Workforce Investment Act Non-Experimental Net Impact Evaluation

FINAL REPORT

December 2008

Lead Investigators

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

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

Kenneth R. Troske
University of Kentucky

Project Director

Jacob M. Benus
IMPAQ International, LLC

Acknowledgements

The lead investigators wish to acknowledge the central role in this project played by the staff at IMPAQ International, which served as the primary contractor on this project. 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. Daver Kahvecioglu was instrumental in implementing the statistical methods of analysis. Eileen Poe-Yamagata served as project manager, overseeing all aspects of the project and contributing to the final editing of the manuscript. 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. Special thanks are due to Kyung-Seong Jeon, who undertook independent analyses for this project at the University of Missouri and worked with IMPAQ staff to implement the statistical methods. Finally, the investigators 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.

Workforce Investment Act
Non-Experimental Net Impact Evaluation
FINAL REPORT

Table of Contents

Executive Summary

This study reports results of a nonexperimental net impact evaluation of the Adult and Dislocated Worker programs under the Workforce Investment Act (WIA). The key measure of interest is the difference in average earnings or employment attributable to WIA program participation for those who participate. The analyses focus on three levels of service: *Core services*, including outreach, job search and placement assistance, and labor market information; *Intensive services*, involving comprehensive assessments, development of individual employment plans, counseling, and career planning; and *Training services*, occupational and basic skills training that link customers to job opportunities in their communities.

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 only Core or Intensive services, and the incremental impact of Training services. State-of-the-art nonexperimental methods are used to compare WIA program participants with matched comparison groups of individuals who are observationally equivalent across a range of demographic characteristics, social welfare benefit receipt and labor market experiences but who either did not receive WIA services or did not receive WIA training.

In evaluating the impact of an intervention, such as WIA, it is necessary to identify both the outcome for participants and the outcome that would have occurred had they not participated. In order to estimate the outcome for individuals had they not participated in the program, propensity score matching techniques are used to identify individuals in comparison groups who are similar to the individuals who participated in the WIA program. An extensive set of variables on individual demographic characteristics, labor market experience and geographic location are used to predict the probability that an individual participates in the program. Individuals are then matched using these estimated probabilities. Previous work has shown that this method works well when researchers use data with a set of covariates that accurately capture the likelihood that an individual participates in the program. The rich data available for this study means that this is an ideal methodology to use to conduct this nonexperimental evaluation.

Although differences across states are substantial, important similarities in the patterns of estimated impacts are observed. The results for all participants in the WIA Adult program (regardless of services received) show that participating is associated with a several-hundred-dollar increase in quarterly earnings. The analysis of participants who receive only Core/Intensive services suggests that their benefits may be as great as $100 or $200 per quarter over the period of study, which is substantial compared to the small costs of those services. Adult program participants who obtain Training services have lower initial earnings than those who don’t receive Training services, but they catch up within 10 quarters, ultimately registering large total gains per quarter. The marginal benefits of training may exceed $400 in earnings each quarter.

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 do ultimately overtake the comparison group, although the analyses suggest that the benefits they obtain are smaller than those in the Adult program. The return they experience from training also appears to be appreciably smaller than that obtained by Adult program participants.

Women appear to obtain greater benefits for participation in both the Adult and Dislocated Worker programs, with the quarterly earnings increment exceeding that of males. The value of training appears to be greater for females as well, especially over the long run.

Impacts are also estimated separately for various subgroups (and males and females within them), focusing on those who are overrepresented among WIA participants or who face special challenges or barriers to working in the labor market, including: nonwhite non-Hispanics; Hispanics; those under 26 years of age; those 50 years of age or above; and veterans. The estimated effects for the subgroups are similar to the estimated effects for all WIA participants; there is essentially no evidence that any of the subgroups considered have experiences that differ from the average in important ways. Sampling error for these groups is quite large, so the findings are also consistent with at least moderate variation across groups.

It is important to note that these estimates are averages, and differences across states are substantial. However, estimates of overall WIA program net impacts are positive in almost all states. Although it is not possible to rule out the possibility that some estimates may be influenced by systematic selection that has not been controlled by these methods, the general pattern of the results almost surely reflects actual program impacts on individual participants.

Although the WIA participants in the 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 a large fraction of all 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.

I. Introduction

The Workforce Investment Act (WIA), enacted in August 1998, created a new, comprehensive workforce investment system. WIA is distinguished from its public employment and training program predecessors by the introduction of a One-Stop service delivery system designed to: improve coordination and integration of services; provide universal access to services for adults; promote the use of Individual Training Accounts in training services; increase local flexibility in determining training strategies; and establish additional requirements for performance measurement and accountability.

In 2005, the U.S. Office of Management and Budget, using its Program Assessment Rating Tool (PART), assigned the WIA program an overall rating of “adequate.” Although the detailed PART report indicated that the program exceeded its central goals of helping participants find jobs and keep them (as measured by the entered employment rate, job retention rate, and average earnings increases of WIA participants between 2000 and 2004), the program received relatively low marks for its evaluation efforts. The report suggested that independent evaluations had not been of sufficient scope and rigor to determine WIA's impact on participants' employment and earnings.

The nonexperimental net impact evaluation of WIA described in this report addresses the immediate need for a rigorous evaluation of the WIA program to assess the effectiveness of WIA services in improving the labor market outcomes of WIA participants. The evaluation focuses on the Adult and Dislocated Worker programs under Title I of the Workforce Investment Act. The Adult and Dislocated Worker programs offer three levels of service to WIA participants: *Core services*, including outreach, job search and placement assistance, and labor market information; *Intensive services*, involving comprehensive assessments, development of individual employment plans, and counseling and career planning; and *Training services* (both occupational and basic skills training) that link customers to job opportunities in their communities.[1] Administrative data from 12 states are used to analyze the impacts of Core/Intensive and Training services in these programs, focusing on participants who entered WIA in the period July 2003-June 2005.

The study employs state-of-the-art nonexperimental methods that reflect the considerable advances made in the last two decades to improve the accuracy of impact estimation in the absence of random assignment. Within each state, WIA program participants are compared 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. Participants and comparison group members are compared within state and state-established workforce investment areas to assure

[1] These services are typically made available sequentially, so that receipt of training services follows intensive services, and only after core services are provided.

that they are facing similar local labor markets,[2] and all outcomes are measured comparably 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. A Department of Labor (DOL) project to implement an experimental evaluation of the WIA program has just now begun, and results will be available in approximately seven years. Although these results will be of great interest, the evaluation undertaken in this study, completed within a 15-month time frame and at a much lower cost, can provide important information for policy makers who must make decisions in the immediate term.

It must be recognized, however, that, in the absence of data drawn from a representative sample of the population of WIA participants, the study does not 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. 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. The study covers 159,131 WIA participants in the Adult and Dislocated Work programs, or slightly less than a fifth of the participants during the period under consideration.

Although this analysis is based on a nonrandom sample of WIA participants, there are several reasons why important policy conclusions can be drawn from these results. First, even if these estimates of program impact are only meaningful for those in the sample, the sample contains one in five WIA participants in the U.S. in the period of the study, so the results reported are relevant for a large number of individuals. Equally important, given that there is a diverse sample of states in the data, both in terms of location as well as in terms of the emphasis of the program, it seems likely that results in other states would display similar patterns. That is, even in the absence of formal analysis, it seems reasonable to conclude that the process by which WIA operates in these 12 states is similar to how it operates in other states. Therefore, the estimated effects across demographic groups and types of WIA services could be used by administrators in other states when deciding on how to allocate their own employment and training resources. Equally important is the fact that the estimated long-run impacts of WIA provide administrators and policy makers with information that simply cannot be obtained from the short-term performance measures that are currently produced.

