES earnings are only 50 percent higher. Furthermore, whereas ES earnings are much lower than Dislocated Workers earnings, UI claimants' earnings are approximately equal to Dislocated Worker earnings. Overall, it is clear that the ES serves a population that is closer to that of Adult workers than does the UI program. In the three ES states, 17 percent of WIA Adult program participants had no employment during the six quarters up to participation, whereas the number is 14 percent for the comparison group of ES participants. In the full sample, the Adult figure is similar to that reported in Table III.4, but for the comparison group, the number is only 4 percent.

Table III.4
Summary Statistics for WIA and ES: Program Participants in 3 States

	WIA Adult			WIA Dislocated Worker			
	Overall	No Training	Training	Overall	No Training	Training	Comparison Group
Sample size							
Unique individuals	14,715	7,394	7,321	11,288	5,741	5,547	884,894
Total quarters of participation	15,582	7,984	7,598	11,432	5,834	5,598	1,561,121
Demographic	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Male	0.366	0.389	0.343	0.432	0.423	0.442	0.543
Black	0.405	0.551	0.251	0.232	0.285	0.177	0.244
Hispanic	0.033	0.028	0.038	0.033	0.029	0.036	0.040
Age	34.017	36.001	31.931	42.590	44.163	40.951	36.898
Years of education	12.312	12.131	12.504	12.795	12.913	12.671	12.203
Employment							
Employment-employment	0.242	0.221	0.265	0.421	0.464	0.376	0.333
Employment-not employed	0.272	0.255	0.291	0.400	0.354	0.448	0.321
Not employed-employed	0.320	0.374	0.263	0.134	0.144	0.124	0.298
Not employed-not employed	0.166	0.150	0.182	0.044	0.038	0.051	0.136
Earnings second year prior	8919	9260	8564	28632	30544	26561	14406
Earnings in prior year	7332	7294	7371	26178	28068	24207	13408
Earnings following year	8713	9763	7609	12966	16111	9689	10343
Earnings second year after	11201	11185	11218	17118	19607	14525	11766
Program Experience							
WIA in prior two years	0.030	0.035	0.024	0.017	0.022	0.011	0.008
Comparison program participation in prior two years	0.486	0.480	0.493	0.559	0.531	0.587	0.604

It is unclear whether differences between the UI and ES comparison samples are important for this analysis. Both samples are very large, and, given that detailed earnings and employment information is available in both, good matches for most WIA participants are available in either file. As noted above, if sufficiently detailed information is used in the matching process, and if the comparison sample allows for good matches to be chosen, estimates are valid regardless of the comparison sample used.

IV. Matching Analysis: Details of Implementation

1. Matching Strategy

For the purposes of these analyses, an individual will be counted as a treated case in any quarter that the individual enters the WIA program. A very small number of individuals enter WIA more than once, and each WIA entry is treated as a separate case.[15]

Matched samples. For comparisons 1-2 listed in Table II.1, the comparison group comprises participants in an alternative program—either UI claimants or participants in the ES program. A comparison case is defined as a quarter in which an individual had contact with the comparison program, either filing a claim or receiving UI benefits (where UI claimants make up the comparison pool) or receiving some job search service (where ES participants make up the comparison pool). A particular individual in the comparison sample therefore contributes a case for every quarter of participation. Because details of program participation in prior quarters are controlled, each quarter contributed by a given comparison individual differs in terms of the attributes used for matching. For example, if an individual receives UI benefits in two consecutive quarters but has no prior UI experience, the case corresponding to the first quarter will match with WIA entries occurring in that quarter that do not have prior UI experience. The comparison case corresponding to the second quarter of UI experience will match WIA entries occurring in that second quarter where that WIA participant also received UI benefits in the prior quarter. Hence, a given comparison individual will offer multiple potential matches, reflecting differences in the flow of experience over time.

In most of these analyses, WIA cases are matched only to comparison cases with program participation in the same quarter, so all cases with participation in a given quarter correspond to a "hard match" group. All matching, including calculation of the propensity score, is undertaken separately for each such group. In some cases, the sample is too small for this approach, and different quarters are combined, but in each case, dummies for quarters are included as matching variables.

Where the comparison group is derived from alternative program participation (comparisons 1 and 2 in Table II.1), by definition treated and comparison cases have different experiences during the quarter of participation. Both are likely to be experiencing similar employment difficulties, but it is not possible to match them according to the specifics of their experience in that quarter. However, it is clear that matching by the details of prior experience may be of substantial value. In contrast, for comparison 3, since the estimate focuses on the impact of training, both treatment and control cases are WIA participants. The approach therefore controls not only for the prior experience in the alternative program but also for any experience in the current quarter.

Although a WIA entry is identified as a treated case even when the individual subsequently reenters WIA, any potential comparison case where the individual subsequently enters WIA is

[15] Measures of overall program impact therefore identify the incremental impact of program entry. In practice, the number of multiple entry individuals is so small that results are insensitive to the way they are treated.

omitted as a potential comparison case. Comparison cases are therefore analogous to controls in a random assignment study where the controls are precluded from participating in the program.

Matching variables. We fit a logit model using all the control variables to produce the propensity score, which is then used to match WIA participants to the comparison group. Theory assures that if the propensity score fully and properly identifies the probability that an individual is a treated case, then matching on this score will yield exactly matching distributions on all independent variables. In order to ensure that WIA entrants match prior recipients as closely as possible in all relevant ways, a highly flexible parameterization that includes nearly 100 independent variables is used.

Critical to these methods is that controls for the WIA and comparison samples be coded in a fully comparable fashion. In the case of demographic and geographic measures, policy differences in program data gathering and coding procedures were of concern. Every effort was made to assure comparability between treatment and comparison group variable coding and to restructure variables when necessary.[16] Since individual employment information was based on UI wage records that were matched with both WIA and comparison cases, it is clear that they are measured comparably. Similarly, information on prior experience in the comparison program (a control for both WIA and comparison program participants) was available in a symmetrical fashion for WIA and comparison group cases.

Standard errors. Conventionally, standard errors of propensity score matching estimates are obtained using bootstrap methods. With large samples such as those available to this study, it is not feasible to calculate bootstrap standard errors for all estimates (see, e.g., Lechner, 2001).[17] Following Abadie and Imbens (2006a), Imbens and Wooldridge (2008), and Imbens (2008), the analyses here use an analytical formula for calculating standard errors that is asymptotically correct.

However, in pilot analyses undertaken for two states, we compared several approaches that have been suggested for estimating standard errors. The first, recommended by Imbens and Woodridge (2008) and Imbens (2008), produces a conditional standard error, which provides an estimate of the variation in an impact estimate conditional on the independent variables.[18] Abadie and Imbens (2006a) suggest an approach for estimating the unconditional standard error, which provides an estimate of the variation in the impact estimate that would result if the sample were chosen repeatedly from the full universe, with values on independent variables varying from sample to sample.

When implemented in two states, these standard error estimates were very similar. Looking across all comparisons, in both states the mean difference was less than 1 percent. The absolute

[16] Differences between states are unavoidable and of less importance.

[17] Abadie and Imbens (2006b) show that, for matching estimates using a fixed number of matches, bootstrap standard errors are asymptotically biased. It is unclear whether these problems are relevant for the matching estimators used here, as there is no work indicating whether bootstrap standard errors for radius matching methods are consistent. See Imbens and Wooldridge (2008).

[18] In the implementation here, following Imbens (2008), it is conditional on the propensity score estimate.

difference was less than 1 percent in one state and 2.5 percent in the other state. It is clear that no substantive conclusion depends on the choice between these measures. In the analysis below, the conditional standard error is reported. This is referred to as the "Imbens standard error."

Although both of these standard error estimates are asymptotically correct given the assumptions on which they are based, the design of the analysis is expected to increase sampling error in ways they do not capture. Methods used in pilot analyses for estimating the bootstrap standard errors attempted to incorporate all sources of error that could influence the estimates.[19] On average, the bootstrap standard errors we calculated for two states were quite close to the conditional (Imbens) standard errors. In one state, the bootstrap standard errors averaged 4 percent higher and in the other state they averaged 7 percent higher. There was some variation in the difference across estimates. The absolute percentage difference between the bootstrap and Imbens standard errors was between 15 and 20 percent in these states. These experiments with bootstrap standard errors do not suggest that the results obtained using the Imbens standard errors are misleading. These conclusions are similar to those of Heinrich (2007), who reported only modest differences between alternative standard error estimates.

Radius Choice: Application of Cross Validation Methods. As described above, matching methods require that criteria be established to determine when cases are viewed as sufficiently close matches. The analysis here employs radius matching by the log odds of the propensity score. The primary concern is to assure that the matching procedure balances the independent variables for treated and comparison cases, so that impact estimates compare cases that are very similar on all independent variables. The discussion below on matching diagnostics describes specification searches used to attain balance. However, it is also possible that estimates may be improved by taking into account the dependent variable in choosing the radius size. Galdo, Smith and Black (2008) show that it may be useful to choose a radius to minimize the mean integrated standard error (MISE) for the comparison cases in a comparison sample weighted to correspond to the distribution of the treated sample. In essence, this approach trades off the benefits of averaging the dependent variable across comparison cases against the bias due to including cases that match less well.

Galdo et al. provide several methods for implementing their approach. We considered the MISE based on weighting the comparison sample by the inverse odds of the propensity score. Comparisons on two states were undertaken, using earnings in the fifth and tenth quarters after program entry as the dependent variables. MISE was estimated for radius values of 0.001, 0.003, 0.01, 0.03, and 0.1. For large samples, radius values larger than 0.1 were extremely time-consuming to estimate, so, in most cases, estimating MISE for larger radius values was not feasible. In general, the larger radius values produced smaller MISE values and also matched a

[19] Each replication was selected from the original samples of unique individuals, forming units of analysis from these individuals and undertaking the matching procedure on the resulting sample. Bootstrap standard errors were estimated both using the initial matching estimates and after undertaking bias adjustment (see Section IV.3). In each case, estimates were based on 25 replications, reflecting the computer-intensive nature of these replications. Across the two states chosen for these analyses, bootstrap standard errors were estimated for earnings and employment outcomes (up to 16 quarters after program entry), for each of the comparisons by gender (6 comparisons by 2 genders). Hence, over 300 estimates were obtained in each state.

larger portion of the treated cases. Based on these results, in the analysis that follows, the radius value was set to 0.1 as an initial default.

However, in choosing the radius for the analyses reported below, these considerations do not override the need that the approach successfully match on independent variables. Where a specification fails the balancing tests, the danger of bias requires that impact estimates be rejected, since there is no assurance that estimates obtained from an unbalanced match are appropriate for all dependent variables. Nonetheless, in deference to our findings reported above, larger radius values are chosen where tests suggest similar levels of balance.

2. Matching Diagnostics

The matching model specification was determined separately for each of the 12 states. Three dimensions were of concern: the appropriate radius value, the specification for independent variables, and whether exact matches would be required for quarter of entry.

In each state, the analysis began with a default specification. First, the sample was divided into subgroups by quarter of entry, producing a total of eight groups over the two program years, the period used in most states. In fitting the logit for the propensity score, all available independent variables were included in linear form along with selected interaction measures identifying particular patterns of employment and program participation over time. The radius was set so that any WIA entrant was matched to all comparison cases for which the log odds of the propensity score was within 0.1.

All entry subgroups were then combined and tests were performed to determine if the means for the independent variables for the treated cases differed from the matched comparison cases. In addition, tests of statistical significance were performed on differences between means for the squares of the continuous variables and selected interactions between the variables. In most states, there were between 70 and 100 initial variables that were tested, approximately 25 square terms, and up to 200 interaction terms. Approximately 5 percent of differences are expected to be statistically significant at the 0.05 level if there were, in fact, no real differences. A matching procedure was viewed as successful if fewer than 8 percent of the differences were statistically significant.

An additional concern was the proportion of WIA entries that were successfully matched. It may not be possible to find matching comparison cases for treated cases whose characteristics place them in the sparsely populated portion of the comparison case space. In general, if at least 90 percent of the WIA entries were matched, the matching proportion was taken as acceptable.

If the initial specification successfully balanced the sample and matched a sufficient proportion of treated cases, the specification was accepted. In such a case, although interaction terms were not included in the logit specification, given the success of the balancing tests, their inclusion was not necessary. If the specification failed the balancing test, interaction terms were added to the logit specification. If balance continued to be a problem, the radius size was reduced, first to

0.03 and then to 0.01, forcing cases to match more closely. This approach often reduced the number of treated cases that matched.

If fewer than 90 percent of the WIA entries were matched, subgroups based on quarters of WIA entry were combined, in some cases combining quarters into two groups representing program years and in other cases into a single group. The benefit of matching separately by entry quarter is that a WIA participant is always matched with a comparison case in the same quarter, so that prior and subsequent employment measures all apply to the same period. However, by requiring matches to occur for cases within the same quarter, this approach limits the number of potential matches for each entry, and may reduce the likelihood of a match.

In most cases, this search produced a specification that passed these balancing criterion and matched over 90 percent of the treated cases. However, in a few comparisons, it was not possible to find a specification meeting these criteria, and a specification was ultimately accepted that matched as few as 80 percent of the cases. In some cases, the balancing criterion was relaxed, so that specifications were accepted for which as many as one in ten mean differences were statistically significant. In a small number of instances where the number of treated cases was relatively large, even small differences in variable means were statistically significant. Specifications were then accepted for which more than one in ten differences was statistically significant so long as the mean absolute standardized difference was less than 0.02 (i.e., the average difference was less than 2 percent of the standard deviation of the variable).

In almost all cases involving comparisons 1 and 2 (comparing WIA participants with the comparison program), it was possible to find a specification that met the above standards. In a few cases, an acceptable balance was obtained only by reducing the radius to the point where too few treated cases were matched. In a substantial number of the states, matching failed for one or more of the groups considered in comparison 3. Comparison 3 focused on training, with the treated group defined as WIA participants receiving Training services and the comparison group as WIA participants receiving Core and Intensive services. Several of the smaller programs provided training to a large share of their participants, so that few comparison cases were available. For example, several programs had fewer than 500 male participants in the Dislocated Worker program who had received training, and fewer than 200 others. In such cases, even where those receiving training were not very different from others, no close matches were available for a substantial number of treated cases.

As noted in the last section, the analysis examines nearly 3 million individuals in the comparison program, contributing over 6 million matching quarters, and approximately 160,000 cases in the WIA Adult and Dislocated Work programs. The very large reservoir of comparison cases assures that close matches are feasible for most WIA participants, but differences in the distributions of variables identified in Table III.1 suggest that many comparison cases will not be appropriate matches. In fact, in the smaller state WIA programs, the vast majority of comparison cases do not match with any WIA case. More significant, however, is that a very large share of treated cases is matched to some comparison case.

For each comparison, Table IV.1 lists the total number of treated cases by gender and program, as well as the proportion of cases that were omitted from the analysis because comparison matches were not available. The second column identifies cases that were omitted because the comparison was not performed in certain states, due primarily to small sample sizes. The third column indicates the proportion of cases (in a state where the comparison was performed) that were omitted because they did not match. Both percentages are expressed relative to the total number of treated cases, so the sum of these percentages indicates the total loss of cases due to matching problems. For comparisons 1 and 2 (WIA participants vs. comparison group members), in most cases the total loss (omitted plus failed to match) is in the range of 4-7 percent, although it is close to 11 percent for female adult participants in comparison 1. The omitted cases are generally those with very minimal employment activity in the two years prior to participation, for which no similar comparison cases (usually UI claimants) were available.

Table IV.1
Matched Treated Cases Available for Estimation

Comparison	Treated Cases	Cases Omitted (Percent): State Analysis Omitted	Cases Omitted (Percent): Failed to Match	Cases Analyzed
Adult Program				
Females				
1. WIA vs Comparison Group	56612	4.9%	5.7%	50657
2. WIA Core/Intenstive vs.Comparison Group	38671	0.0%	6.4%	36190
3. WIA Training vs. WIA Core/Intensive	17941	24.4%	5.6%	12564
Males				
1. WIA vs Comparison Group	40940	0.0%	5.0%	38894
2. WIA Core/Intenstive vs. Comparison Group	31041	0.0%	5.6%	29292
3. WIA Training vs. WIA Core/Intensive	9899	43.2%	5.5%	5087
Dislocated Worker Program				
Females				
1. WIA vs Comparison Group	33174	0.0%	3.9%	31876
2. WIA Core/Intenstive vs.Comparison Group	22190	0.0%	3.7%	21378
3. WIA Training vs. WIA Core/Intensive	10984	25.0%	5.3%	7655
Males				
1. WIA vs Comparison Group	30915	0.0%	4.0%	29690
2. WIA Core/Intenstive vs. Comparison Group	21704	0.4%	3.6%	20840
3. WIA Training vs. WIA Core/Intensive	9211	32.4%	5.9%	5676

In the case of comparison 3, the total proportion excluded is much greater, nearly 50 percent for males in the Adult program and over 30 percent in others. This reflects the fact that states with high proportions of individuals receiving training provide very few comparison cases that are available for analysis. There are simply too few WIA participants without training to allow valid matches for up to half of those receiving training.