[2] Mueser et al. (2007) find that labor markets within the same state are sufficiently similar to allow valid controls, although their analysis focuses on a single state.

The report is organized as follows. Section II describes the nonexperimental methodology employed in the evaluation as well as the study's implementation of the method. 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 presents results for separate subgroups. Section VIII provides a brief summary and conclusion, and section IX discusses further research to aid in developing job training policy.

II. Methodology

This section provides a discussion of how the evaluation problem is conceptualized in statistical terms. It then turns to a general review of matching methods used in the analysis reported here. The third subsection presents the study plan for implementing these methods.

1. Basic Evaluation Problem

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

One of the most extensive discussions of methods for estimating job training program effects is provided by James Heckman, Robert LaLonde and Jeffrey Smith in their contribution to the 1999 *Handbook of Labor Economics*. Discussions in Manski (1996) and Rosenbaum (2002) are also particularly insightful. The discussion here begins by considering the most basic models as they relate to the choice of nonexperimental methods for the WIA program evaluation.

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

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

Evaluations employing random assignment methods assure that the treatment is independent of Y_0 and Y_1 and the factors influencing them. The average program effect for those subject to random assignment may be estimated as the simple difference in outcomes for those assigned to treatment and those assigned to the control group. Where D is not independent of factors influencing Y_0, participants may differ from nonparticipants in many ways, including the effect of the program, so the simple difference in outcomes between participants and nonparticipants need not identify program impact for any definable group.

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

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

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

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

where Y_1 - Y_0 = ΔY is understood to be the program effect for a given individual and the expectation is across all participants with given characteristics. (This is known as the conditional independence assumption or the assumption of unconfoundedness in econometrics.) Matching and regression adjustment methods are all based on some version of (2). They differ in the methods used to obtain estimates of $E(Y_1|D=1, X)$ and $E(Y_0|D=0, X)$.[3, 4]

Until the late 1990s, the most common approach (e.g., Barnow, Cain and Goldberger, 1980) to estimating program impact was based on a simple linear specification assuming that the earnings function was the same for participants and the comparison group. Program impact was estimated, along with a vector of parameters of the linear earnings function, by fitting the equation

$$Y = X\beta + \delta D + e, \tag{3}$$

with e an error term independent of X and D. Although this approach can be pursued using more flexible functional forms, estimates of program impact rely on a parametric structure in order to compare participants and nonparticipants.

The critical question for regression adjustment is whether the functional form properly predicts what post-program wages would be for participants if they had not participated. Even under the maintained assumption in (1) that outcomes for participants and the comparison group do not differ once observed characteristics are controlled, if most of the comparison sample has characteristics that are quite distinct from those of the participants, regression adjustment will be

---

[3] Where concern focuses on program impact for nonparticipants or other subgroups, a stronger assumption than (1) is required. Normally, it is assumed that, conditional on X, both Y_0 and Y_1 are independent of participation.

[4] Although it is convenient to explicate estimation techniques in terms of a single population from which a subgroup receives the treatment, in practice treatment and comparison groups are often separately selected. The combined sample is therefore "choice based," and conditional probabilities calculated from the combined sample do not reflect the actual probabilities faced by individuals with given characteristics in the original universe. However, the methods used here can apply under choice-based sampling. In particular, if (1) applies in the population from which the treatment and comparison groups are drawn, (1) will also apply (in the probability limit) in the choice-based sample where the probability of inclusion differs for treated and untreated individuals but is otherwise unrelated to individual characteristics. The methods outlined here can be shown to be consistent for such a sample.

predicting outcomes for participants by extrapolation. If the functional relationships differ by values of X, the regression function may be poorly estimated, resulting in a potential bias.

2. Matching Methods

In recent years, more attention has focused on matching methods, which are designed to ensure that estimates of program impacts are based on outcome differences between comparable individuals. Although the basic idea behind matching can be traced to early statistical work, the development of matching methods in the last 25 years (Rosenbaum and Rubin, 1983) and recent increases in computing power have facilitated their wider implementation. (See also Rosenbaum, 2002 and Imbens, 2004 for general discussions of matching methods.)

The simplest form of statistical matching pairs each participant to a comparison group member with the same values on observed characteristics (X). In most cases, there are too many observed values of X to make such an approach feasible. A natural alternative is to compare cases that are "close" in terms of X, for which several matching approaches are possible. 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. The development of the propensity score approach is one of the most important innovations in developing workable matching methods. 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

$$Y_0 \perp\!\!\!\perp D \mid \mathrm{X} \implies Y_0 \perp\!\!\!\perp D \mid P(X).$$

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. Rather than attempting to match on all values of X, cases can be compared on the basis of propensity scores alone. In particular,

$$E(\Delta Y | P) = E_p(E(\Delta Y | X)),$$

where E_P indicates the expectation across values of X for which $P(X)=P$ in the combined sample.

This implies that

$$\begin{aligned} E\left(\Delta Y \middle| D=1\right) &= E_{X|D=1}\left(\Delta Y \middle| P\left(X\right)\right) \\ &= E_{X|D=1}\left(E(Y_1 \mid P(X), D=1) - E(Y_0 \mid P(X), D=0)\right), \end{aligned} \qquad (4)$$

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. Although other balancing scores could serve the same function, the propensity score is very convenient.

In practice, the propensity score must be estimated. Normally, a logit or probit function is used for this purpose, and it is critical that a flexible functional form be used. 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 in a period following participation and earnings prior to program participation, and for comparison cases the earnings difference is calculated over the same periods. Even if individuals who participate in WIA differ in important ways from those in the comparison group, so long as such differences are stable over time in their impact on earnings, this specification can eliminate bias resulting from differences between participants and others.

The approach can be illustrated by considering the case where individuals are selected into the program on the basis of an unmeasured factor that influences earnings. In particular, take prior earnings Y_0^{-1} as well as subsequent earnings to be influenced by the individual-specific factor *δ,*

$$Y_0^{-1} = Y_0^{-1\prime} + \delta$$
$$Y_0 = Y_0' + \delta$$
$$Y_1 = Y_1' + \delta$$

where primes denote unmeasured components of earnings that satisfy condition (1). If selection into the program is based both on *X* and *δ,* condition (1) will not be satisfied. If, for example, those with higher values of *δ* are more likely to participate in the program,

$$E(Y_0 \mid D = 1, X) > E(Y_0 \mid D = 0, X),$$

with the result that $E(Y_1 | D=1, X) - E(Y_0 | D=0, X)$ is an upward biased estimate of the program effect. In this case, however, the change in earnings over time, $Y_0 - Y_0^{-1} = Y_0' - Y_0^{-1\prime}$, can be shown to meet the condition:

$$Y_0 - Y_0^{-1} \perp D \mid X \,. \tag{5}$$

In short, if the independence assumption (1) is violated because fixed effects differ for those receiving services and the comparison group, assumption (5) will still hold for the differenced measures. The program impact can then be estimated as

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

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.

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. Only if growth in earnings is not associated with a particular personal characteristic is it appropriate to omit that measure.

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. If prior earnings incorporate transitory shocks that differ for participants and comparison group members, since difference-in-difference estimation interprets such shocks as representing stable differences, estimates will contain a transitory component that does not represent the true program effect. More generally, the difference-in-difference estimates need to be understood as one of several estimates that make different assumptions. Where alternative estimates produce different results, the ability to infer actual program effects will be limited.