The large proportion of omitted cases in comparison 3 influences estimates in two ways. First, by reducing the sample size, the precision of impact estimates is reduced. This is reflected in standard errors, which will be increased. Second, and more important, by omitting certain states, it calls into question whether the resulting estimates are representative of the 12 states in the sample. If omitted states have systematically different impact estimates than those included, these estimates may be misleading. The extent of this problem cannot be determined, but the omitted states do not appear to be selected in any clear way—except that they represent small states with relatively large proportions trained. Still, of the seven states with high training levels, in each of the comparisons reported, at least three provide valid data.

3. Presentation of Results

Bias adjustment was also applied to impact estimates that are reported, although, given that the matches are very close, it makes very little difference.[20]

In each case, estimates are obtained for the WIA program's impacts on average earnings and level of employment in the 16 quarters following program impact. All earnings have been adjusted for inflation to correspond to the first quarter of 2005. Once impact estimates specific to a state are obtained, the mean across states is estimated by weighting the estimate for a given state by the number of participants who were matched in that state. The resulting weighted mean provides an estimate of the average impact for WIA participants who entered the program during the periods specified in the states of this analysis. The estimate omits individuals for whom acceptable matches were not found, i.e., participants whose characteristics differ from those of any comparison program individual.

For each state, estimates were also obtained for the difference between the treated sample and the matched comparison sample on outcome variables in *prior* quarters. We focus on earnings and employment 10 quarters prior to program entry, and earnings and employment 16 quarters prior to program entry. These estimates provide a specification test for our model, since, if the program "effect" on prior measures is significantly different from zero, this implies that participants are different in ways not captured by the matching criteria. Estimates of the impact on subsequent outcomes are therefore suspect. In each case where estimates are discussed, estimates on prior measures are also presented to provide an indicator of possible bias in estimates.

In the case where the model fails such a specification test, one approach is to calculate a "difference-in-difference" estimate, which can be obtained by subtracting the prior earnings effect from the conventional estimates. As noted above, such estimates can be unbiased if differences between treated and control cases are due to fixed differences. Such estimates do, however, rely on strong assumptions that are difficult to test.

[20] Following Abadie and Imbens (2006a), we fit a linear model in the comparison sample, and then use coefficients estimated in this sample to adjust for any differences in independent variable means that exist between the treatment and matched comparison cases. Where sample sizes were very small, the bias adjustment was not possible, and estimates reported do not include the bias adjustment.

Associated with each state impact estimate is an Imbens (conditional) standard error, which is combined in the conventional way to form the standard error for the weighted average across states. For the main estimates, confidence intervals are provided, allowing conventional significance tests to be performed.[21]

V. Adult Program: Impact Estimates

This section presents estimates of the average impacts of the WIA Adult program across individuals; that is, the difference in average earnings or employment following program participation as compared to the average if these individuals did not participate in the program. The comparison group provides an estimate of what outcomes would be if participants had not participated. This estimated effect of participating is calculated as measured earnings or employment in a given quarter for WIA entrants minus that for individuals who are similar in terms of demographic characteristics, prior employment (previous eight quarters), and prior program experience.

1. Overall Program Impacts

The first set of analyses focuses on how earnings and employment for all Adult program participants are affected by the program. WIA's Adult program focuses on disadvantaged workers. Table III.1, above, shows that participants are, in fact, disadvantaged relative to both those in WIA's Dislocated Worker program and to those in the two comparison programs. This is reflected in differences in educational attainment, race, and gender, but most importantly in differences in prior employment success.

The first step in producing aggregate impact estimates is to obtain separate estimates of program impact for quarters 1-16 following program entry for each of the 12 states for which data are available. Table V.1 provides summary statistics for estimates. For each state, the simple average of impact estimates for quarters 1-5 is presented, as well as the average for quarters 11-16.[22]

There are substantial differences in average impacts in the first five quarters, with two states registering small or negative impacts for both genders, and two states yielding estimates near or exceeding $1000 for both genders. But the finding of a substantial positive statistically significant effect is widespread. Of the 24 estimates (12 states x 2 genders), 18 are positive and statistically significant, with the smallest at $208 per quarter, and only one is negative and statistically significant. The same basic pattern applies for quarters 11-16, although effects are larger for most states.

---

[21] The graphs present confidence intervals based on values two standard errors below the point estimate and two standard errors above, corresponding to the 95.5 percent confidence interval.

[22] Recall that data use agreements preclude revealing state identities. States are ordered by the size of the average effect in quarters 11-16 for females. For some states, impact estimates are only available for a subset of quarters 11-16, and, in these cases, averages are based on available quarters. In three states, no estimates after quarter 10 are available and the quarter 10 estimate is presented.

Table V.1
Adult Program Treatment Effect on Quarterly Earnings by State: WIA versus Comparison Program

	Females		Males	
State	Quarters 1-5	Quarters 11-16	Quarters 1-5	Quarters 11-16
1	-140	-165	-187	235
2	208*	168*	69	-35
3	528*	396*	452*	290
4	409*	418*	354*	-12
5	302*	588*	475*	835*
6	302*	624*	359*	483*
7	476*	909*	120	197
8	-129*	949*	38	371*
9	241*	1094*	360*	964*
10	721*	1198*	892*	840*
11	1187*	1283*	1233*	892*
12	908*	1426*	1203*	1211*

*Statistically significant at the 0.05 level.
Note: Average effects for specified quarters. Where estimates are not available for a given state, the average is calculated on available quarters. In the case of three states, estimates are not available for quarters 11-16, and the reported estimate applies to quarter 10.

Although the results in Table V.1 show that there is substantial variation across states that cannot be explained by sampling error, sampling error is still large in many cases. Figures V.1 and V.2 provide estimates for women and men, respectively, combining the estimates from all 12 states. The horizontal axis extends from 1 to 16, identifying the quarter following program entry. The vertical axis is in dollars, indicating the difference between average earnings in a quarter for the WIA Adult program participants and matched comparison program participants, which is the primary measure of program impact.[23] Also on the graph are dashed lines that show the confidence interval for each estimate. The lower dashed line subtracts twice the Imbens standard error from the estimate, and the upper dashed line adds twice the standard error. Also presented in this figure are the estimates of "impact" on earnings 10 quarters prior and 16 quarters prior to program entry. This measure can be interpreted as providing a specification test for the direct estimate, and its value can be used to calculate a difference-in-difference estimate.[24]

[23] For some states, data are not available for more than 10 quarters, so the number of states on which estimates for subsequent quarters are based is variable. These estimates have been adjusted so they are not influenced by the relative impacts in the particular states that contribute data. However, results for these later quarters will be less reliable. The confidence intervals have been modified to reflect uncertainties associated with the adjustment.
[24] See Section IV.3

Figure V.1
Adult Program Treatment Effect on Quarterly Earnings for Females, WIA versus Comparison Group

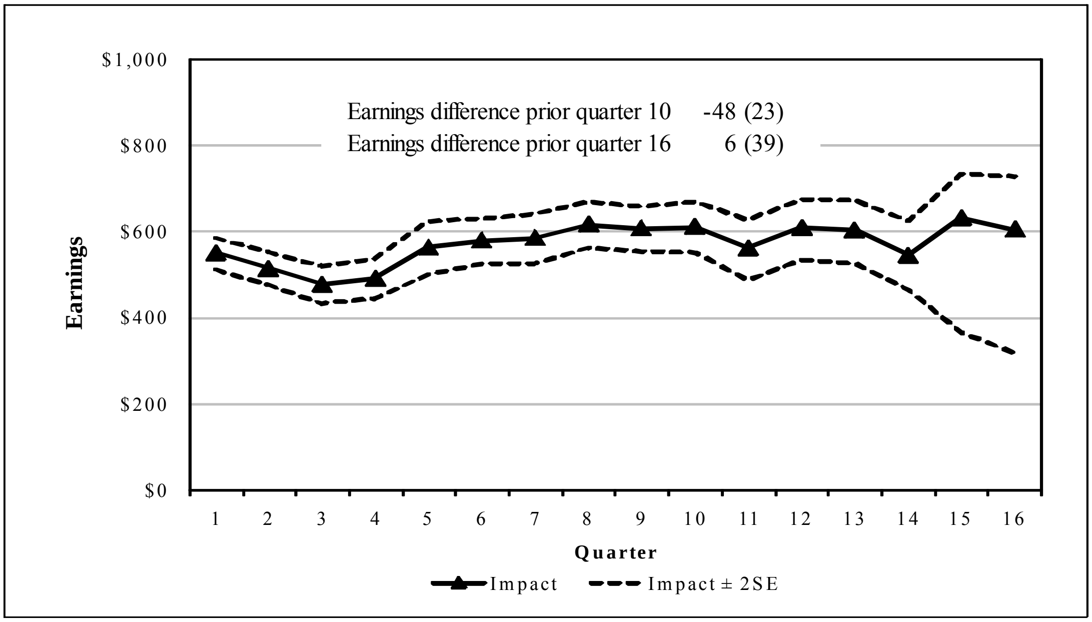
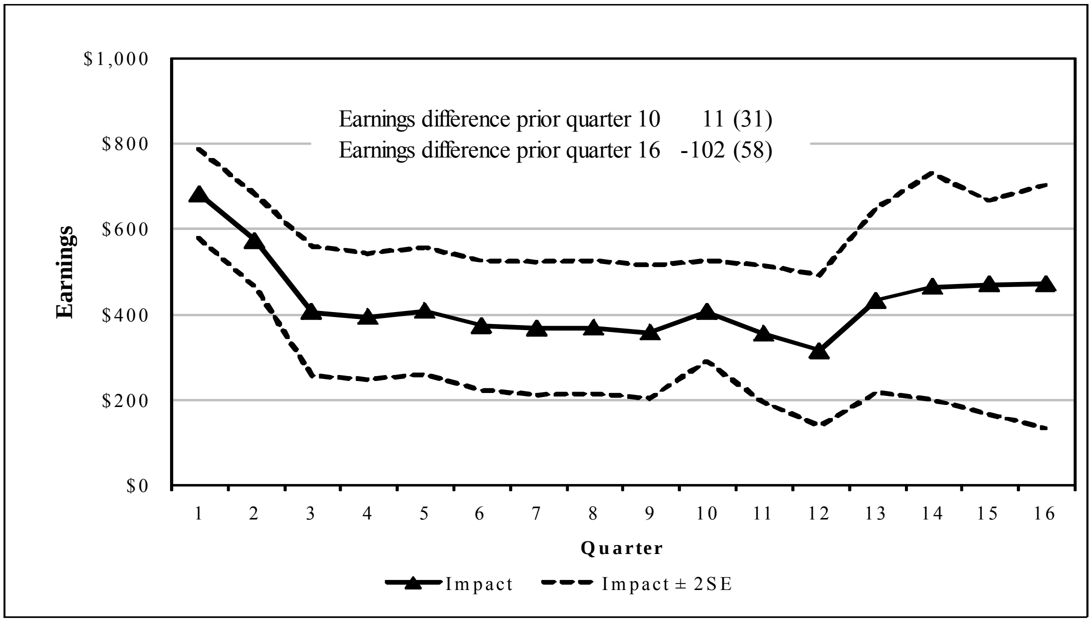


Figure V.2
Adult Program Treatment Effect on Quarterly Earnings For Males, WIA versus Comparison Group



The estimates reported in the figures imply that, for both genders, participants earn between $400 and $700 more per quarter than comparable individuals in the comparison program. For women, the return over most of the 16 quarters is between $500 and $600 per quarter, whereas for men there is a decline in the first three quarters in this return, with the level settling in the range of $400. Given that average quarterly earnings for WIA Adult participants are between $2000 and $3000 in the first year after program entry, ultimately approaching $4000 after four years, the returns are in the range of 10 to 30 percent.

Figures V.3 and V.4 provide analogous estimates for employment. The method used to obtain estimates reported in these figures is identical to that for earnings, except that here the proportion employed in the quarter (identified as having received positive earnings in the quarter) is taken as the dependent variable. Each value can be read as the difference between the employment rate for Adult program participants and the matched comparison cases. For example, the value is 0.13 for females in the first quarter after participation, implying that the employment rate for participants is 13 percentage points higher than that for matched comparison cases. The basic pattern of results is quite similar to that for earnings. In particular, female participants' levels of employment—relative to the comparison group—decline from 13 percent to about 8 percent within a year and ultimately to about 6 percent. Male impacts are one or two percentage points lower, although, given the confidence interval, it is unwise to place much emphasis on the observed differences.

As noted in Section III, there are substantial differences in the proportion of individuals receiving training across the state programs. It might be expected that patterns of returns would be different for programs with different levels of training. First, the total resources per participant will be appreciably higher in such states, so long-run program impact could be higher if more intensive services produce greater returns. Second, a large share of the value may well occur with a greater lag, since training benefits presumably accrue over an extended period. Figures V.5 and V.6 provide impact estimates for the seven states that provide training to more than half of their participants. In these states taken together, 68 percent of Adult program participants receive training.

The initial returns—during the first several quarters after program entry—in these seven states are very similar to the aggregate for all states (as seen in Figures V.1 and V.2). In contrast, however, among females, growth in earnings up through the first 10 quarters is notable, with the ultimate quarterly earnings increment reaching $1,100. Although there is no growth in returns for men, neither is there a decline in the first 10 quarters, as occurs for the whole sample. In short, there is at least weak evidence suggesting that high-training states produce benefits that endure longer. The basic pattern for employment impacts is similar to that for earnings, and so that graph is not presented.

Taken at face value, these results imply that the program has strong and substantial impacts with little or no lag. These could reflect aggressive actions by program staff to help workers obtain employment initially, with training assuring benefits that accrue over an extended period. Skeptics will argue, however, that the findings of such large initial impacts call into question the appropriateness of the comparison group and ultimately the validity of the results. With most

Figure V.3
Adult Program Treatment Effect on Quarterly Employment for Females, WIA versus Comparison Group

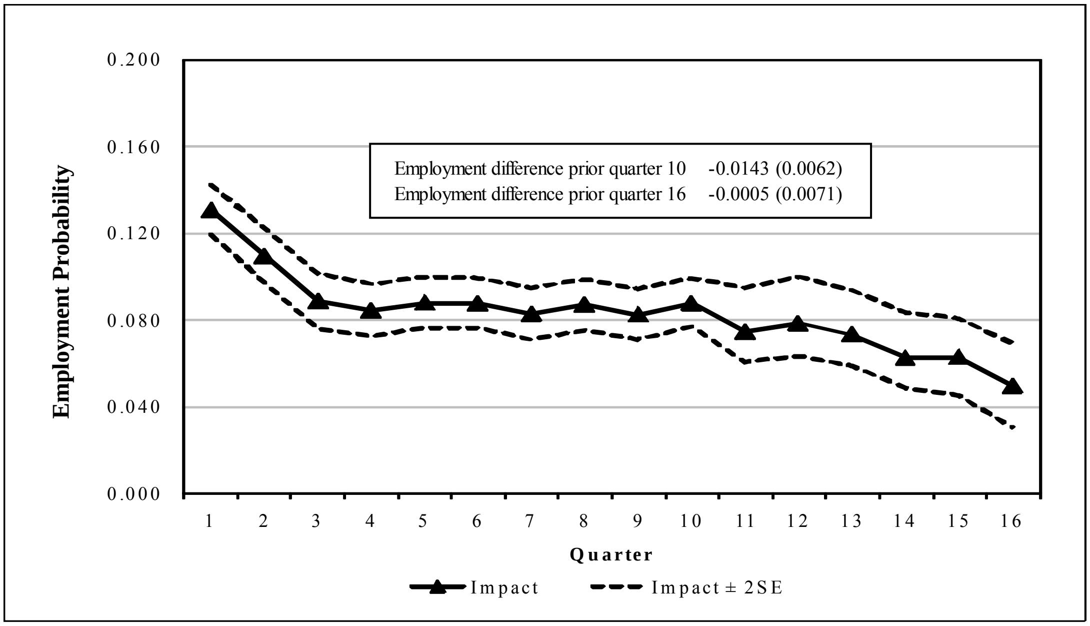


Figure V.4
Adult Program Treatment Effect on Quarterly Employment for Males, WIA versus Comparison Group