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. For each treated case, a case in the comparison group that was most similar to that case—in terms of propensity score, in the case of propensity score matching—would be matched with it. In effect, the program impact for a particular treated case *i* was estimated as $Y_{1i} - Y_{0j(i)}$, where $Y_{0j(i)}$ is the outcome for the comparison case that is matched with the treated case *i*. The estimated program impact was obtained as the average of this measure over all treated cases. Such pairwise matching was usually performed using sampling from the comparison group without replacement, meaning that each comparison group member could be included as a matched case only once. More recently, studies have used sampling with replacement, allowing for one comparison case to serve as the match for more than one treated case.

Although pairwise matching is most intuitive, in recent years alternative approaches have been recognized as superior. In particular, in contrast to matching one comparison group case with a given treated case, 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. As suggested earlier, 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

$$E(\Delta Y) = \frac{1}{N}\sum_{i=1}^{N}[Y_{1i} - \overline{Y}_{0j(i)}] \qquad (6)$$

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.

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. With choice-based sampling, the number of treated and comparison cases does not reflect the likelihood that an individual with given characteristics participates in the program in the full universe but rather is determined by the various factors outside the control—and knowledge—of the researcher. Matching on the log odds of the propensity score assures that results are invariant to choice-based sampling.

Matching on the log odds of the propensity score also has the advantage that it "spreads out" the density of very low or very high propensity scores. In a design with a very large sample of comparison cases, Mueser et al. found that in order to assure good matches it was necessary to use much smaller bandwidths at low propensity scores. Use of the log odds allows for a consistent bandwidth to be used. 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 is expected to translate into the same metric at different propensity score levels.[5]

Although the theory underlying propensity score matching implies that, in the limit as a sample grows, matching on the propensity score also matches on all control variables, in any given application with a finite sample, there is no assurance that matches will be close enough to remove significant differences. In addition, since applications by necessity use a parametric structure to calculate the propensity score, inadequacies in the estimation method may cause further deviations. It is therefore necessary to compare the treated cases with the matched comparison cases. In general, if differences are too great, it may be necessary to alter the caliper used in the analysis or to modify the details of how the propensity score is estimated.

Alternative matching approaches. In light of recent research (Mueser et al., 2007; Imbens and Wooldridge, 2008) suggesting that with sufficient sample overlap and well-balanced covariate distributions, impact estimates should be relatively insensitive to the details of how matching is undertaken, only a brief review of alternative matching approaches is provided here. (See Mueser et al. (2007) for a detailed comparison of alternative matching methods based on a state sample of JTPA participants and a matched sample of job searchers that correspond closely to the data used in the current evaluation.)

Mahalanobis distance has the property that matching will reduce differences between groups by an equal percentage for each variable in X, assuming that the covariance matrix for X is the same for the two groups. This ensures that the difference between the two groups in any linear function will be reduced (Rosenbaum and Rubin, 1985). In terms of theory, Angrist and Hahn (1999) argue that there is little basis for preferring propensity score matching to Mahalanobis matching or to other metrics, but Mueser et al. (2007) found that Mahalanobis distance matching was appreciably less efficient than propensity score matching.

As far back as Rosenbaum and Rubin (1985), it has been suggested that matching be undertaken on the propensity score combined with other variables. Zhao (2004) has proposed alternative metrics that weight variables by coefficients reflecting their impact on outcome measures as well as propensity score. Such an approach improves matches for those variables that could cause outcomes to differ for treatment and comparison cases. His Monte Carlo experiments suggest that in small samples his alternative metrics may have benefits, but there is no approach that

[5] Since the odds ratio is a continuous transformation of the propensity score in the relevant range, where participant and comparison cases are matched closely, results using this approach will be very similar to methods that use propensity score, so long as the matching works well.

dominates all the others across environments.

Validity of impact estimates. Dehejia and Wahba (1999, 2002), applying matching methods to data from the National Supported Work demonstration project (originally analyzed by La Londe, 1986), present a strong case in support of these methods for evaluating job training programs. 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.[6]

Bloom, Michalopolos and Hill (2005) summarize studies that use experimental evaluations of program impacts and compare these with nonexperimental studies. They observe 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 a juncture 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 more 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

[6] Mueser et al. (2007) compare nonexperimental JTPA program impact estimates in a single state with random assignment estimates in a dispersed set of sites.

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. 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.

A recent attempt to summarize the literature addressing comparison group study design appears in the materials distributed by the Evidence-Based Policy Coalition: "If the program and comparison groups differ markedly in key pre-program characteristics (e.g., employment, earnings), the study is unlikely to produce valid results…Importantly, this is true even when statistical methods such as propensity score matching and regression adjustment are used to equate the two groups."[7] This incorrectly interprets these studies' results. Although similarity in pre-matching measures may, in many cases, increase the likelihood that unmeasured differences are small, it is neither necessary nor sufficient to assure this condition.

In their widely cited work on the use of matching to nonexperimentally estimate program impacts, Heckman, Ichimura and Todd (1997) explain that "the evaluation problem is a missing data problem." Most basically, individuals both participating and not participating in programs at the same time cannot be observed. It is also common, however, even in experimental evaluations, for other problems to arise that lead to missing data and contribute to bias in the estimation of program impacts. For example, in the National JTPA Study (Orr et al., 1996), not all of the individuals randomly assigned to the experimental "treatment" group received JTPA services, and data on about one-fourth of the original sample were not available for the 18-month impact evaluation. In other words, any attempt to estimate public employment and training program impacts will almost always compare imperfectly matched participants and nonparticipants.

As is the case with any nonexperimental evaluation method, the richness and relevance of the data available for the evaluation have important implications for the performance of the estimators. More recently, 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. A key insight of these studies for this evaluation is that if relevant 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 should be of relatively minor import.

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? As noted

[7] "Brief Overview: Which Comparison-Group Study Designs Are Most Likely to Produce Valid Estimates of a Program's Impact?" attachment provided to reviewers evaluating program evaluation study designs, December 2008, p. 1.

above, 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 this kind of 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, the analyses 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. [8]

3. Plan of Analysis

As noted above, 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. This study evaluates two WIA programs: the Adult program,

---

[8] The National JTPA Study (NJS), which used a randomized experimental design to estimate program impacts, likewise made no claim to estimate impacts of the full JTPA program. Rather, the NJS merely noted that their sample of sites (JTPA service delivery areas) "collectively reflect the diversity of the JTPA system," including geographic, environmental and operational diversity (Doolittle et al., 1993, p. 9). The NJS report noted important ways in which those experimental results were unlikely to generalize to the national JTPA system, including the fact that no large, central city was included among its sites. The sample used here has important advantages over the NJS in terms of its representativeness of the population of program participants. The NJS included just 16 (2.5%) of more than 600 local service delivery areas (in 16 states), whereas this current evaluation includes data from more than 100 of 568 local Workforce Investment Areas in 12 states. In addition, compared to the combined sample of experimental and comparison group members of 14,441 individuals in the JTPA impact evaluation, the current study analyzes nearly 160,000 WIA participants and nearly 3 million comparison group members.

serving largely disadvantaged individuals, and the Dislocated Worker program, serving workers who have lost jobs.[9] 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) 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.[10] As of 2005, about 43 percent of WIA exiters nationwide are coded as receiving Training services.[11] 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.[12] 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 allow for the identification of such differences.

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

---

[9] The Youth program, which serves those under age 22, was not evaluated. Given that the character and size of the Youth program differs across states more than the programs serving adults, and that program goals are not clearly focused on the labor market, ETA chose to omit evaluation of this program. However, analyses of those under age 26 in the Adult and Dislocated Worker programs are undertaken in the subgroup analysis.

[10] The strict sequential structure of services may not be followed in all sites.

[11] Of the 230,446 exiters from the Adult program for April 2005-March 2006, 105,457 are reported to have received Training; and of 210,117 exiters from the Dislocated Worker program during that period, 83,699 are reported to have received training. These figures are based on analyses of the WIASRD data undertaken by Social Policy Research Associates (2007, pp. 36, 114).