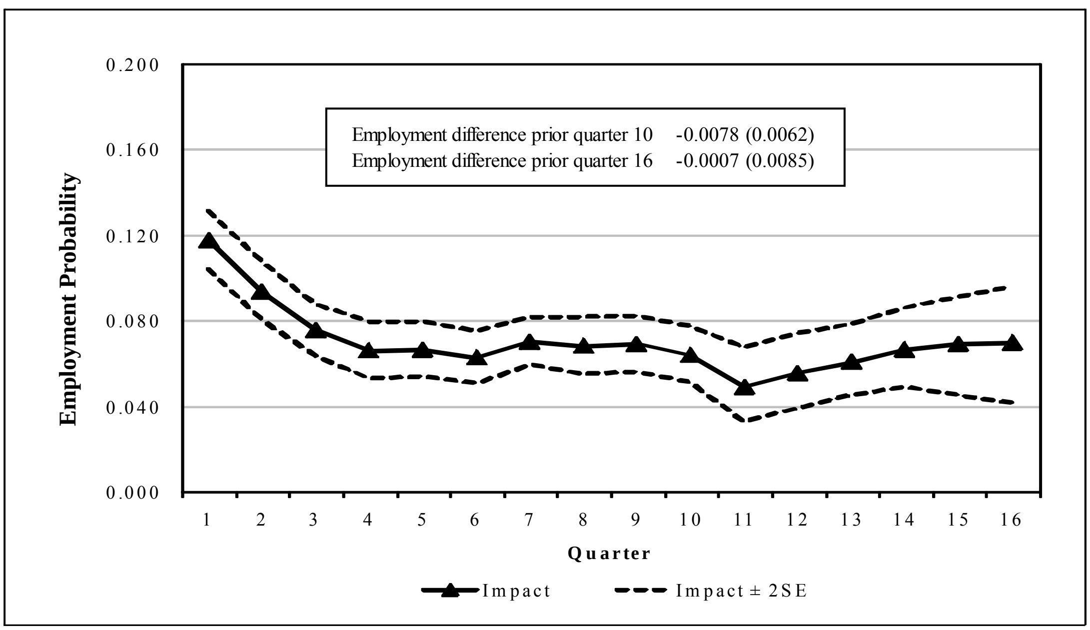


Figure V.5

Adult Program Treatment Effect on Quarterly Earnings for Females, WIA versus Comparison Group in 7 High-Training States

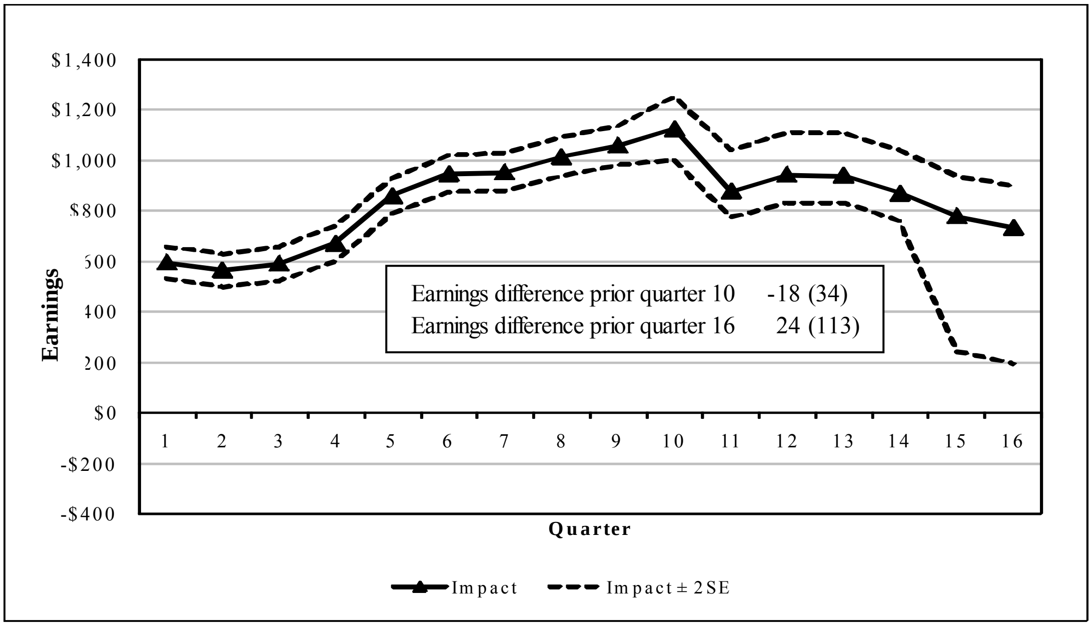
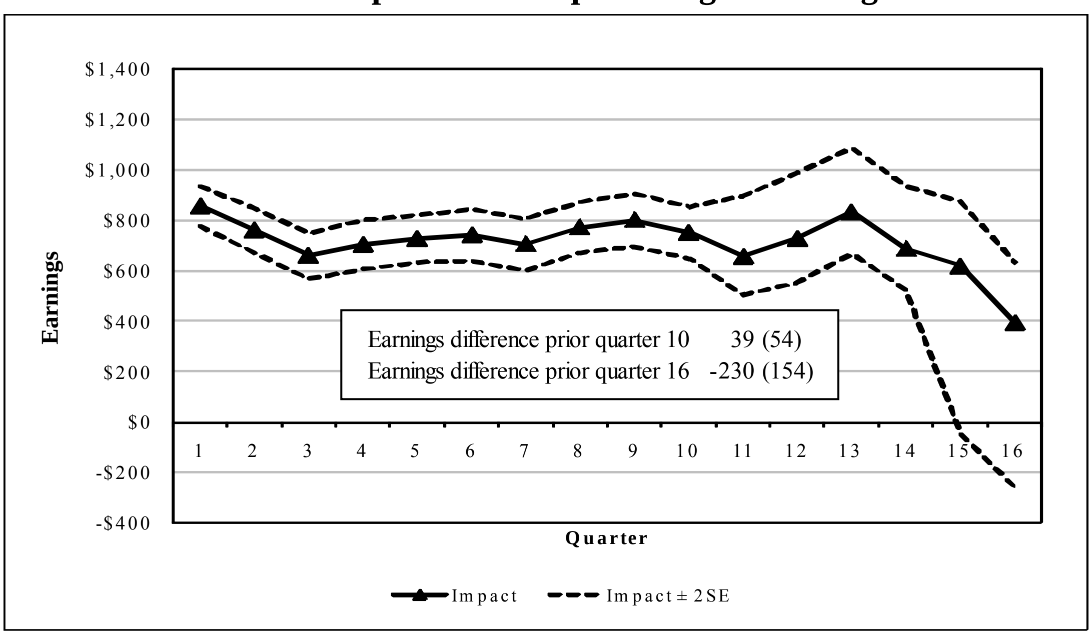


Figure V.6

Adult Program Treatment Effect on Quarterly Earnings for Males, WIA versus Comparison Group in 7 High-Training States



training programs, initial participants experience reductions in earnings as they engage in training activities that supplant employment that would otherwise occur, so called "lock-in" effects (Dyke et al., 2006; Hotz et al., 2006). In these data, the median person exits the program around three quarters after entering. Hence, it is clear that estimates of program impact in the two quarters after entry identify a point in time when most people are still participating in the program.

In order for selection to cause these results, it must be the case that WIA participants have unmeasured attributes that make them more likely than those in the comparison program to obtain employment or higher earnings. For example, staff admission criteria or individual self-selection would need to select entrants who were appreciably more likely to obtain employment than other individuals with similar characteristics, employment and program participation histories.

A first test for selection is provided by analyses that predict prior earnings. As noted in Section IV, although controls are included for earnings in the eight quarters prior to entry, if there are stable factors that improve the employment prospects for treated cases relative to matched comparison cases, earlier earnings would be higher for the WIA cases. In each figure, we report the difference in earnings or employment between treated and comparison cases for measures applying to the tenth and sixteenth quarters prior to entry (standard errors are in parentheses). In the case at hand, these difference estimates show that earnings and employment are *not* higher for WIA participants; in most cases, differences are small. The largest difference is for male WIA participants 16 quarters earlier, for which it appears that WIA participants have earnings that are about $100 *below* those of the comparison group. Such a difference would tend to downwardly bias estimates; estimates from a difference-in-difference model would produce a program impact estimates that was $100 greater. For males in the seven high-training states, earnings are $230 lower. It is therefore clear that if selection is causing spurious positive impact estimates, selection is unlikely to be based on stable individual characteristics.

One alternative explanation would be that there are transient differences between WIA participants and others. The comparison group members receiving unemployment compensation may include a substantial portion of individuals who are not seeking employment. UI recipients classified as awaiting recall are not required to search for employment, and many others may have little interest in getting a job—despite formal requirements—until benefits are about to expire. According to this view, those obtaining UI benefits are in a phase where their short-term employment levels are expected to be depressed, perhaps in part reflecting the incentives due to UI benefits, which would discontinue if a job was obtained. WIA participants, in contrast, have chosen to select into a program with the purpose of improving their employment prospects.

If the problem stems from the differences between WIA participants and UI claimants, it might be expected that such differences would be less important for the other comparison group, those seeking ES services. Although most UI claimants are required to register for ES services, those awaiting recall are exempt from this requirement, so the ES sample removes one group whose interest in employment may be modest. Since any individual seeking support for employment search can obtain ES services, this sample is expected to include self-motivated job searchers.

Figure V.7
Adult Program Treatment Effect on Quarterly Earnings for Females, WIA versus ES Participants in 3 States

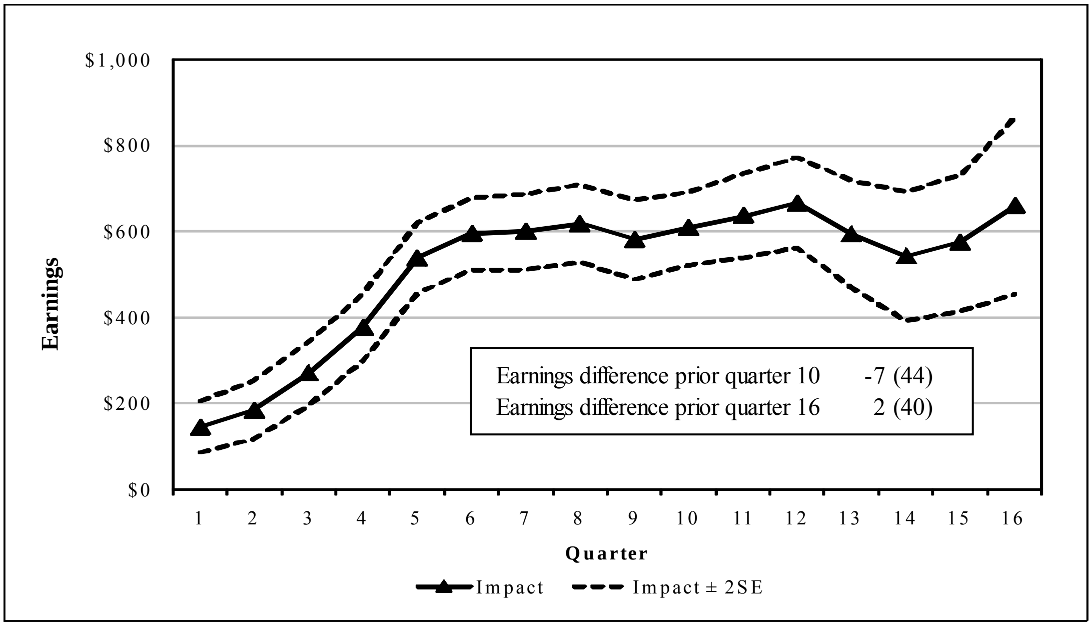
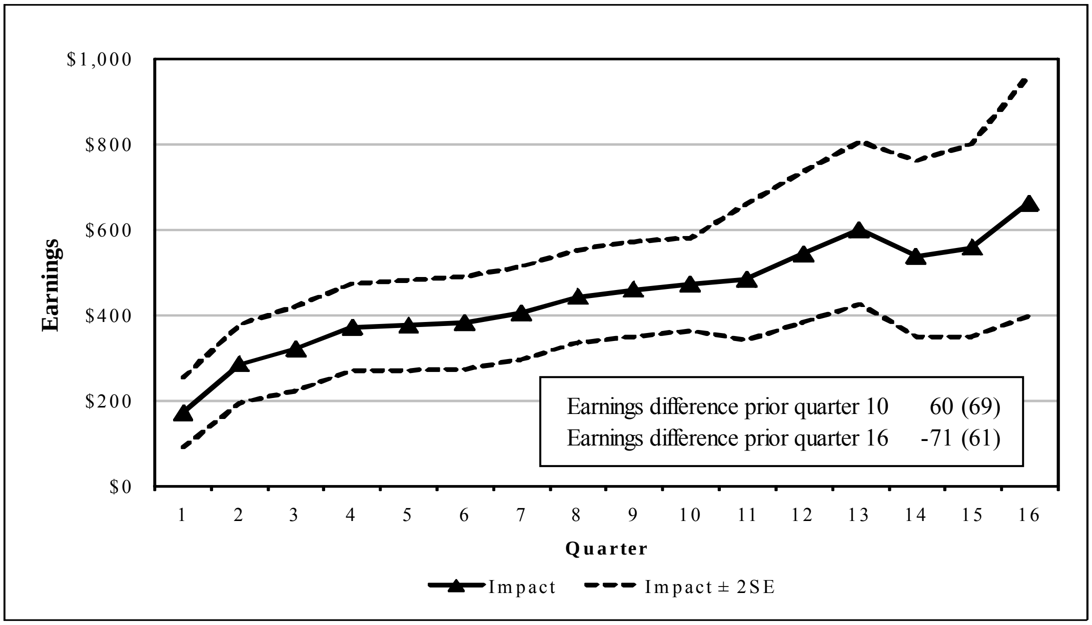


Figure V.8
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA versus ES Participants in 3 States



Figures V.7 and V.8 provide earnings impact estimates for the three states where ES recipients form the comparison group. There are important differences in these patterns as compared with the full sample of states. Perhaps most notable, impacts in the first few quarters after entry are somewhat smaller, in the range of $200 for both men and women. There is a fairly steady increase in program impact up through the last quarters.

These results therefore support the view that the large impacts on earnings and employment in the quarters immediately after WIA entry could be at least partly due to selection differences between WIA and the UI claimant comparison group. Of the nine states for which UI claimants are the comparison group, initial program impact is similarly small in only two of them.

In conclusion, estimates of overall program impact are positive in almost all states, although variation across programs is substantial. While the patterns suggest the possibility that estimates may be partly spurious, none of the selection explanations considered would fully explain these estimates.

2. Impact Estimates for WIA Core and Intensive Services

This subsection reports estimated impacts for WIA Adult program participants who did not receive training. The lowest level of services provided by WIA is Core services, which are generally similar to the services available without restriction as part of the ES program. The next level of services is termed Intensive services, which may involve short courses—usually of no more than a few day's duration—as well as assessment and counseling provided by WIA staff. Although some of these services are available as ES services, WIA Intensive services generally involve greater staff time and extend over longer periods. Such services are clearly distinguishable from training services, which require much more extensive time investment, both by the center and the participant.

It is natural to ask why recipients of Core and Intensive services would experience benefits if closely-related services are readily available outside WIA. Because of the way that the program is structured and the interaction of this structure with the performance standard system, WIA participants may receive a level of attention from staff that is not fully reflected in the service profile. Performance of exiting participants is a primary way in which the agency is evaluated, and staff are expected to be made conscious of the importance of client employment outcomes. As a result, in contrast to ES participants,[25] who usually have relatively weak attachment to the program, those in the WIA program are more likely to receive continuing staff attention until they exit.

Figures V.9 and V.10 show that, for WIA participants who don't obtain training, the initial earnings increment in the quarter following program entry is approximately $550 for women and nearly $700 for men. These impact estimates are essentially the same as those for all WIA participants. However, following the initial quarter, the results differ somewhat. For women,

---

[25] In the ES data available in this study, entry and exit from the program are not meaningfully defined.