[12] In many areas, WIA Core services are not readily distinguished from services provided under ES, which are available to anyone who seeks them. To some degree, the estimate of the impact of WIA obtained in the current analyses will be "watered down" because of this, but it nonetheless properly represents the overall impact of WIA. Insofar as WIA fails to offer services that extend beyond those readily available elsewhere, it is appropriate that impact estimates are reduced.

earnings and employment, reflecting participants' involvement in program activities rather than employment. These have been labeled "lock-in" effects. Later earnings effects are expected to be positive, as skills obtained during the program interact with job experience. The 16-quarter follow-up from program entry allows the growth and persistence of program effects to be examined and produces more accurate estimates of the benefits of participation. Where data are available, the difference-in-difference model is estimated, taking the dependent variable as the difference between earnings or employment in the outcome quarter and the level prior to program participation. In particular, two difference-in-difference codings will be used, taking the dependent variable as the growth in earnings from the tenth quarter prior to participation and from the sixteenth quarter prior to participation.

Comparison pool. The analysis uses 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). 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.[13]

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 will be 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.

The estimation approach used in the current analysis 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) (page 4), or equation (5) in the difference-in-difference specification (page 7). If this assumption is violated, that is, if a matched

[13] Even if their effects are small, ES services may well offer benefits that are significant relative to their very modest costs.

treated case and comparison individual would have had different earnings in the absence of the treatment, the impact estimate will contain bias.[14] 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. Several recent analyses (Friedlander and Robins, 1995; Heckman and Smith, 1999) have stressed the importance of choosing a comparison group in the same labor market. 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.[15] 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.[16] 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.

In addition to providing separate analyses by gender, in states with substantial Hispanic populations, and in states with substantial (non-Hispanic) nonwhite populations, separate analyses for these groups are undertaken. Separate analyses for those under age 26 and those ages 50 and above are included to determine if the benefits for younger and older individuals differ from those of other participants. Finally, separate analyses are undertaken for U.S. military veterans.

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

[15] 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, Ichimura and Todd, 1997; Heckman, et al., 1998; Heckman, LaLonde and Smith, 1999; Heckman and Smith, 1999).

[16] 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 the next section.

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 (if available)
- program participation history (WIA, UI, ES)
- current and prior TANF recipient
- time since layoff (when available)

Treatment and comparison samples. Table II.1 shows treatment samples and the comparison groups used in each case. 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 identified below.

Line 1 lists comparisons of WIA participants—regardless of services received—with samples of 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 in the same state. This comparison allows one to identify the extent to which training, per se, is associated with employment and earnings outcomes.

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 (6) on page 8.

Given the benefits noted above, matching is undertaken on the log odds ratio, which is a simple monotonic function of the propensity score. 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. 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, is 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 always based on a comparison of individuals during the same time period.

Linear adjustment. Whatever matching algorithm is used, there will generally be some difference in the conditioning variables between participant and comparison cases. Although earlier work occasionally employed local linear regression adjustments to address this issue (Heckman, et al., 1998), adjustments based on a linear model have now become common following recent theoretical work by Abadie and Imbens (2006a) arguing that such an approach removes an asymptotic bias inherent in simple matching estimators. The method fits a linear model in the comparison sample, and then uses coefficients estimated in this sample to adjust for any differences in characteristic means that exist between the treatment and comparison samples. Although the discussion in Abadie and Imbens suggests that such "bias adjustment" is likely to be less important in large samples like those here, these adjustments are made to ensure consistency with the existing literature.

Common support diagnostics. Frequently, matching estimates are limited by a failure of overlap in the distributions of variables for treated and comparison cases. This occurs when some individuals in the treatment group have characteristics or combinations of characteristics that cannot be matched by those of individuals in the comparison group. In such cases where "common support" fails, regression techniques produce impact estimates, but their validity is suspect.

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). Following Abadie and Imbens (2006a), Imbens and Wooldridge (2008), Imbens (2008), the analyses here use an analytical formula for calculating standard errors that is asymptotically correct.

Although Abadie and Imbens (2006b) argue that bootstrap standard errors may produce flawed estimates of sampling errors, it is unclear whether these problems are relevant for the matching estimators used here.[17] Given that the small sample properties of these alternatives are not known, in order to assure the validity of estimates, bootstrap standard errors are calculated for selected subsamples.

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.[18]

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 with a state's employment exchange agency). In addition to facilitating the construction of comparison groups, these data were used to control program participation prior to the quarter of program entry. In all but three states, at least six quarters of such information is available prior to the first quarter of program participation.

[17] Abadie and Imbens (2006a) show that for matching estimates using a fixed number of matches, bootstrap standard errors are asymptotically biased. However, there is no work indicating whether bootstrap standard errors for radius matching methods are consistent. See Imbens and Wooldridge (2008).

[18] 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, selection may be an issue for 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 through June 2005.

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.[19]

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.[20]

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; 63,515 individuals entered the Dislocated Worker program, producing a total of 64,089 total program entries.[21] The rightmost column identifies the number of individuals who participate in comparison programs and are available to be matched to program participants. The upper column indicates that approximately 2.9 million unique individuals are available, contributing nearly 6.2 million quarters of program activity.[22] A very large number of comparison cases is available for matching.

[19] Further discussion of the UI and ES samples is provided in section II.3 above.

[20] In one state, 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.

[21] Where an individual entered the program more than once during a quarter, this was considered to be only a single entry. Data cleaning eliminated multiple entries when these appeared to be due to data entry errors or when they pertained to the same set of services. For example, when a second entry date occurred prior to the exit date associated with the prior entry, the second entry date was omitted.

[22] 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			Comparison Group
	Overall	No Training	Training	Overall	No Training	Training	
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, it can be seen 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.[23] 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

[23] Measures of statistical significance are not reported because, with these large sample sizes, even substantively unimportant differences in means are statistically significant.

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 the difference is 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 much 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.

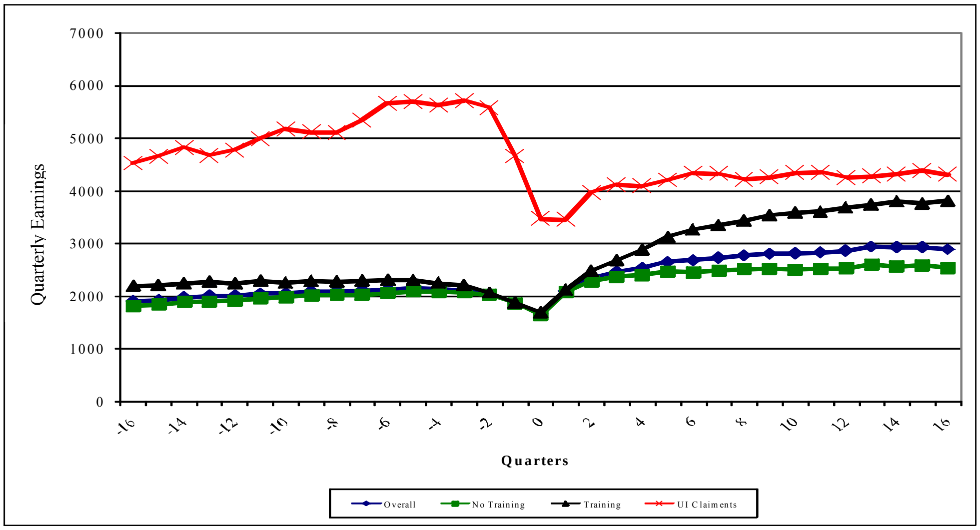
The bottom panel of the table shows that 4 to 5 percent of WIA entrants had previously participated in WIA, and that the number participating in the comparison program was substantial. About a fifth of Adult program participants had prior comparison program experience, compared to over two-fifths of Dislocated Workers. By definition, a comparison case participates in the comparison program in the specified quarter. The figure 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.[24] 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.