Figure V.9
Adult Program Treatment Effect on Quarterly Earnings for Females, WIA Core/Intensive versus Comparison Group

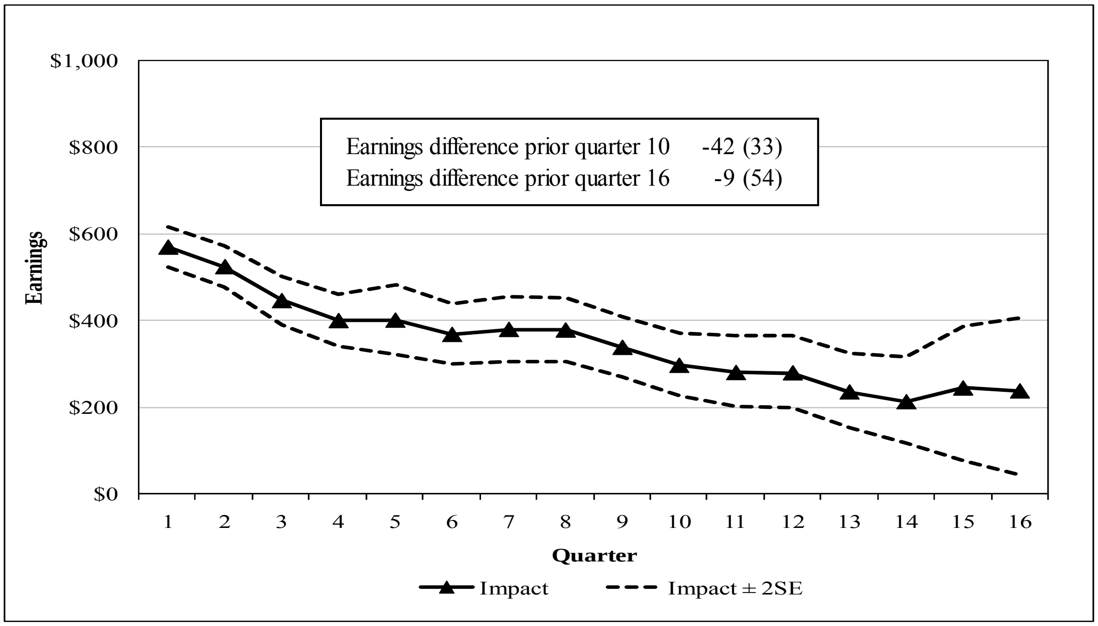


Figure V.10
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA Core/Intensive versus Comparison Group

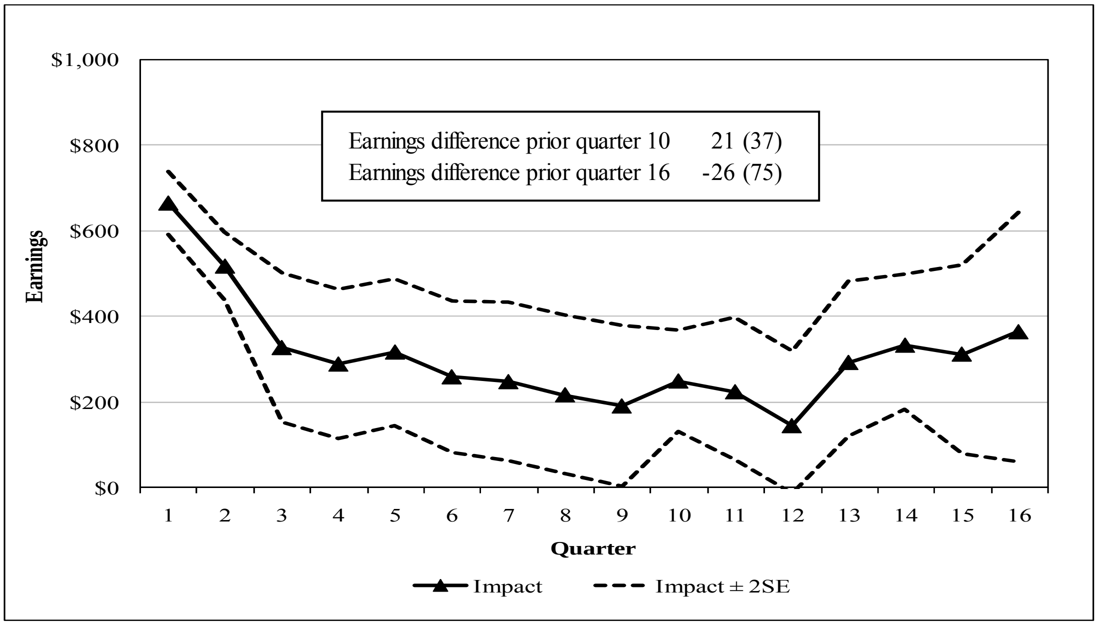


Figure V.11
Adult Program Treatment Effect on Quarterly Employment for Females, WIA Core/Intensive versus Comparison Group

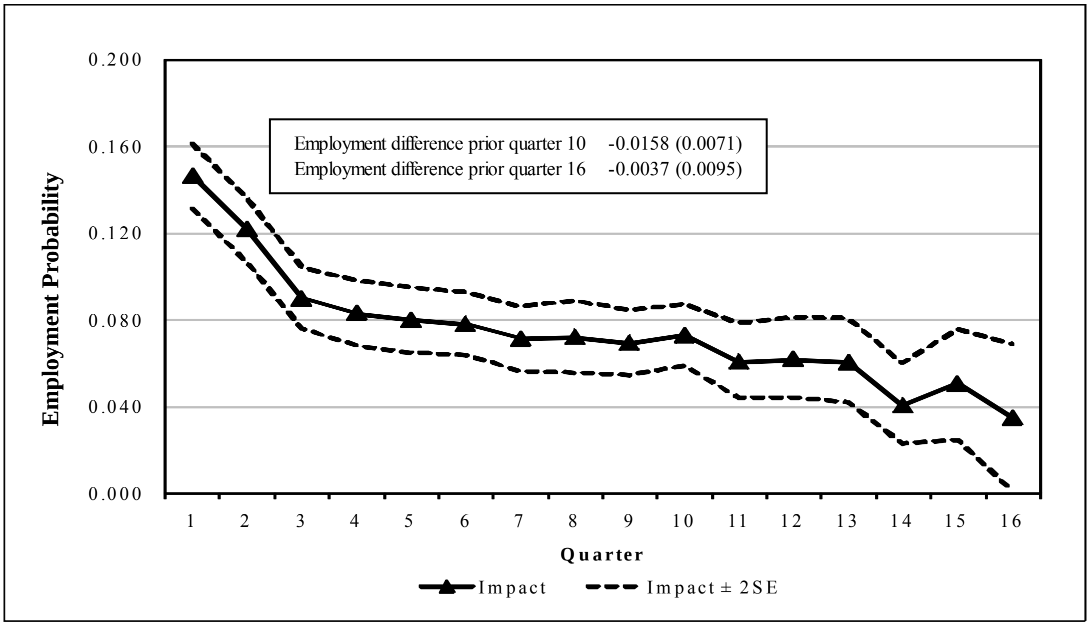
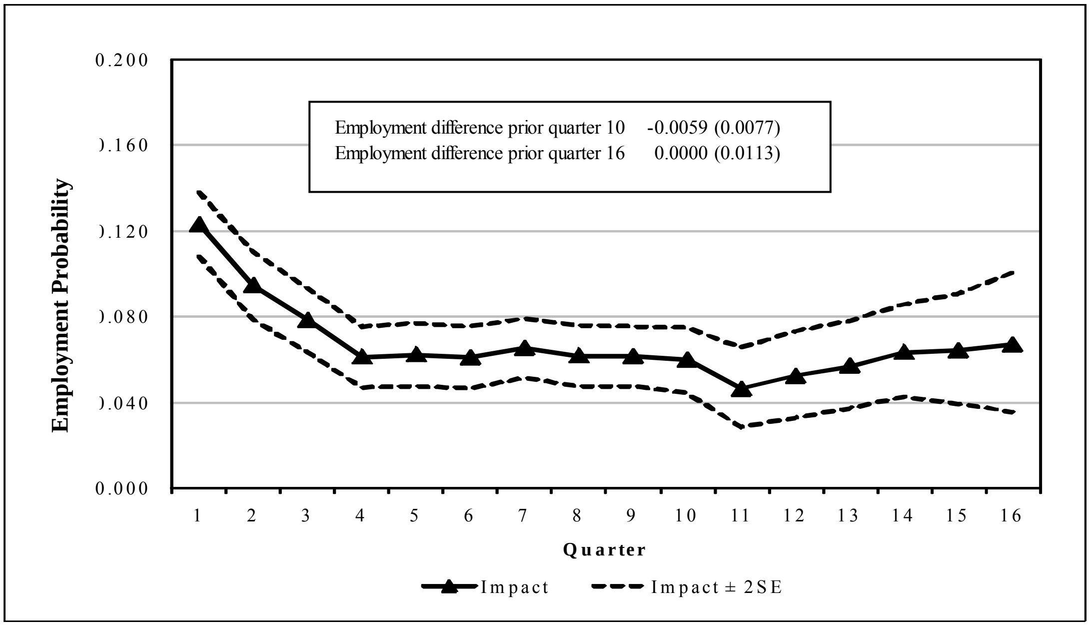


Figure V.12
Adult Program Treatment Effect on Quarterly Employment for Males, WIA Core/Intensive versus Comparison Group



the impact declines continuously, approaching $200 by the end of the period, in contrast to the steady or growing number observed earlier. For men, the impact declines quickly, with the earnings increment after quarter 10 erratic but mostly in the range of $200-$300. Nine of the 12 states exhibit declining or constant impact patterns—similar to the aggregate graph. Employment effects are presented in Figures V.11 and V.12, and the basic pattern of results is clearly the same.

These results are, of course, consistent with the view that individuals who do not receive training receive effective short-term counseling that enables them to gain an immediate advantage in the labor market but that this advantage declines over a relatively short period. Certainly, this kind of return would be substantial, especially given that a benefit of $200-$300 per quarter continues for an extended period. The estimates are also consistent with selection into the program, in which individuals whose immediate employment opportunities—perhaps as judged by a counselor—are particularly good (independent of their employment histories) and are admitted into the program. In this case, earnings would not reflect program impacts but rather counselors' success at choosing applicants who can find jobs.

A third possibility is that selection does not occur upon entry into the program but that those who obtain high-paying jobs choose not to pursue training. In this case, the high earnings in the initial quarters after entry are not an indicator of program impact but rather reflect selection after entry into the program. Although the estimates presented in Figures V.9 through V.12 would not indicate causal impacts, there need be no bias in the impact estimates presented in Figures V.1 through V.4.

The estimate of the "effect" of program participation on prior earnings for those who do not receive training is essentially zero, that is, treated individuals' prior earnings correspond to those of the comparison group. Hence, as was the case in the prior section, any selection effects do not appear to be occurring on the basis of stable characteristics that are reflected in prior earnings. Difference-in-difference estimates of program effect would be essentially the same as those reported here.

Differences by the training policies of states may be of interest. Figures V.13 and V.14 present results for the seven states that are more likely to provide training. The decline in program impact is less immediate in these states. For women, there appears to be a $600 increment in quarterly earnings for participants extending for more than two years after program entry; men experience an increment of about $400. It would appear that states that train a larger share of their applicants produce longer-lasting benefits for those individuals who do not receive training services.

Finally, the observed impact estimates for the three states that use ES participants as the comparison group (not shown) are much more modest than for the sample as a whole—especially in the initial quarters after program entry—in the order of $300-$400, with little trend over the first 10 quarters. This supports the view that the positive effects observed for the full sample may result from selection effects.

Figure V.13
Adult Program Treatment Effect on Quarterly Earnings for Females, WIA Core/Intensive versus Comparison Group in 7 High-Training States

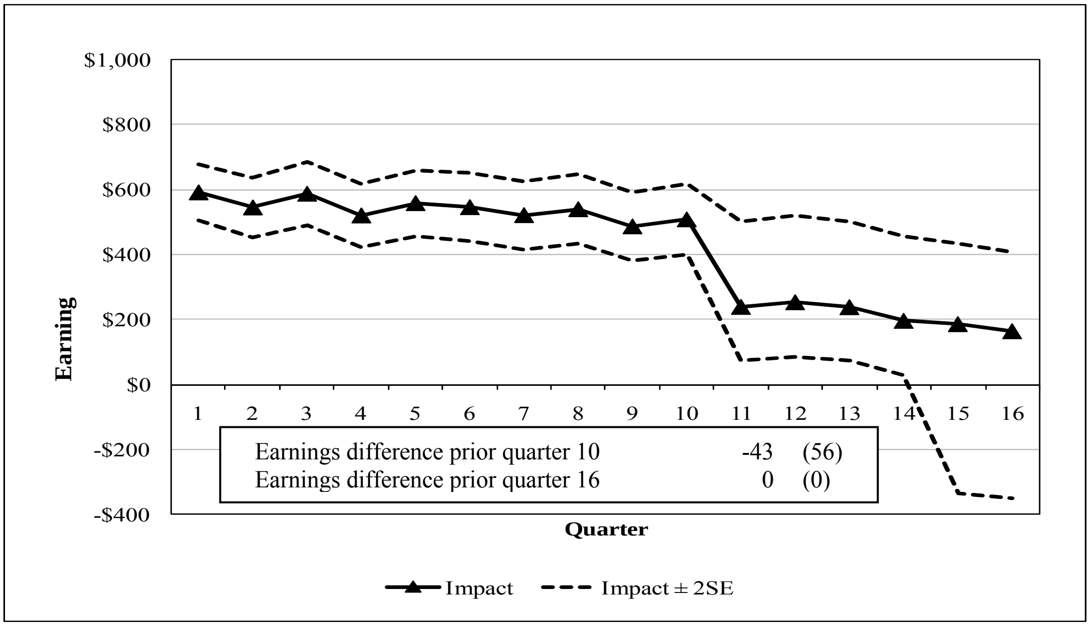
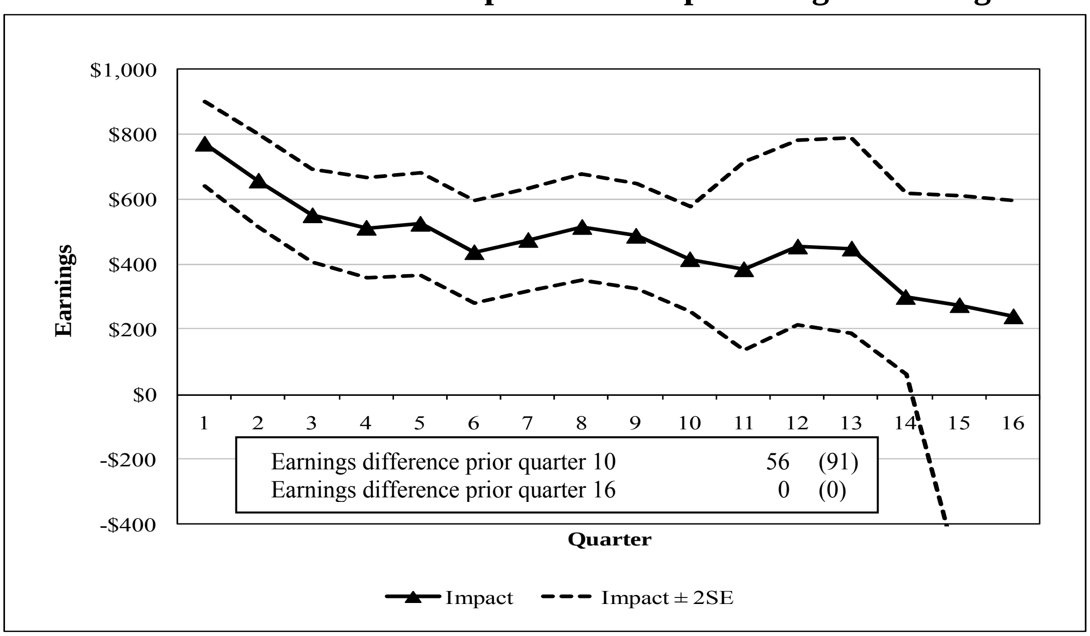


Figure V.14
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA Core/Intensive versus Comparison Group in 7 High-Training States



In sum, the estimates for the full sample of WIA Adult participants receiving only Core or Intensive services are very large. Given the modest extent of the services received by participants and the observed decline in impact over time, it is reasonable to suggest that these are partly due to selection effects. The more modest impact estimates for the ES states tend to support this interpretation. It is therefore not appropriate to take these large estimates at face value. An educated guess would be that gains in earnings are unlikely to be more than $200 per quarter over the four years following program entry. Nonetheless, even such modest impacts of Core and Intensive services for Adult participants may be sufficient to justify their fairly low costs.

3. Impacts of Training

The heart of WIA services is in the basic and vocational skills training provided to individuals. Although a variety of training opportunities are widely available outside of WIA, for a large share of WIA Adult participants, the alternatives available are inferior or costly. It is clear that acceptance into WIA alters the type and extent of training an individual ultimately obtains. If training is of value in the job market, entry into the program would be expected to influence earnings or employment.[26]

Figures V.15 and V.16 present impact estimates of training based on comparison 3, where the comparison group is individuals in the same WIA program who did not receive training. Earnings impact estimates for females imply a $200 decrement in the first quarter after program entry, as would be expected if time in training limited employment options initially. Earnings, however, catch up to others three or four quarters later, with a positive increment over $800 by the end of 10 quarters. In contrast, males who receive training appear to have positive returns—in the range of $200 immediately after entry—with the increment remaining in the $500-600 range for the next 10 quarters.[27] The initial values would appear to be implausible, given that those who receive training are in the program for an average of over three quarters; relatively few individuals would have exited the program within two quarters.

Figures V.17 and V.18 show that estimates for impacts on employment are, however, more consistent with expectations. For females, employment is about 5 percentage points lower for those receiving training, and the employment rate for participants only catches up four quarters after entry. By the tenth quarter the increment is in favor of training recipients by about 5 percentage points. For males, the pattern is quite similar, although the increment is close to zero for six or seven quarters after entry. The ultimate increment is slightly smaller, in the range of 3-4 percent. Interestingly, the pattern of results does not vary substantially by whether states train a large share of their participants. Nor are results substantially different for ES states. (These results are not presented.)

---

[26] There is no way to determine the extent to which comparison group members receive training or related services outside the WIA program. Reported impact estimates are therefore incremental relative to services received by the comparison group, some of whom undoubtedly receive training.
[27] The very high estimates in quarters 15 and 16 should be discounted given the large standard errors.

Figure V.15
Adult Program Treatment Effect on Quarterly Earnings for Females, WIA Training versus Comparison Group

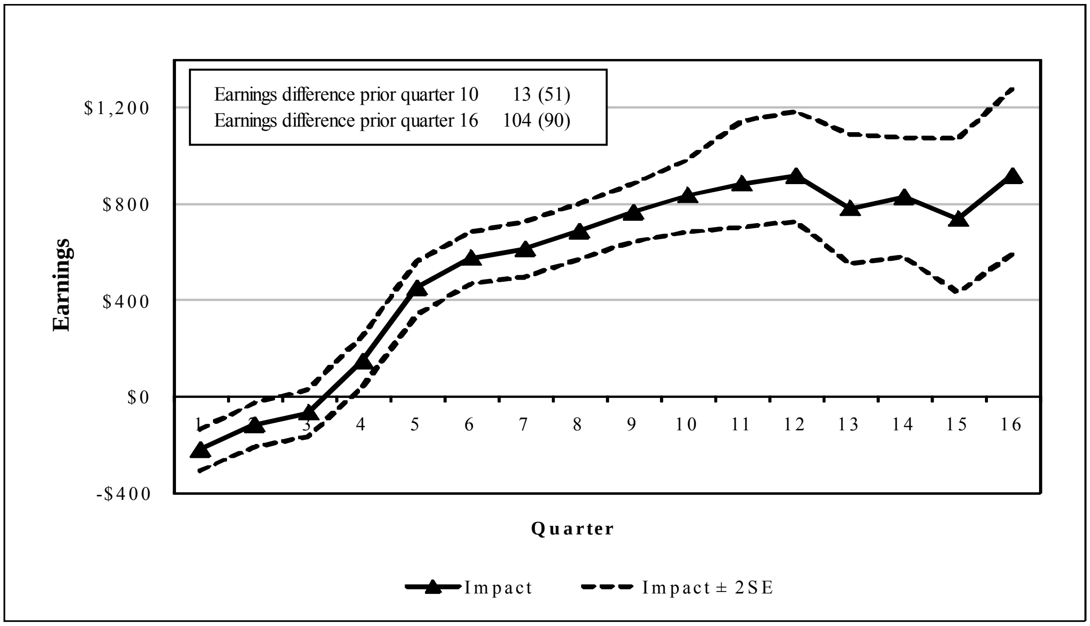


Figure V.16
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA Training versus Comparison Group

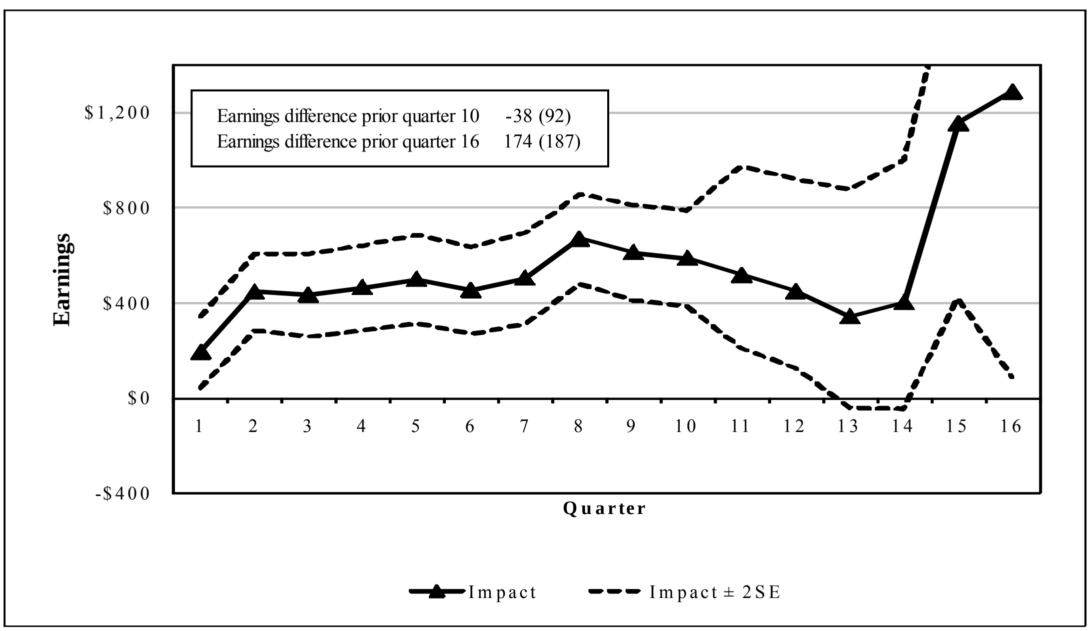


Figure V.17
Adult Program Treatment Effect on Quarterly Employment for Females, WIA Training versus Comparison Group

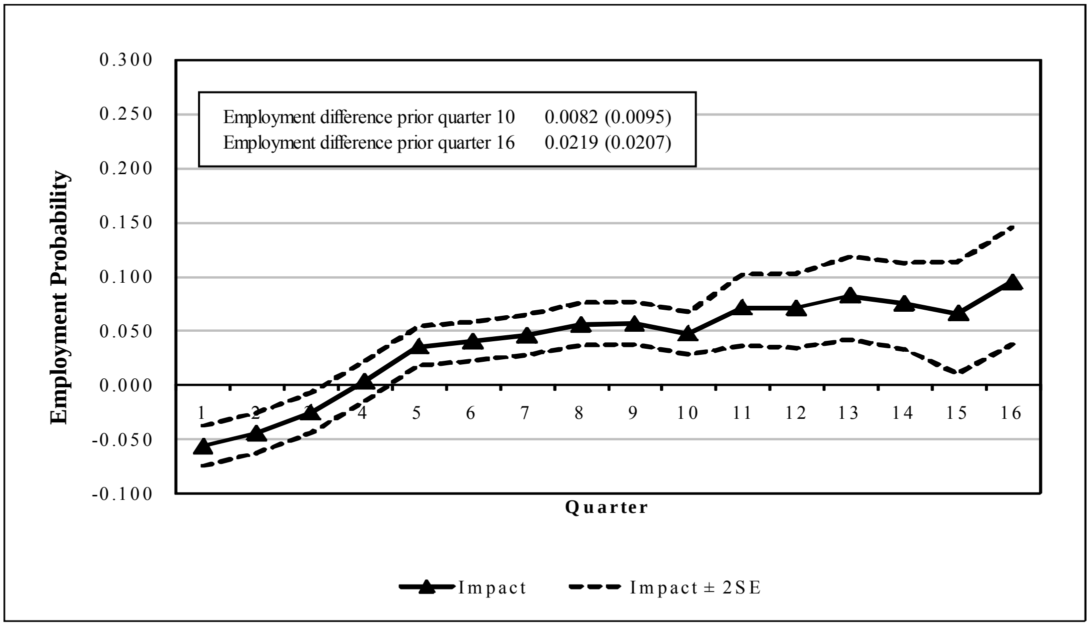
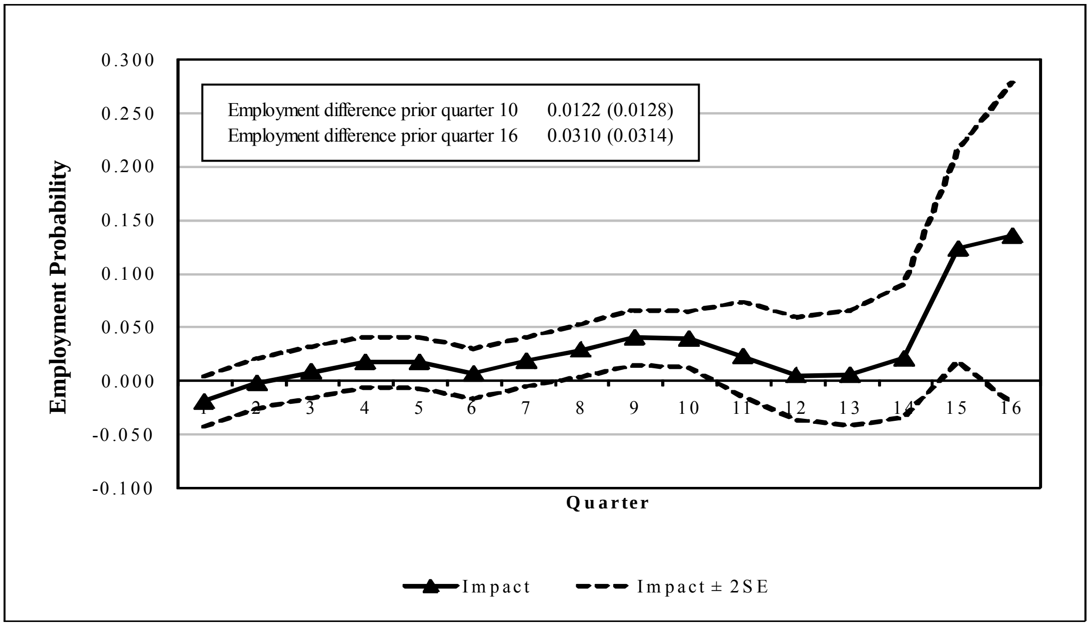


Figure V.18
Adult Program Treatment Effect on Quarterly Employment for Males, WIA Training versus Core/Intensive



For both males and females, there is some indication that selection into training on the basis of stable differences may affect results. For females, those receiving training have earnings that are $100 higher in the quarter four years prior to program entry. For males, the difference is nearly $200. In the employment comparison, the difference is approximately 1 percentage point for females and 3 percentage points for males. If a simple difference-in-difference model is estimated based on this measure, estimated impacts are reduced by these amounts.

4. WIA Adult Program Impacts: Summary

Taken at face value, the results reported above imply large and immediate impacts on earnings and employment for individuals who participate in the WIA Adult program. The short-term effects are greatest for individuals who do not receive training services, although the benefits that accrue to them tend to degrade over time. Those who obtained training services have lower initial returns, but they catch up to others within ten quarters, ultimately registering larger total gains.

For each of these summary statements, a selection story can be constructed to explain away estimated effects. A reader with strong prior beliefs that the program has no effect will not find evidence that clearly rejects this view. But objective observers will likely have difficulty believing that selection alone can explain the observed patterns. In particular, growth in earnings for those receiving training would appear to reflect growth that has been widely observed in related programs.

Results for WIA Adult participants who do not receive training are highly uncertain, although it is reasonable to infer that program impacts are likely to be no more than $100 or $200 per quarter over the four years following program entry. At the same time, WIA training impacts could be substantially greater. By quarter 10, credible impact estimates suggest average benefits of over $400 per quarter.

We estimated impacts separately for various subgroups (and males and females within them), particularly those that make up a larger share of WIA participants than their share in the overall population or who face special challenges or barriers to working in the labor market. The include nonwhites, Hispanics, individuals under 26 years old, those 50 years old or older, and male veterans. For the most part, estimated effects for these subgroups were similar to the estimated effects for all WIA participants. In fact, there is essentially no evidence that there are any substantial differences for any of these subgroups. On the other hand, sampling error for many of these groups is quite large, so differences could well exist.

VI. Dislocated Worker Program: Impact Estimates

The WIA Dislocated Worker program serves individuals who have lost their jobs, in contrast to the Adult program, which seeks to serve disadvantaged workers. Despite this important difference in the targeted group, the formal structures of the Adult and Dislocated Worker programs are similar. Individuals receive Core services as a matter of course, and may be recommended for Intensive services, or, ultimately, Training services. Interestingly, the proportion of individuals who receive training for a given state in the Dislocated Worker program is generally very similar to the proportion in the Adult program, so that the ranking of states by training is essentially the same.

Table II.1 shows that Dislocated Workers are more likely to be male, white, and are, on average eight years older than Adult program participants. Their prior earnings are more than twice those of Adult participants. Overall, they look very much like the comparison group, with earnings in the prior two years within 10 percent of earnings for the aggregated comparison group. However, they are more advantaged than the U.S. Employment Service (ES) comparison group, with average prior earnings that are twice a great (Tables II.1 and II.4).

1. Overall Program Impacts

Estimates of state-specific effects for participants in the Dislocated Worker program are provided in Table VI.1. Here it is clear that impact estimates in the initial quarters are much smaller than comparable estimates for the Adult program. Five states display impact estimates for at least one gender that are negative and statistically significant, implying that those who participate in the program experience lower earnings during the first five quarters after program entry as a result of their program participation. Such results would occur if program activities supplanted employment during initial participation, as might be expected. In only three states is the estimate for these quarters positive and statistically significant for at least one gender.

For quarters 11-16, estimates of impacts in most states are positive and statistically significant for at least one gender. Only two states have values that are negative and statistically significant for either men or women. Despite these apparently positive impact estimates, there are some indicators that these estimates should not be taken at face value. For the ten states where a specification test is possible, we generally find a positive impact on prior earnings—implying that individual participants in the Dislocated Worker program may be advantaged relative to the comparison group in ways not captured by control variables.[28] However, in examining individual states, sampling error is too large to allow this possibility to be investigated. It is therefore necessary to turn to estimates based on aggregating the state results.

---

[28] In five states earnings 16 quarters prior to entry are available, and in five states earnings 10 quarters prior to program entry are available. Among the 20 estimates (10 states by gender), 17 are positive and three are negative.

Table VI.1
Dislocated Worker Program Treatment Effect on Quarterly Earnings by State: WIA versus Comparison Program

	Females		Males	
State	Quarters 1-5	Quarters 11-16	Quarters 1-5	Quarters 11-16
1	-348	143	-561*	130
2	-1813*	-345*	-1666*	-283*
3	142*	368*	211*	254*
4	150	693*	90	624
5	577*	914*	481	941*
6	59	670*	21	568*
7	-619*	286*	-832*	327
8	-1251*	38	-1657*	-517*
9	-191	780*	72	897*
10	-541*	215	783	776
11	888*	1292*	674*	1270*
12	-53	992*	-115	899*

*Statistically significant at the 0.05 level.
Note: Average effects for specified quarters. Where estimates are not available for a given state, the average is calculated on available quarters. In the case of four states, estimates are not available for quarters 11-16, and the reported estimate applies to quarter 10.

Figures VI.1 and VI.2 graph estimated program impacts for participants in all 12 states in the Dislocated Worker program. Participant earnings in the quarter following entry are $200-$300 below the comparison group, but relative earnings increase continuously over the 16 quarters of follow-up analysis. In the fifth or sixth quarter after program entry, participant earnings are equal to those of the comparison group. Ultimately earnings grow to exceed those of comparison group workers by up to $400 per quarter. Despite the similarity in basic pattern, male earnings peak at around 10 quarters, whereas female earnings appear to grow until the end of the four-year window. Long-term female returns appear to be slightly higher as a result.

Figure VI.1
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Females, WIA versus Comparison Group

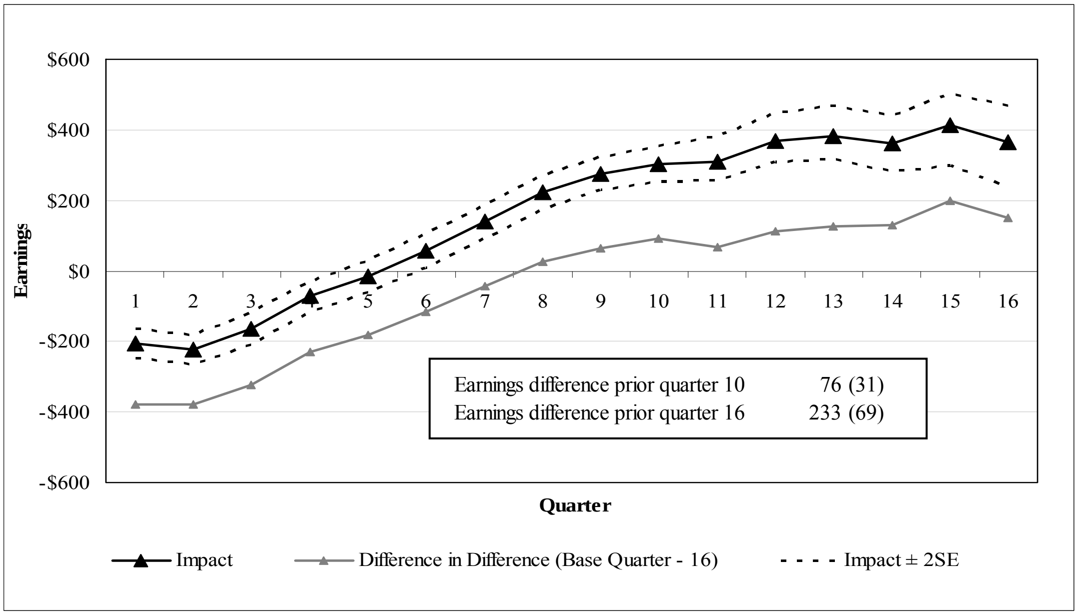
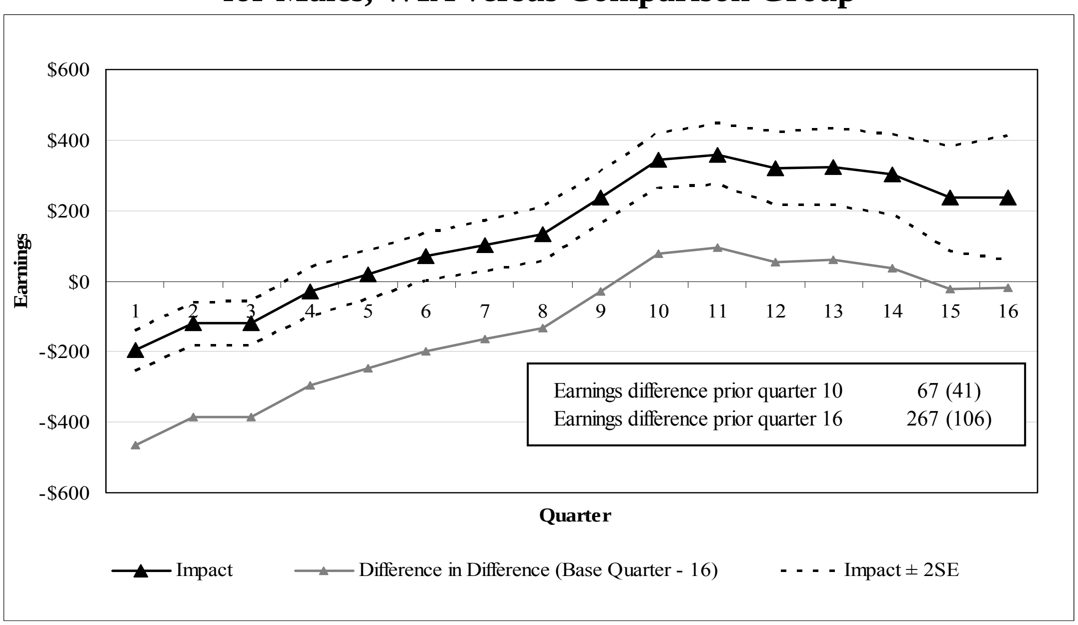


Figure VI.2
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA versus Comparison Group



Figures VI.3 and VI.4 show that for women, initial employment is approximately 2 percentage points below the comparison group, and catches up within about three quarters, and ultimately, employment is nearly 8 percentage points above the comparison group. In contrast, for males there is no initial employment difference, although the growth over time is smaller, with the positive increment after three years peaking at about 6 percentage points.