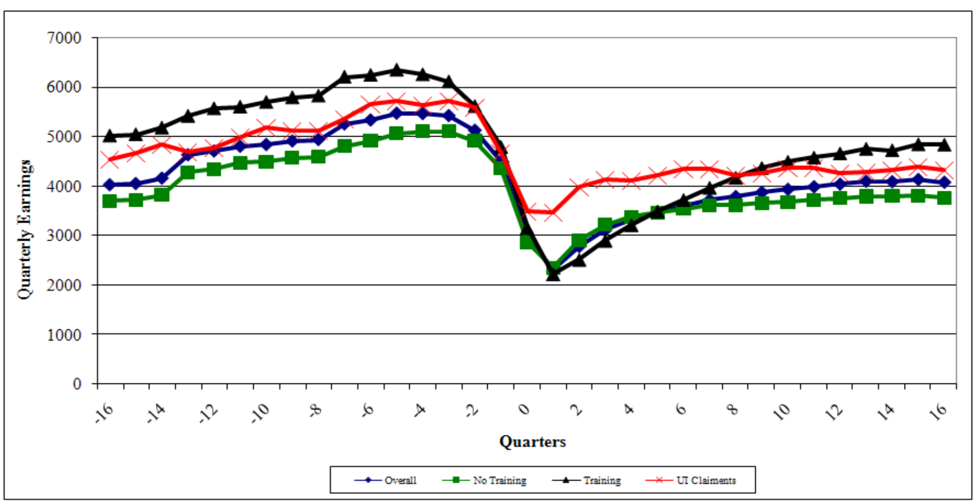
Notwithstanding these differences, there are important similarities in the patterns of earnings for individuals in these states. Figure III.1 graphs quarterly earnings for WIA Adult program participants and the sample of individuals in the comparison group. Figure III.2 provides comparable plots for 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

[24] Any WIA participant who does not receive Training services is coded as receiving Core/Intensive services only. Data codes available do not allow individuals who enter the WIA program but leave without receiving any services to be identified.

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



who receive Training services, WIA participants who do not receive Training services and comparison cases.[25]

The most notable pattern in both figures is the decline in earnings that occurs in the several quarters prior to entry into WIA, 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 as well.

There remain differences between the Adult and Dislocated Work patterns, of course. The dip in earnings is quite modest in the case of Adult program participants, and much more pronounced for Dislocated Workers. In the Adult program (Figure III.1) earnings are markedly lower than the comparison group, although the difference in earnings is much smaller in the quarters after entry into the programs. In the case of Dislocated Workers (Figure III.2), prior earnings are quite similar to the comparison group, but participants had appreciably lower earnings than the comparison group immediately following program participation.

One way to estimate the impact of a program would be to consider the change in WIA participants at the start of the program. The growth in earnings at the point of program entry would be taken as evidence that the program increased earnings. This is clearly inappropriate, however, as those in the comparison group experienced earnings growth as well. The decline in earnings and then recovery are consistent with program selection, since even in the absence of program effects, earnings are expected to return to prior levels.

An alternative is to compare the path for WIA participants with that for the comparison group. Prior earnings for comparison group members are appreciably above WIA Adult participants, and the gap declines following program participation. Although such an approach is clearly an improvement over an examination of simple trends, it implies a comparison of individuals who are different in terms of prior earnings as well as a variety of other characteristics, as indicated in Table III.1. Such differences may well cause their experiences to diverge.

The matching analysis introduced above and implemented 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. It is important to recognize that differences between the WIA participants and the comparison group

[25] 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 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 sample differences. Appendix 5 provides mean earnings by program participation and level of service.

are in themselves not of importance if matching is successful. Those comparison group cases with incomes that are higher than those for WIA participants are omitted by the matching procedure, so their presence in the original comparison group is unimportant. In the analyses presented below, average earnings are essentially identical for treated individuals and the matched comparison group.

The validity of the matched comparison group rests on the particular variables that are used for matching. 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. Controls for prior labor market experience and standard demographic measures are consistent with the recommendations of the literature.[26]

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 between the treated and matched comparison groups will be eliminated. Specification tests and comparisons of results for various combinations of states are an essential part of the analyses that follow.

The aggregate numbers presented in Table III.1 hide large 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.[27] 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.

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 receive training. The comparison group comprises 1.6 million individuals who contribute 3.5 million records available for matching.

---

[26] See the discussion in section II.2.

[27] 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

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

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 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.

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 relative economic position of ES participants relative to the WIA treatment groups. Whereas in Table III.1, comparison group (three-quarter of them UI claimants) prior earnings were two to three times the size of earnings for Adult program participants, ES earnings are only 50 percent higher. Furthermore, ES earnings are much lower than Dislocated Workers earnings; in contrast, 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.

It is possible that the differences between the UI and ES comparison samples are not 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. Insofar as results are similar for UI and ES states, this supports the view that differences in the samples are not important.

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

IV. Matching Analysis: Details of Implementation

This section provides a detailed discussion of how the matching methods were implemented. The subsection below describes exactly how WIA and comparison group units used for matching are structured, as well as the detailed coding of variables. It also describes the approach used to estimate standard errors, describing some of pilot analyses, as well as initial work undertaken to determine the appropriate radius value for matching. The next subsection provides statistical results related to the matching procedures, including discussion of the diagnostics used to assure that the matching results are valid. The third subsection provides an explanation of how results are presented.

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.[28]

Matched samples. For comparisons 1-2 listed in Table II.1, the comparison group comprises participants in an alternative program—either as UI claimants or as 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

[28] Measures of overall program impact therefore identify the incremental impact of program entry. For an individual who enters WIA a second time after the initial entry, the impact estimate for the first entry includes the effect of both spells of WIA participation, whereas for the second entry, the impact is for that entry alone. In practice, the number of such cases is so small that results are insensitive to the way in which multiple entry individuals are treated.

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 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. The discussion in Section III makes clear that WIA participants, especially those in the Adult program, differ in important ways from those in the comparison groups. In addition to gender, differences in race, age and educational attainment could produce substantial differences in outcomes. Even more important are differences in prior employment experience. As noted above, earnings of participants in the WIA Adult program in the year before program entry are less than half of those in the WIA Dislocated Worker program or comparison program participants. Differences in prior program experience may also indicate differences in how individuals confront their labor market opportunities that will be associated with their ultimate earnings. The benefit of these data is that they provide extensive controls for these factors.

This method fits 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 the methods of this study 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.[29] 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

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

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. A more detailed discussion of variable coding is provided in Appendix 1.

Standard errors.[30] It is necessary to obtain an indicator of the importance of sampling error in the estimates. Initial analyses undertaken for two states 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.[31] 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. The true value of the unconditional standard error must exceed the conditional standard error, but there is no certainty this will be the case for the estimates obtained in any one sample. Both approaches are somewhat involved and require use of a matching algorithm that is as computer-intensive as that required to obtain program impact estimates.

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 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. For example, these methods require that the propensity score be estimated, which induces a source of error for which there is no explicit accounting. Also, each individual in the comparison program may contribute multiple cases to be matched with treated cases, so the assumption that each case is independent is not met. Bootstrap standard errors may capture such factors. However, there are several important limitations to bootstrap standard errors. As is the case for estimation procedures for analytical standard errors, theorems supporting the use of bootstrapping show that bootstrapping produces error estimates that are *asymptotically* unbiased. For finite samples, there is no certainty that estimates will be unbiased. An additional problem of bootstrapping is the intensive computer resources required to estimate them. In the case of the analyses reported here, estimating bootstrap standard errors for all of the states included in these analyses was simply not feasible.