In the three states for which ES participants are the comparison group, the pattern is almost exactly the same. However, the seven states offering high levels of training display a rather more extreme pattern (Figures VI.5 and VI.6). In the first three to four quarters following program entry, quarterly earnings are more than $800 below the comparison group. Relative earnings do increase, but they only equal the comparison group earnings after eight or nine quarters. For both males and females, earnings exceed the comparison group earnings by about $200, three years (12 quarters) after program entry. Interestingly, participants in these seven states experience a substantially reduced employment rate as well: Figures VI.7 and VI.8 show that Dislocated Workers have employment rates that are 8 to 10 percent lower than others in the first two quarters after program entry, and employment rates exceed the comparison only six to seven quarters later. After 10 quarters, however, employment rates for both genders are about 5 percentage points above the comparison group.

Figure VI.3
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA versus Comparison Group

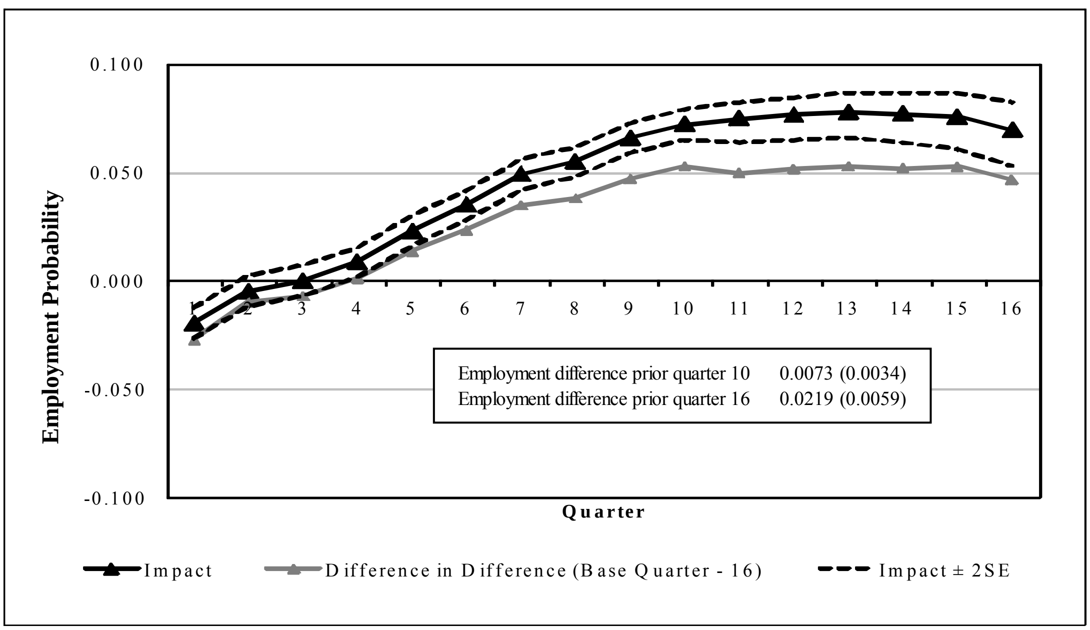


Figure VI.4
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA versus Comparison Group

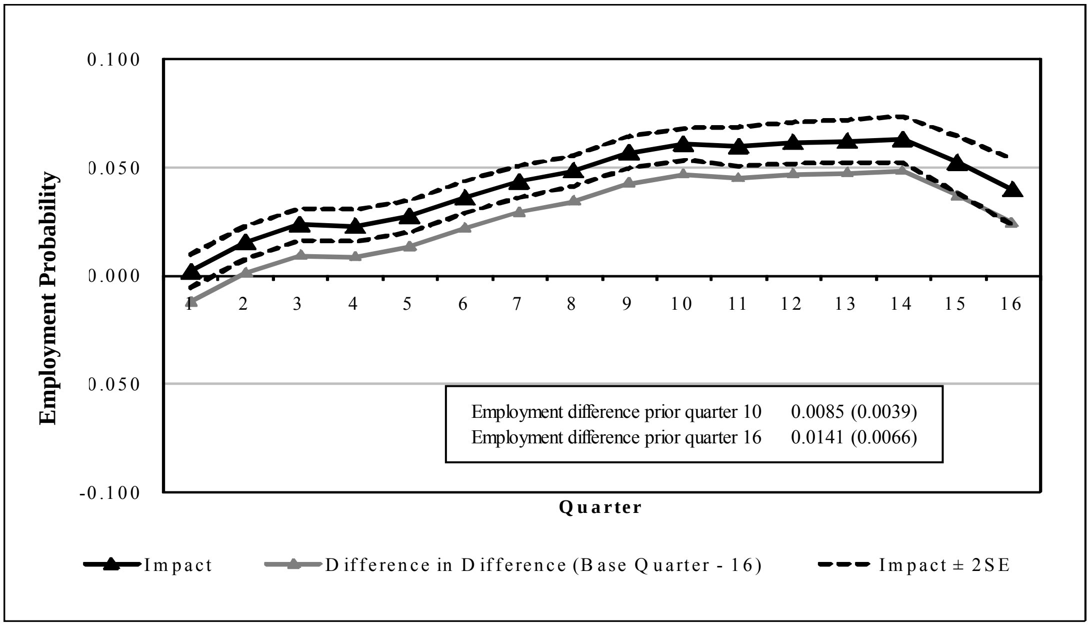


Figure VI.5
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Females, WIA versus Comparison Group in 7 High-Training States

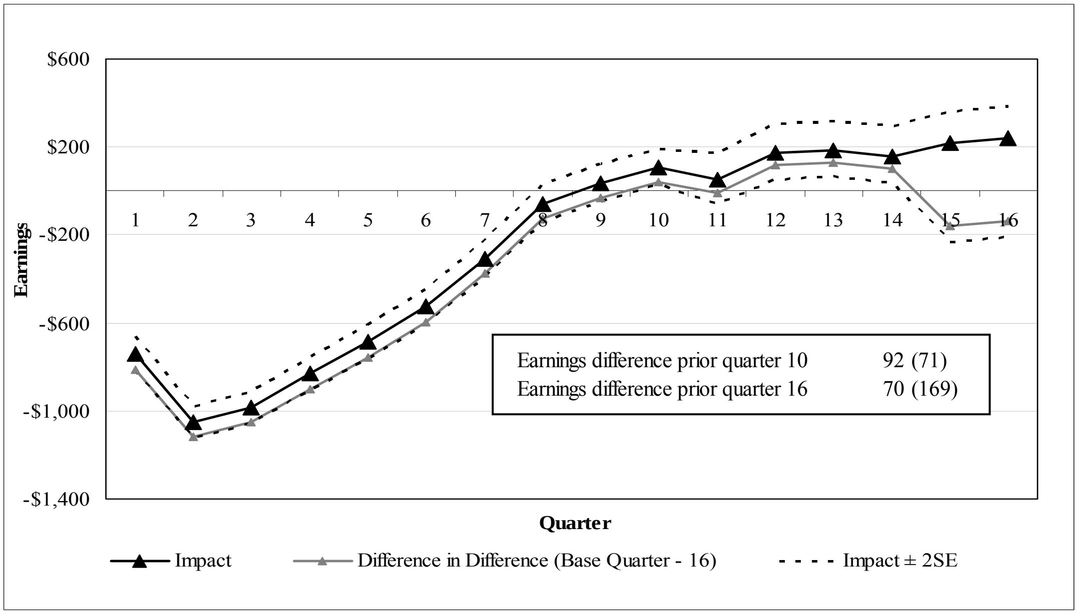


Figure VI.6

Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA versus Comparison Group in 7 High-Training States

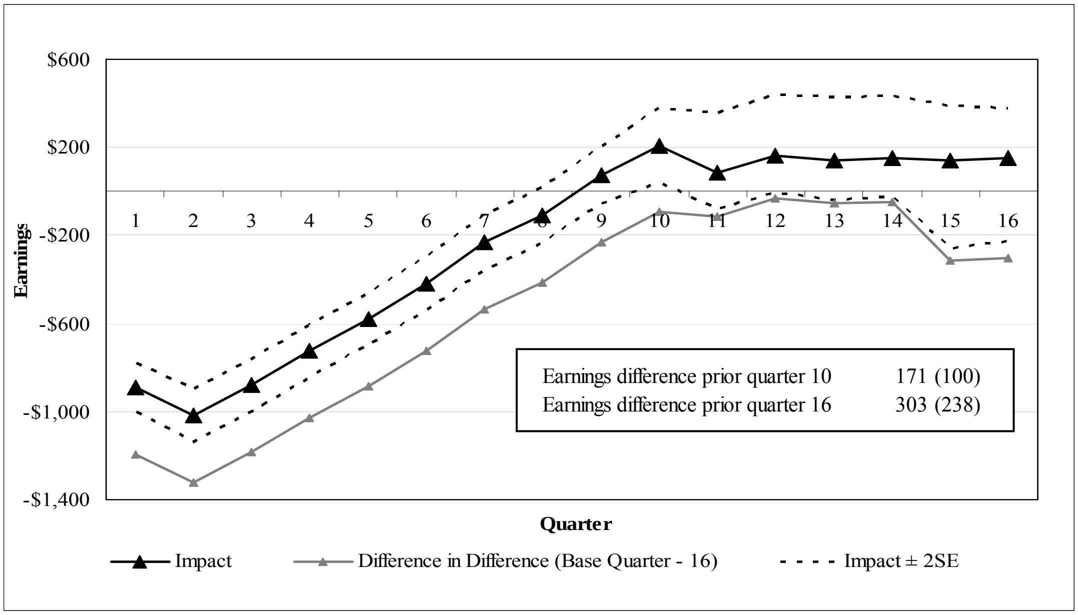


Figure VI.7

Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA versus Comparison Group in 7 High-Training States

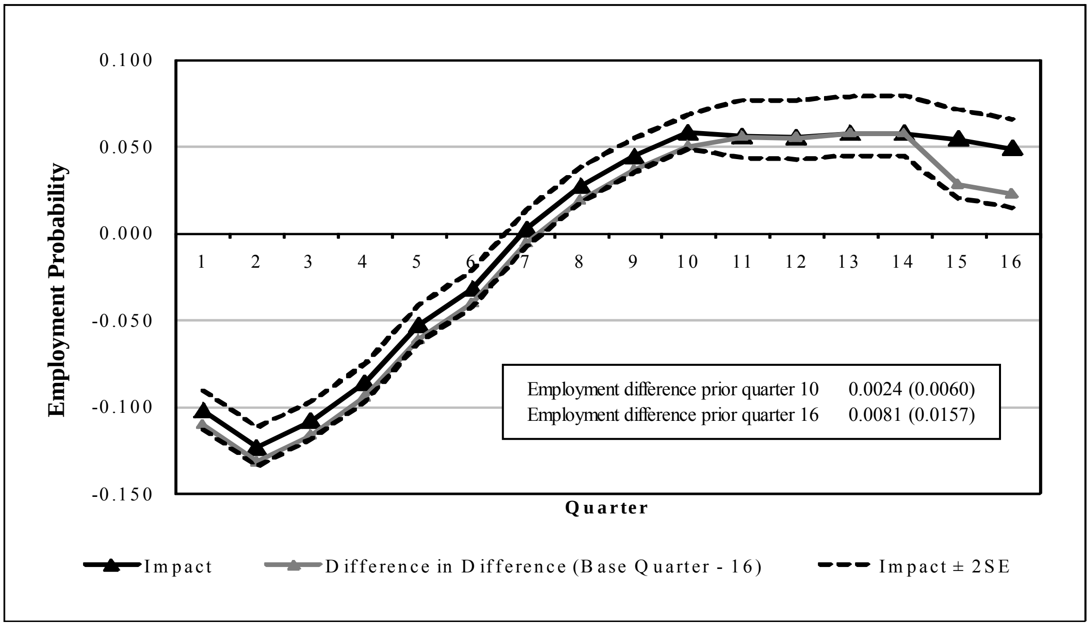
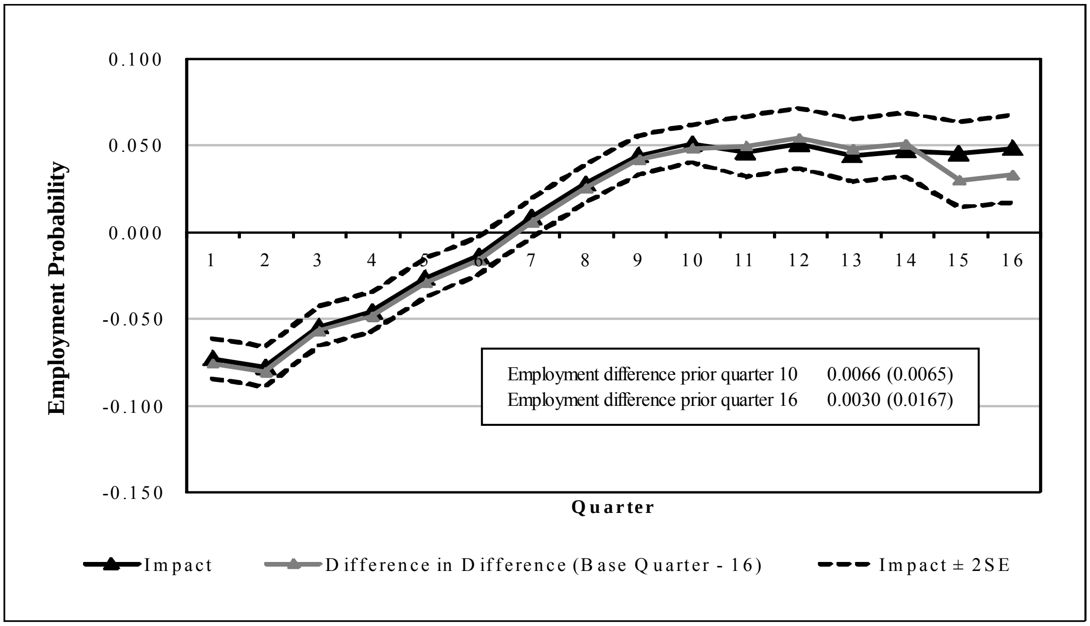


Figure VI.8
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA versus Comparison Group in 7 High-Training States



It is important to recognize that Dislocated Worker program participants are relatively high wage individuals who are faced with permanent job loss. Their initial negative impact estimates imply that their earnings are below unemployed workers with similar prior incomes and work histories. This is what would be expected if involvement in training activities precludes or reduces employment, inducing lock-in effects. Earnings growth observed over the three following years is consistent with the attainment of skills with training.

This interpretation is based on the assumption that Dislocated Workers are similar in unmeasured ways to the comparison group. Our specification test, based on predicting prior earnings, suggests this is not the case. There are substantial differences between the participant and comparison group 16 quarters earlier, with participant earnings more than $200 higher (Figures VI.1 and VI.2). Standard errors (in parentheses) imply that these estimates are statistically significant. Prior employment is also several percentage points higher for program participants. That matched participants have higher prior earnings suggests the possibility that their earnings in later periods may not reflect program impact but rather unmeasured factors that become apparent in the three years after program entry.

Although these findings contribute to greater uncertainty as to the true impact, it is possible to obtain some indication of the possible extent of the bias. The difference-in-difference estimate subtracts this prior increment from all impact measures. In the case here, the impact estimate for prior quarter 16 is subtracted from the simple estimate reported in the table to produce the difference-in-difference estimate. As discussed in Section II, this estimator provides a valid estimate of program impact if the program selects on the basis of stable characteristics (with the same impact on prior and subsequent earnings) that are not captured by measures of earnings in the prior eight quarters. (In the analyses of the Adult program, in Section V, the prior increment measures were of modest size, so the difference-in-difference estimates were not included in the figures.) The difference-in-difference estimates imply that participants' earnings catch up to those of nonparticipants with a longer delay and that the ultimate impact on earnings is more modest. For women, earnings exceed those of nonparticipants only after eight quarters, and the positive increment is never over $200. For men, the crossover point is after nine quarters, and the increment is generally less than $100.

Taken together, these results imply those who participate in the program have earnings below nonparticipants for an extended period. Among women, participants are also less likely to be employed. Participants' earnings and employment levels do overtake those of nonparticipants in two to three year's time. Prior earnings differences between participants and the comparison group suggest that there may be selection on unmeasured stable factors, so that ultimate impact estimates may overstate program effects. On the other hand, it is worth noting that initial earnings differentials may well reflect selection on short-run differences. If individuals whose immediate employment opportunities are particularly poor are likely to enter the Dislocated Worker program, their lower initial earnings could at least partly reflect selection.

2. Impact Estimates for WIA Core and Intensive Services

Individuals who enter the WIA Dislocated Worker program but receive no training may be selected in several ways. In addition to program selection, they may be individuals who enter the program but then obtain employment and therefore decide training is of little value. They may also include those who were unable or unwilling to follow through on training requirements. Figure VI.9 and VI.10 provides estimates for this group. Since these are individuals who spent appreciably less time in the program, it is expected that their lock-in period would be shorter. In fact, earnings are not below the comparison group at any point, and earnings do increase over time. Figures VI.11 and VI.12 show that their levels of employment are initially one or two percentage points above those for the comparison group, and that these numbers increase over time.[29]

As above, the prior earnings (10 and 16 quarters prior to entry) of this group of WIA Dislocated Worker participants were substantially higher than the comparison group, with the difference in the range of $200-$300. These differences in prior earnings imply that the specification is not valid. If participants have substantially greater earnings four years earlier, their subsequent earnings may well be higher even if training has no effect. Difference-in-difference estimates provide an alternative measure of program impact. As in the case of the earlier figures presenting Dislocated Worker results, in each of these figures the difference-in-difference estimates are plotted. These estimates suggest a much smaller program impact than the initial estimates.

[29] Separate estimates for the states that train a high proportion of workers do not reveal any important differences. However, since such states have relatively small numbers of individuals who are not trained, the small samples limit possible inferences.

Figure VI.9
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Females, WIA Core/Intensive versus Comparison Group

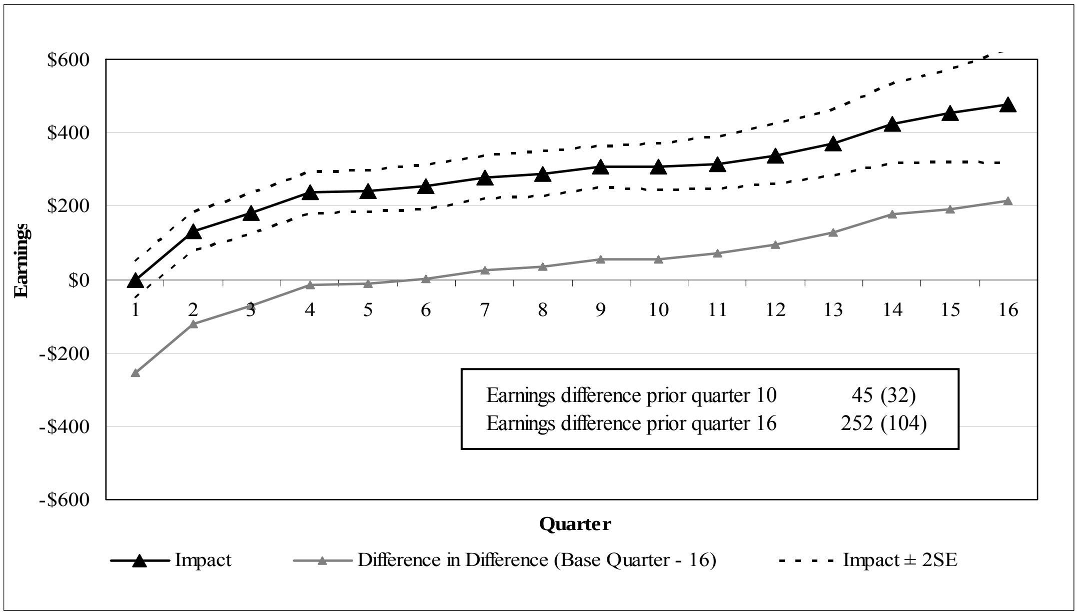


Figure VI.10
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA Core/Intensive versus Comparison Group

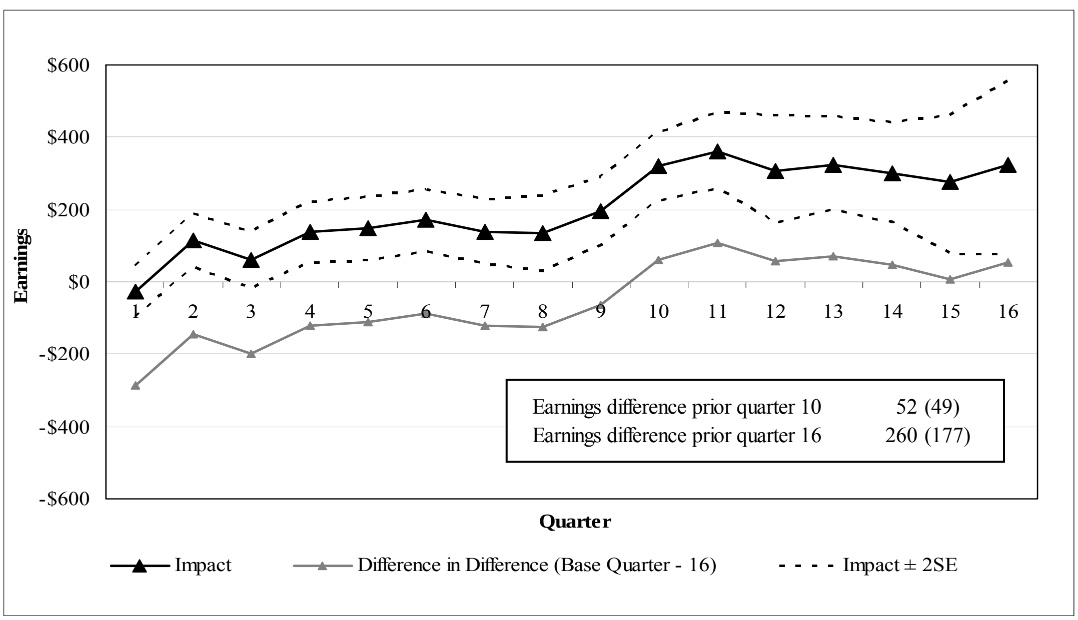


Figure VI.11
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA Core/Intensive versus Comparison Group

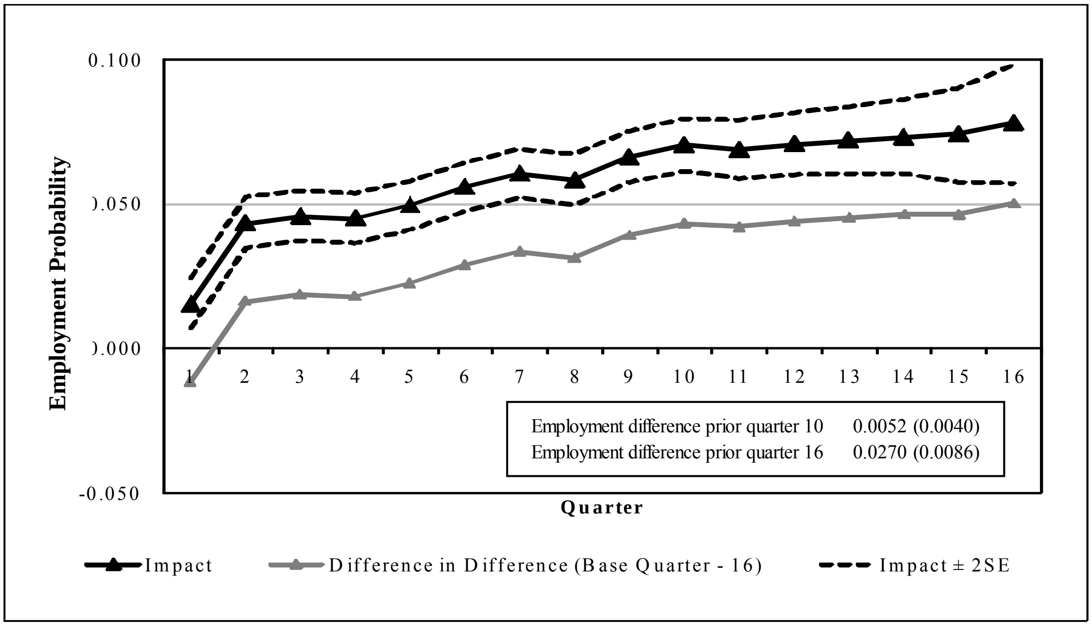
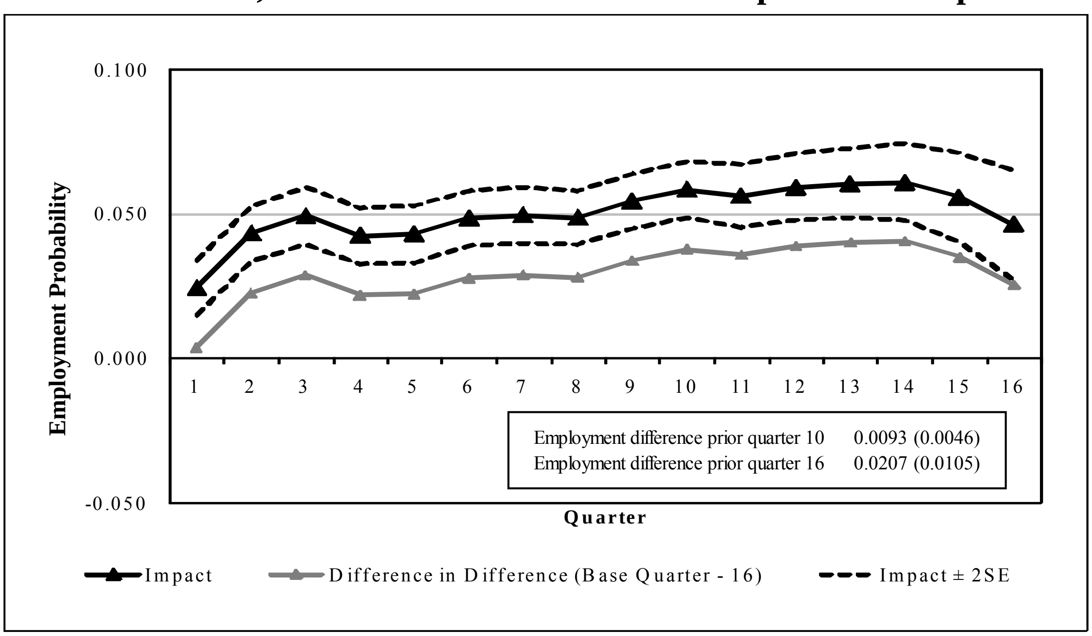


Figure VI.12
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA Core/Intensive versus Comparison Group



3. Impacts of Training

The incremental impact of training is based on a comparison of WIA Dislocated Worker participants who obtain training with those who do not. Figures VI.13 and VI.14 show that initial earnings for those obtaining training are below those of other program participants for eight quarters for women and for more than ten quarters for men. Difference are $1,100 for females in quarters 2 through 4, and $800 for males. A very similar pattern exists for employment (Figures VI.15 and VI.16).

Although the initial negative impact estimate is large relative to sampling error, the confidence interval is large relative to estimated impacts after quarter 10. Estimates for both males and females could easily differ from the point estimate by $200 due to sampling error. Also of possible concern is the difference in earnings prior to entry into the program. For females, the individuals who select into training have lower earnings relative to others in the sixteenth quarter prior to participation. This difference is not, however, statistically significant, so evidence of selection is inconclusive. Estimates for states offering high proportions of training are not substantively different.

Taken at face value, point estimates suggest that those who enter training experience large earnings losses relative to those who do not in their first two years after program entry. Although consistent with a large training lock-in effect, these effects could be at least partly due to selection on short-term employment prospects, with those who land jobs leaving the program without obtaining training. Estimates of effects on earnings and employment three to four years after program entry—more than 18 months after program exit for most participants—show little evidence that training produces substantial benefits. These negative conclusions must be tempered, however, by a recognition that sampling error alone could obscure substantial impacts.

Figure VI.13
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Females, WIA Training versus Core/Intensive

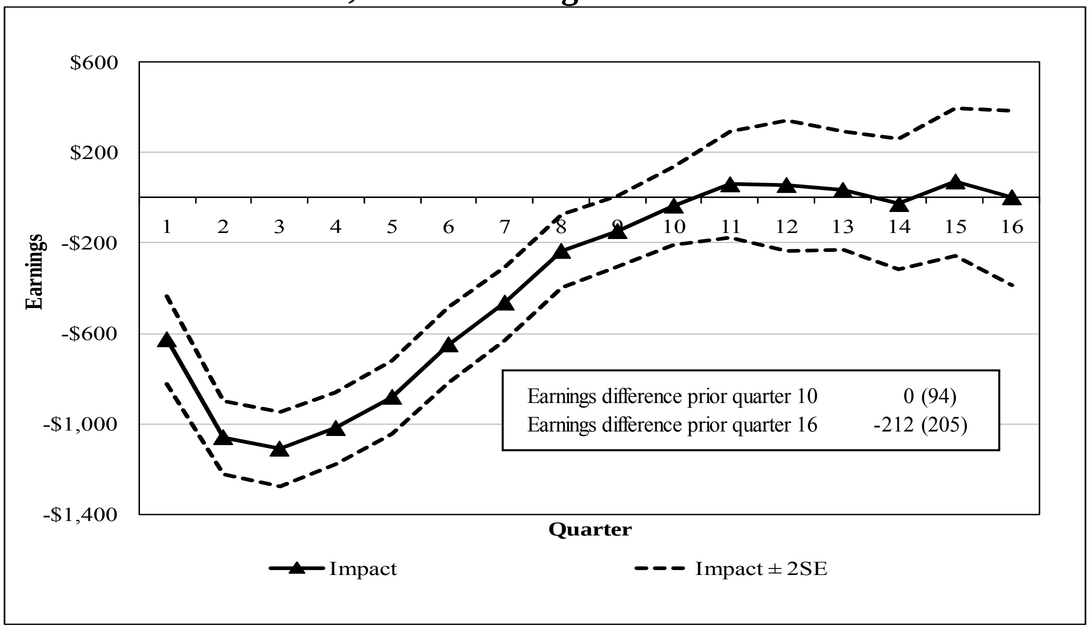


Figure VI.14
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA Training versus Core/Intensive

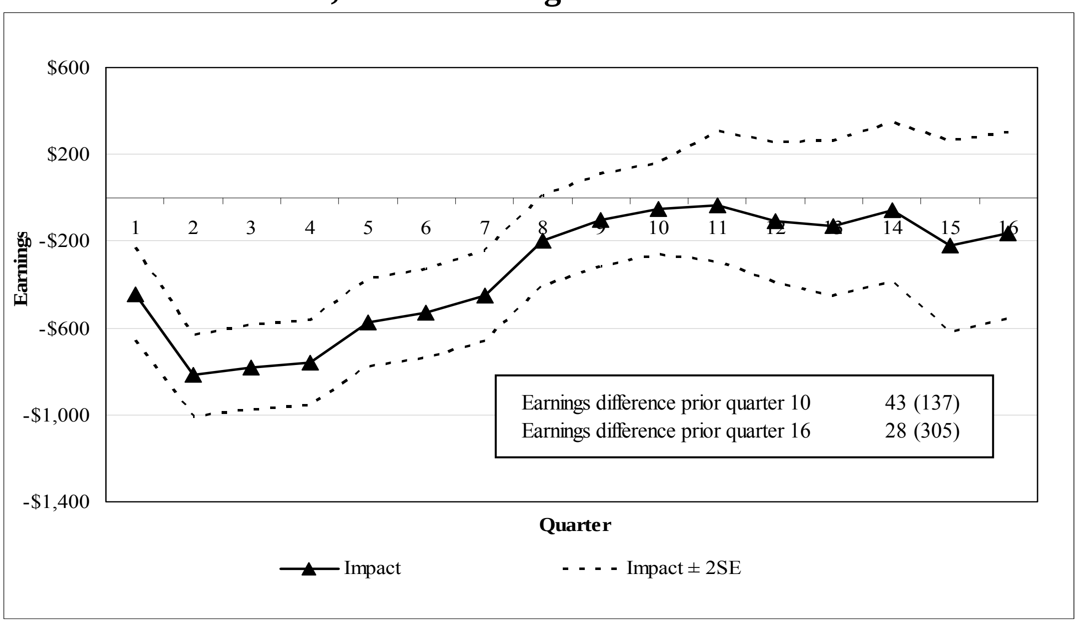


Figure VI.15
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA Training versus Core/Intensive