---

[30] Further detail on the estimation of standard errors is provided in Appendix 2.

[31] In the implementation here, following Imbens (2008), it is conditional on the propensity score estimate. See Appendix 2 for a more extensive discussion of these issues.

Methods used here for estimating the bootstrap standard errors attempted to incorporate all sources of error that could influence the estimates. In general, the bootstrap relies on sampling from the analysis sample with replacement, replicating the analysis multiple times. The estimated standard error is the standard deviation of the estimated impact estimate across replications. Here, 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. In each case, estimates were based on 25 replications, reflecting the computer-intensive nature of these replications. Across the two states chosen for these analysis, 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.

Focusing on estimates without bias adjustment, on average, the bootstrap standard errors 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 was between 15 and 20 percent in these states.

For most comparisons, bootstrap estimates of standard errors for the biased-adjusted impact estimates were of similar size. However, for several of the comparisons involving small numbers of cases, the bootstrap methods were clearly failing, producing standard errors that were often more than 10 times the size of other estimates. This highlights the fact that for small samples and certain designs, the bootstrap estimates may not be meaningful.[32]

These experiments with bootstrap standard errors do not suggest that the results obtained using the Imbens standard errors are misleading. These findings are similar to those of Heinrich (2007), who reported only modest differences between alternative standard error estimates.

[32] The reason for these large bootstrap standard errors in the case of bias adjustment is clear. In each replication, the bootstrap chooses from the original sample with replacement, reducing the effective number of degrees of freedom. Where the sample size is only somewhat larger than the number of independent variables, the bootstrap replications produce data that is perfectly predicted in the linear bias adjustment equation. The bias adjustment is therefore meaningless.

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 to match. 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. The analysis here considers the MISE based on weighting the comparison sample by the inverse odds of the propensity score. Technical details of these methods appear in the Appendix 3. Comparisons on two states were undertaken, using earnings in the fifth and tenth quarters after program entry as the dependent variables. MISE is 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.

Across almost all comparisons, larger radius values produced smaller MISE values, so that the MISE was smallest for the radius value of 0.1. The analysis also showed that, when smaller radius values were used, many treated cases could not be matched. The MISE declined relatively quickly as the radius increased in size from 0.001 to 0.003, but tended to decline progressively more slowly as the radius increased further. Differences in the MISE were relatively modest when the radius value increased from 0.03 to 0.1, with declines in the MISE almost always less than 10 percent. Based on these results, in the analysis that follows, the radius value will be set to 0.1 as an initial default.

However, in choosing the radius for the analyses reported below, these considerations do not override concern that the approach successfully matches 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 any one dependent variable. However, 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.[33]

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. In fitting the logit for the propensity score, all available independent variables were included in linear form along with selected 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 squares, 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

[33] See Appendices 1 and 3 for further detail.

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 three million individuals in the comparison program, contributing over six million matching quarters, and approximately 160,000 in the WIA Adult and Dislocated Work programs. The very large reservoir of comparison cases assures that close matches will be feasible for most WIA participants, but differences in the distributions of variables identified in Table III.1 suggest that many of these 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 omitted because they were among those cases that 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.

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 numbers 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.

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

3. Presentation of Results

The impact estimate based on matching is calculated as the average difference between the outcome measure for treated individuals and that for the matched comparison group. Bias adjustment was also applied to impact estimates that are reported, although, given that the matches are very close, it makes very little difference.[34]

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

[34] For each matching group, a regression is fitted predicting the dependent variable on the matched (weighted) sample of comparison cases using all variables in the logit. The differences in the mean variable values between treated and matched comparison cases are then used in conjunction with the regression coefficients to adjust the impact estimates. In some cases where sample sizes were small, the bias adjustment was not possible, and estimates reported do not include the bias adjustment. In no cases were the bias adjustments applied to estimates large enough to influence the substantive conclusions of the analysis.

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 obtained for the difference between the treated sample and the matched comparison sample on prior earnings measures, focusing 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, since, if the program "effect" on prior earnings 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 earnings are therefore suspect. In each case where estimates are discussed, estimates on prior earnings 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.

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. These standard errors attempt to identify how estimates of average impact would vary if a set of observationally similar individuals were selected from a universe of potential applicants. For the main estimates, confidence intervals are provided, allowing conventional significance tests to be performed.[35]

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

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

terms of demographic characteristics, prior employment (previous eight quarters), and prior program experience.

There is expected to be a great deal of heterogeneity in the effects across individuals. Some individuals receive relatively minimal services, which may differ little from those available through the ES program, whereas others receive extensive help, including training that extends over a period of many months. Equally important, response to training will differ across individuals, with some obtaining valuable skills that produce returns in the labor market, while others never obtain a level of expertise that is useful on a job. Finally, the program may supplant valuable job search and experience, producing a negative effect even for an individual who is relatively successful following participation. Impacts to be estimated are therefore understood to be average impacts across individuals.

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.[36]

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 $100 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 generally somewhat larger for most states.

[36] 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 for states show that there is substantial variation across states that cannot be explained by sampling error, sampling error is still large in many cases. Looking at combined effects allows one to examine overall patterns in a context where sampling error plays a much smaller role. 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.[37] Also on the graph are dashed lines that show the confidence interval for each estimate. The lower dashed line subtracts

[37] 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.

twice the Imbens standard error from the estimate, and the upper dashed line adds twice the standard error. Subject to well-known assumptions, the gap between the two dashed lines provides a measure of the extent to which the true value of the underlying impact may differ from the estimate due to sampling error. When the confidence interval does not contain zero, the impact estimate is statistically significantly different from zero at the 0.05 level of significance.[38] Also presented in this figure is the "impact" estimate for 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.[39]

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.

[38] For a discussion of the Imbens standard error and the basis for using it in the current study, see section IV.1. Formally, it is not possible to conclude that the true value lies within the confidence interval with a certain probability; rather, one knows that if the true value is inside the confidence interval, the observed estimate would not cause one to reject the null hypothesis for that true value at the specified level of statistical significance. The confidence intervals used in the study (two standard errors above and below the point estimate) imply a p value of 0.0455.

[39] See section IV.3

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

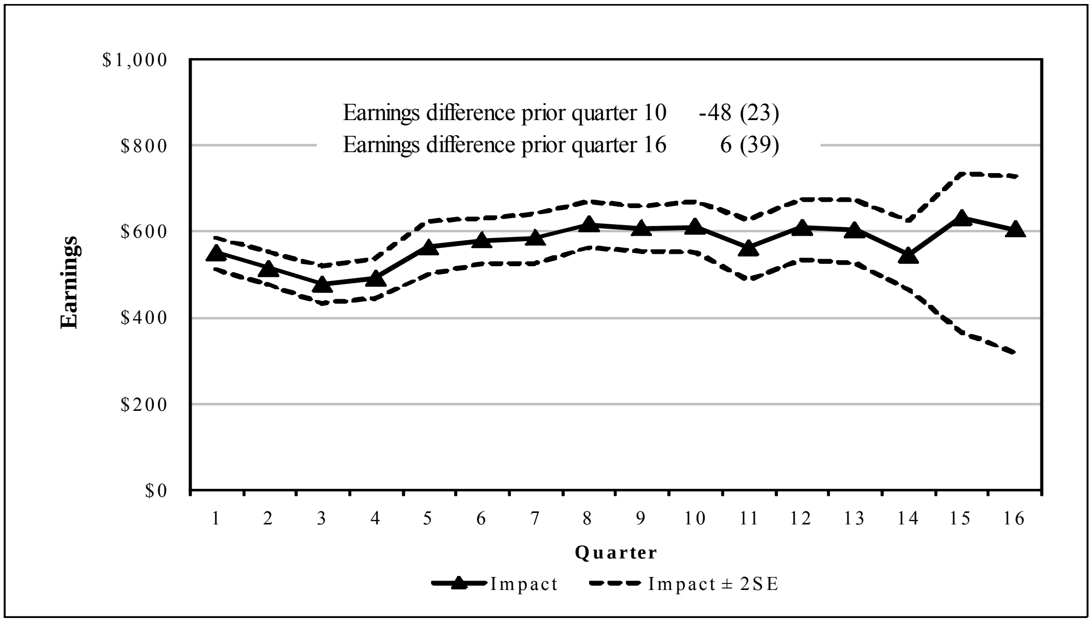
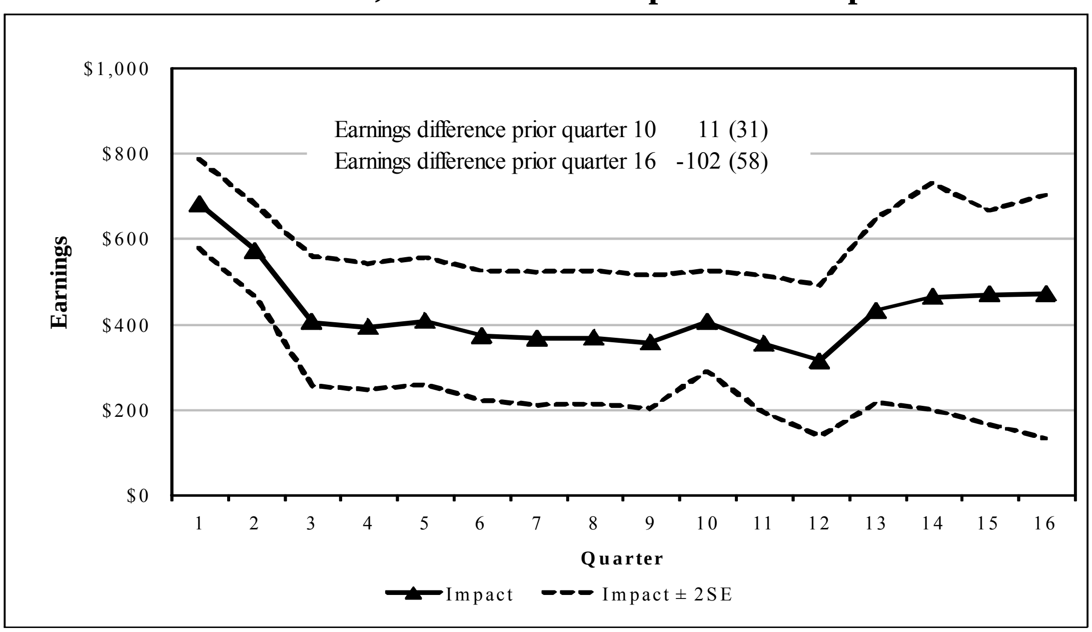


Figure V.2
Adult Program Treatment Effect on Quarterly Earnings For Males, WIA versus Comparison Group



Figures V.3 and V.4 provide analogous estimates for employment. The basic structure of 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 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. Given the confidence interval, it is unwise to place much emphasis on the observed differences, however.

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 somewhat 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 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, 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 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, although there are large differences across states, 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.

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

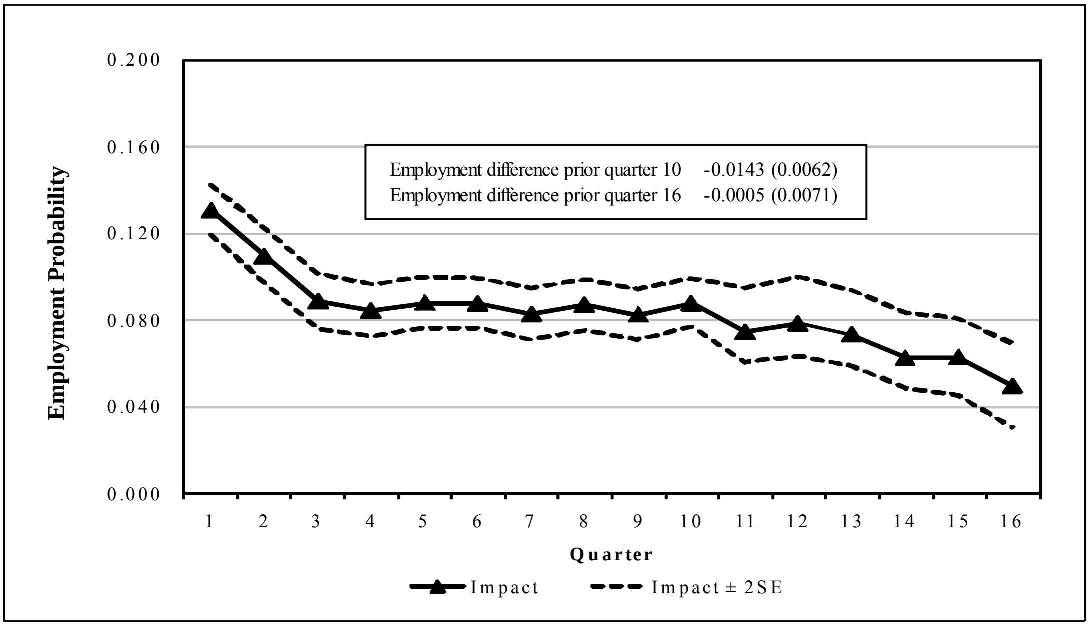


Figure V.4
Adult Program Treatment Effect on Quarterly Employment for Males, WIA versus Comparison Group

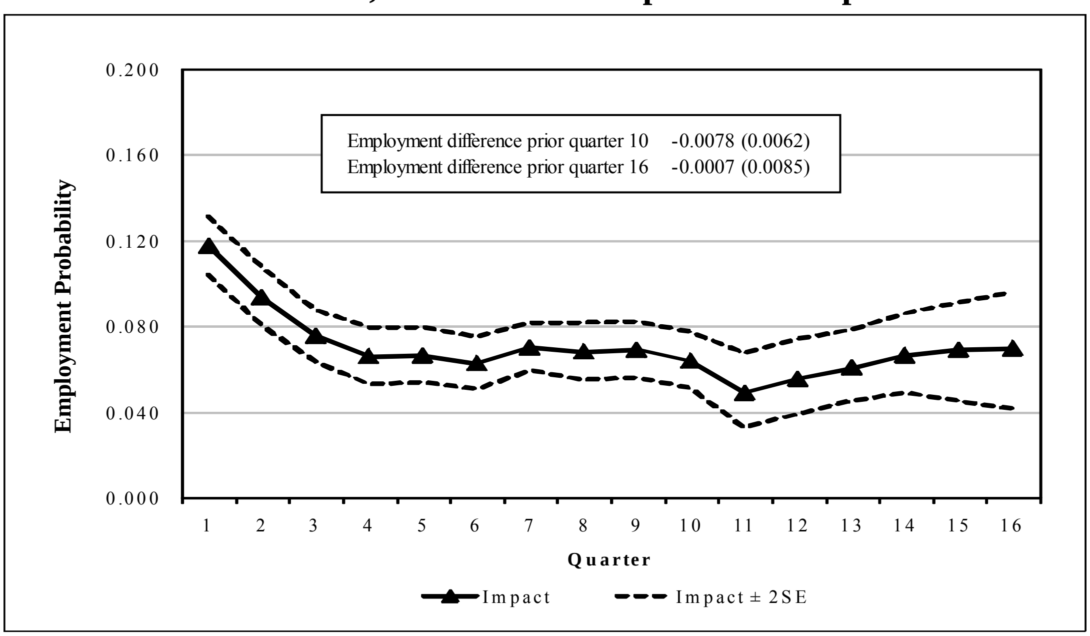


Figure V.5
Adult Program Treatment Effect on Quarterly Earnings for Females, WIA versus Comparison Group in 7 High Training States