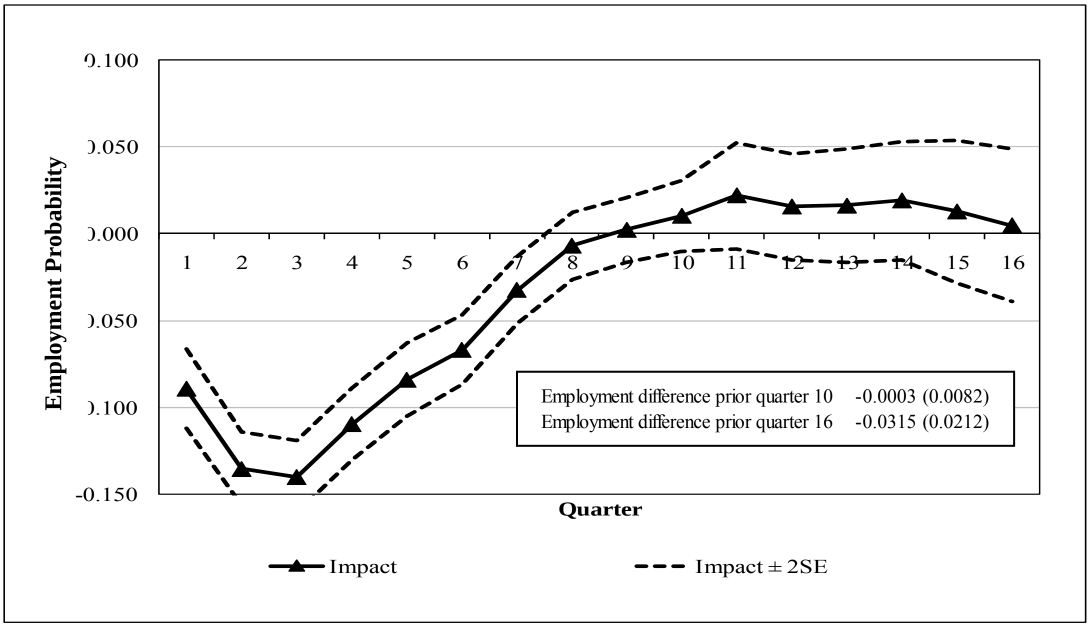
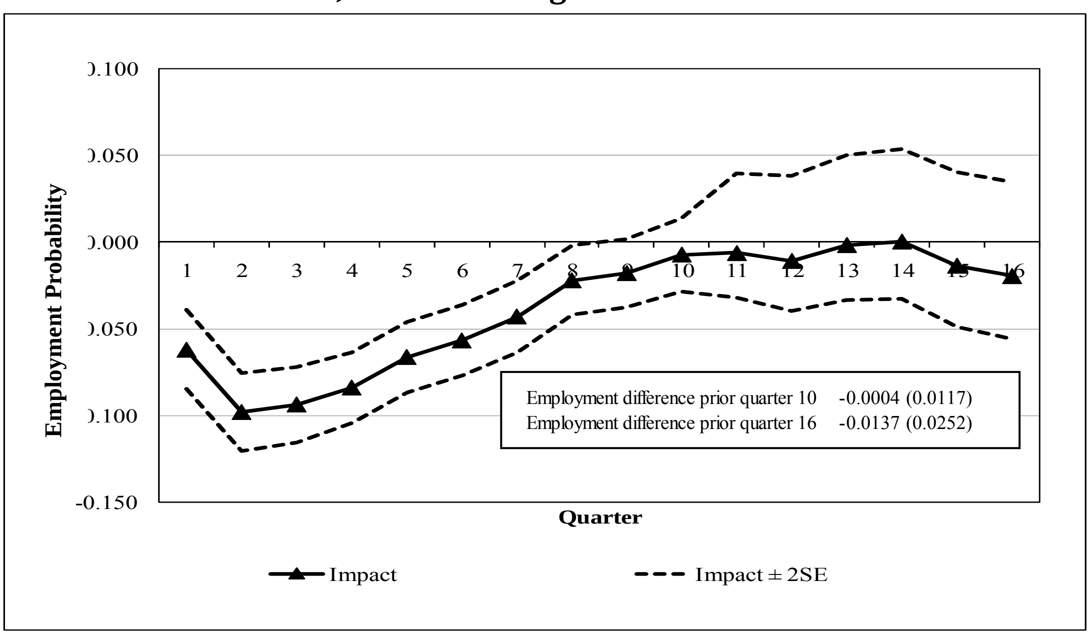


Figure VI.16
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA Training versus Core/Intensive



4. WIA Dislocated Worker Program Impacts: Summary

Dislocated Workers are likely to face serious difficulties in obtaining reemployment, and the kinds of services WIA offers may require time to produce impacts. The pattern of results is consistent with these expectations. However, the extent of any benefits that accrue from participation is particularly hard to judge. Some specification tests suggest that results may be biased toward finding positive program impacts. Difference-in-difference estimates are therefore appreciably smaller than the primary reported estimates. These estimates imply that program participants' earnings do not reach the level of earnings of comparable nonparticipants until more than two years after participation. Perhaps more important, the growth in earnings, relative to nonparticipants, slows at that point. As a result, these estimates imply that the gains from participation are, at best, very modest, even three to four years after entry. Overall, it appears possible that ultimate gains from participation are small or nonexistent. Insofar as there are impacts, females are more likely to benefit than are males.

Where employment is taken as the outcome of interest, estimates of program impact are more supportive of the program. Although the specification tests again suggest that there are unmeasured differences between the treated and matched comparison group, the difference-in-difference estimates of program impact are positive.

Overall, given the results of these specification tests, it is necessary to treat our estimates with caution. Whereas we can conclude with a high degree of certainty that the lower earnings associated with program participation disappear within two years, it is less clear that there are net benefits associated with participation. Although positive program impacts—especially on employment—are consistent with these findings, substantial uncertainty remains.

We estimated impacts separately for nonwhites, Hispanics, individuals under 26 years old, those 50 years old or older, and male veterans, finding no evidence that there are any substantial differences in program impact estimates for any of these subgroups. As in the case of subgroup analysis for the Adult program, sampling error is substantial, and there may be differences that are not statistically discernable.

VII. Summary and Implications

The estimates of WIA program impact presented here are based on administrative data from 12 states, covering approximately 160,000 WIA participant and nearly 3 million comparison group members. Focusing on those entering WIA in the period July 2003-June 2005, estimates are provided for the effect of Core/Intensive and Training services in these programs.

Given variation in the way that states have implemented WIA, it was anticipated that estimates of the WIA program impact would differ across states. The findings of the analysis confirmed notable differences across states in estimated program impacts and patterns of impacts. Such differences likely reflect local economic environment, program structure, labor force composition, and possibly complex interactions of these factors. Some of the observed differences are undoubtedly due to various statistical artifacts, reflecting selection effects occurring both in the WIA programs and the comparison programs; hence, some of the variation may be spurious. Yet many differences are undoubtedly real and identify differences in program efficacy.

Importantly, there are also similarities in the patterns of estimated impacts across states. It would appear implausible that these commonalities in the patterns of results could be explained solely by selection causing differences in unmeasured factors for WIA participants and comparison individuals. Women appear to obtain greater benefits for participation in both the Adult and Dislocated Worker programs, with the quarterly earnings increment exceeding that for males. The value of training appears to be greater as well, especially over the long run. These results are consistent with findings of prior studies, including random assignment experiments (Orr et al., 1996), and generally suggest that it is highly likely that WIA participants gain from their participation.

Adult program participants who receive only Core/Intensive services have initial earnings comparable to those obtained for all Adult participants, but this benefit declines in a relatively short period to a level of approximately \$200-\$300 per quarter. This pattern is consistent with the view that individuals who do not engage in training receive effective short-term counseling that enables them to gain an immediate advantage in the labor market but that this advantage fairly quickly diminishes. One cannot rule out the possibility, however, that this pattern results from selection into the program, in which counselors admit applicants who have better job prospects, or selection into training, whereby those with immediate job opportunities forgo training. In this case, earnings impact estimates would not reflect program effects but rather selection by counselors or participants.

Adult program participants who obtain training services have lower initial earnings, but they catch up to others within ten quarters, ultimately registering total gains of over \$800 for females and \$500-\$600 for males.

The Adult program is more likely to produce tangible benefits for participants than the Dislocated Worker program. Dislocated Workers experience several quarters following entry into WIA with earnings that are below those for the matched comparison group. Although their earnings do ultimately equal or even overtake the comparison group, the benefit of obtaining training for this group is quite small.

In conclusion, overall WIA program net impacts were estimated to be positive in almost all states, although important variation across programs and specific services clearly exists. While the possibility that these estimates may be partly spurious cannot be ruled out, none of the selection explanations considered would fully explain away these patterns of positive estimated program effects.

There are important policy implications of these results that go beyond a simple judgment of whether the program is effective. Program administrators typically look at the cross-sectional or "point-in-time" information that is available to them from performance management systems on a regular basis. They do not have at hand the data analysis tools to examine individual employment and earnings histories and trajectories for more than eight years (33 quarters that include up to 16 quarters of follow-up data) for both program participants and a comparison group, as in this study. The results of this evaluation show that program impacts typically "mature" over time, sometimes increasing in magnitude and sometimes diminishing. Insofar as this work underscores the fact that long-term impacts are of significance and that outcomes of interest may not be apparent for years, it will help to refocus training activities.

This study makes no attempt to formally generalize these results to the full population of WIA participants in the Adult and Dislocated Worker programs. Such generalization would require developing a model to explain differences across states, in essence performing a meta-evaluation that would treat each state's impact estimates separately. Although the WIA participants in our analysis are not a representative sample, there are several reasons why these results are likely to apply to the WIA program more generally. The sample includes about one in five WIA participants during the period of study and the data contain a diverse sample of states. It is reasonable to conclude that the processes by which WIA operates in the 12 states included in the study are similar to those in other states. The results of this study therefore provide information on the long-run impact of the WIA program of potential value in allocating WIA program resources both at the local and national level.

Acknowledgements

The analyses presented here are based on work undertaken for the U.S. Department of Labor and presented in "Workforce Investment Act Non-Experimental Net Impact Evaluation" (IMPAQ International, Final Report, December 2008). The authors wish to acknowledge the central role in this project played by the staff at IMPAQ. IMPAQ staff involved in the negotiation of agreements with states, data acquisition, and initial cleaning of data included Nicholas Bill, Goska Grodsky, Shirisha Busan, and Ted Shen. Eileen Poe-Yamagata served as project manager, overseeing all aspects of the project. As project director, Jacob Benus provided direction and advice at the critical intersection of research strategy and logistics. Numerous individuals outside IMPAQ supported the project in a variety of ways. Thanks are due to the many state agency staff who worked to provide data and responded to numerous requests for further information and clarification, to David Stevens who facilitated provision of data for Maryland, and to Suzanne Troske, who supported data processing in Kentucky. Finally, the authors wish to acknowledge the role played by Adam Tracy, who served as project officer at the Department of Labor, and Jonathan Simonetta, who had direct responsibility for defining many elements of the project underlying this analysis. Notwithstanding the support provided by these individuals and the U.S. Department of Labor, the analysis and interpretation presented in the current paper are the sole responsibility of the authors.

References

Abadie, Alberto and Guido W. Imbens, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica* 74:1 (January 2006a), 235-267.

Abadie, Alberto, and Guido W. Imbens, "On the Failure of the Bootstrap for Matching Estimators," Unpublished paper, John F. Kennedy School of Government, Harvard University (2006b).

Appelbaum, Eileen, Bernhardt, Annette and Richard J. Murnane. 2003. *Low-Wage America: How Employers Are Reshaping Opportunity in the Workplace*. New York: Russell Sage Foundation.

Ashenfelter, Orley C., "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics* 60 (February 1978), 47-57.

Bloom, Howard S., Charles Michaelopoulos and Carolyn J. Hill, "Using Experiments to Assess Nonexperimental Comparison-Groups Methods for Measuring Program Effects," in Howard S. Bloom (Ed.), *Learning More from Social Experiments: Evolving Analytic Approaches*, New York: Russell Sage (2005), 173-235.

Bradbury, Katharine and Jane Katz. 2002. "Are Lifetime Incomes Growing More Unequal? Looking at New Evidence on Family Income Mobility." *Regional Review* (Federal Reserve Bank of Boston) 12(4): 3-5.

Cook Thomas D., William R. Shadish and Vivian C. Wong, "Three Conditions Under Which Experiments and Observational Studies Often Produce Comparable Causal Estimates: New Finidngs from Within-Study Comparisons," *Journal of Policy Analysis and Management* 27 (Autumn 2008), 724-750.

Dehejia, Rajeev H. and Sadek Wahba, "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association* 94 (December 1999), 1053-1062.

Dehejia, Rajeev H., and Sadek Wahba, "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics* 84:1 (February 2002), 151-161.

Dyke, Andrew, Carolyn Heinrich, Peter R. Mueser, Kenneth R. Troske, and Kyung-Seong Jeon, "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes," *Journal of Labor Economics* (vol. 24, no. 3, July 2006), pp. 567-608.

Frank, Abby and Elisa Minoff. 2005. "Declining Share of Adults Receiving Training Under WIA Are Low-Income or Disadvantaged." Washington, DC: Center for Law and Social Policy.

Frölich, Markus, “Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators,” *The Review of Economics and Statistics* 86:1 (February 2004), 77-90.

Galdo, Jose, Jeffry Smith, and Dan Black, “Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data,” Unpublished, May 2008.

Glazerman, Steven, Dan M. Levy and David Myers, “Nonexperimental versus Experimental Estimates of Earnings Impacts,” *Annals of the American Academy of Political and Social Science* 589 (September 2003), 63-93.

Heckman, James J., Robert J. LaLonde, and Jeffery A. Smith, “The Economics and Econometrics of Active Labor Market Programs,” in Orley Ashenfelter and David Card (Eds.) *Handbook of Labor Economics*, Vol. 3 (Amsterdam: North Holland, 1999).

Heckman, James J., and Jeffery A. Smith, “The Pre-programme Earnings Dip and the Determinants of Participation in a Social Programme: Implication for Simple Programme Evaluation Strategies,” *Economic Journal* 109 (July 1999), 313-348.

Heinrich, Carolyn J., “Demand and Supply-Side Determinants of Conditional Cash Transfer Program Effectiveness,” *World Development* 35 (January 2007), 121-143.

Holzer, Harry. 2004. “Encouraging Job Advancement among Low-Wage Workers: A New Approach.” Policy Brief #30. Washington, DC: The Brookings Institution.

Hotz, V. Joseph, Guido W. Imbens and Jacob A. Klerman, “Evaluating the Differential Effects of Alternative Welfare-to-Work Training Compoents: A Reanalysis of the California GAIN Program,” *Journal of Labor Economics* 24 (July 2006), 521-566.

Imbens, Guido W., “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” *Review of Economics and Statistics* 86:1 (February 2004), 4-29.

Imbens, Guido W., “Estimating Variances for Estimators of Average Treatment Effects,” Unpublished, Harvard University (September 2008).

Imbens, Guido W., and Jeffrey M. Wooldridge, “Recent Developments in the Econometrics of Program Evaluation,” Institute for Research on Poverty Discussion Paper no. 1340-08, University of Wisconsin, 2008.

LaLonde, Robert J, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *American Economic Review* 76:4 (1986), 604-620.

Lechner, Michel. “Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption.” Pp. 43–58 in *Econometric Evaluation of Active Labour Market Policies*, ed. M. Lechner and F. Pfeiffer (Heidelberg: Physica, 2001).

Mueser, Peter R., Kenneth R Troske, and Alexey Gorislavsky, "Using State Administrative Data to Measure Program Performance," *Review of Economic and Statistics* (vol. 89, no. 4, November 2007), pp. 761-783.

Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin and George Cave, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study* (Washington D.C.: The Urban Institute Press, 1996).

Osterman, Paul. 2007. "Employment and Training Policies: New Directions." In *Reshaping the American Workforce in a Changing Economy*, Harry J. Holzer and Demetra Smith Nightingale (eds.), Washington, DC: Urban Institute Press, pp. 119-154.

Peikes, Deborah N., Lorenzo Moreno and Sean Michael Orzol, "Propensity Score Matching: A Note of Caution for Evaluators of Social Programs," *The American Statistician* 62 (August 2008), 222-231.

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

Rosenbaum, Paul R., *Observational Studies* (New York: Springer-Verlag, 2002).

Rubin, Donald B., *Matched Sampling for Causal Effects* (Cambridge: Cambridge University Press, 2006).

Smith, Jeffrey A. and Petra E. Todd, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125 (March-April, 2005a), 305-53.

Smith, Jeffrey A. and Petra E. Todd, "Rejoinder," *Journal of Econometrics* 125 (March-April, 2005b), 365-75.

Social Policy Research Associates, PY 2005 WIASRD Data Book: Final, Prepared for the U.S. Department of Labor, August 2007.

Zhao, Zhong, "Data Issue of Using Matching Methods to Estimate Treatment Effects: An Illustration with NSW Data Set," Peking University Working Paper No. #203004 (2003).