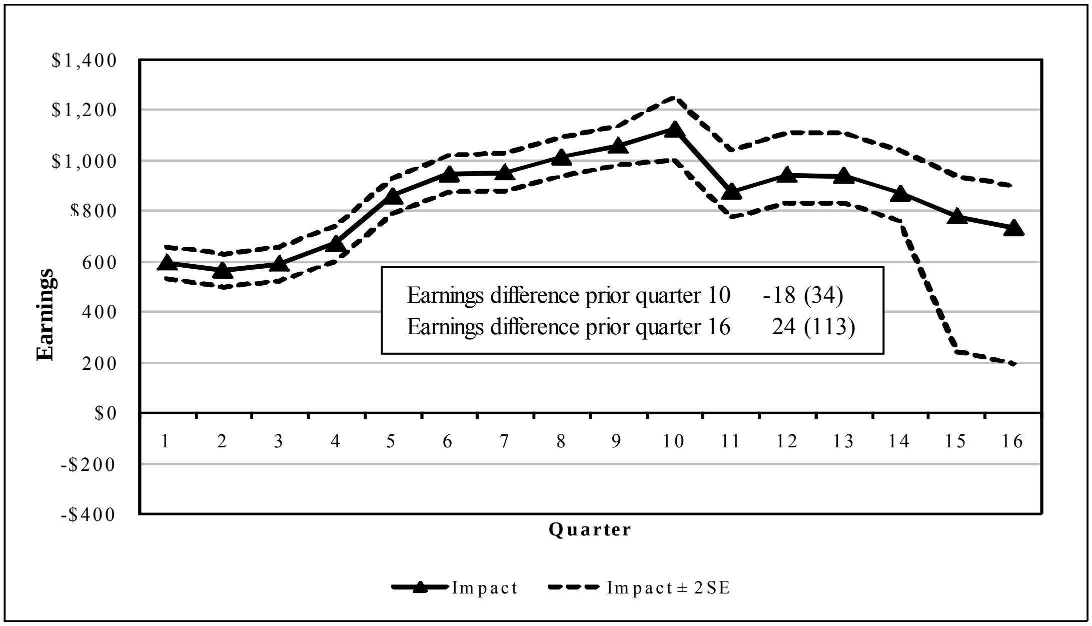
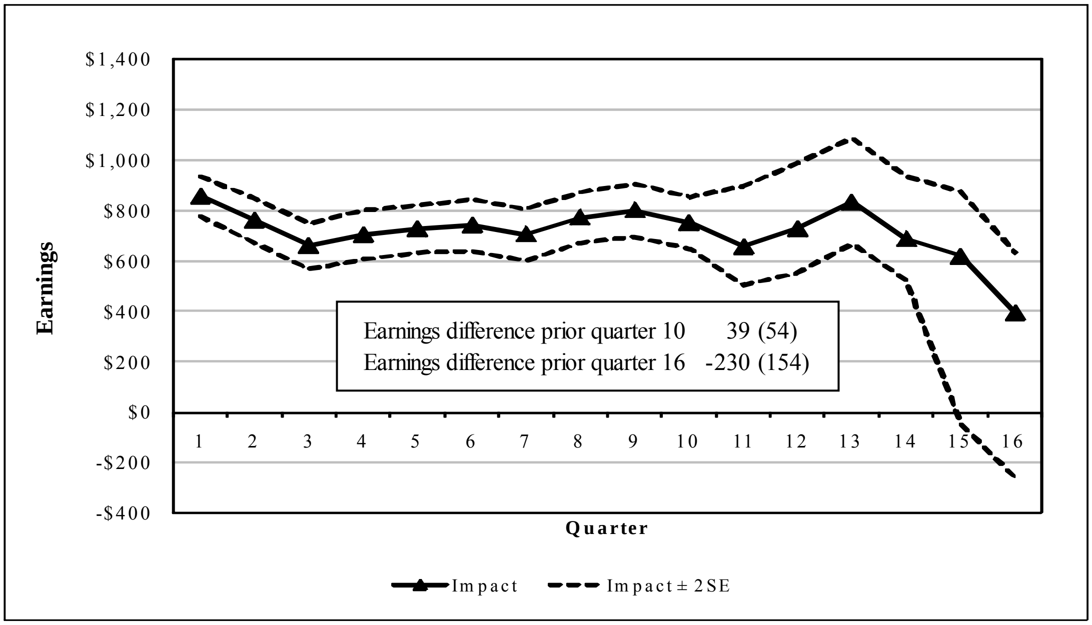


Figure V.6
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA versus Comparison Group in 7 High Training States



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. Each figure reports 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. WIA participants have earnings that are about $100 below those of the comparison group. This means that when a difference-in-difference model is estimated, the calculated program impact is increased by $100. 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 UI claimants are often required to register for ES services, those awaiting recall are exempt from this requirement, removing 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. 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.

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

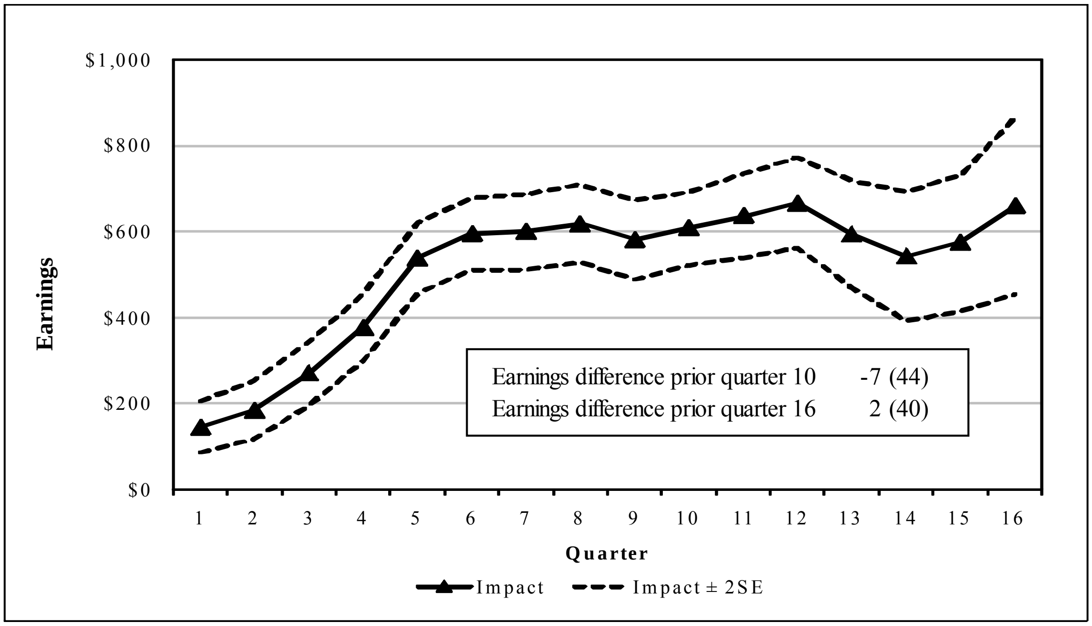
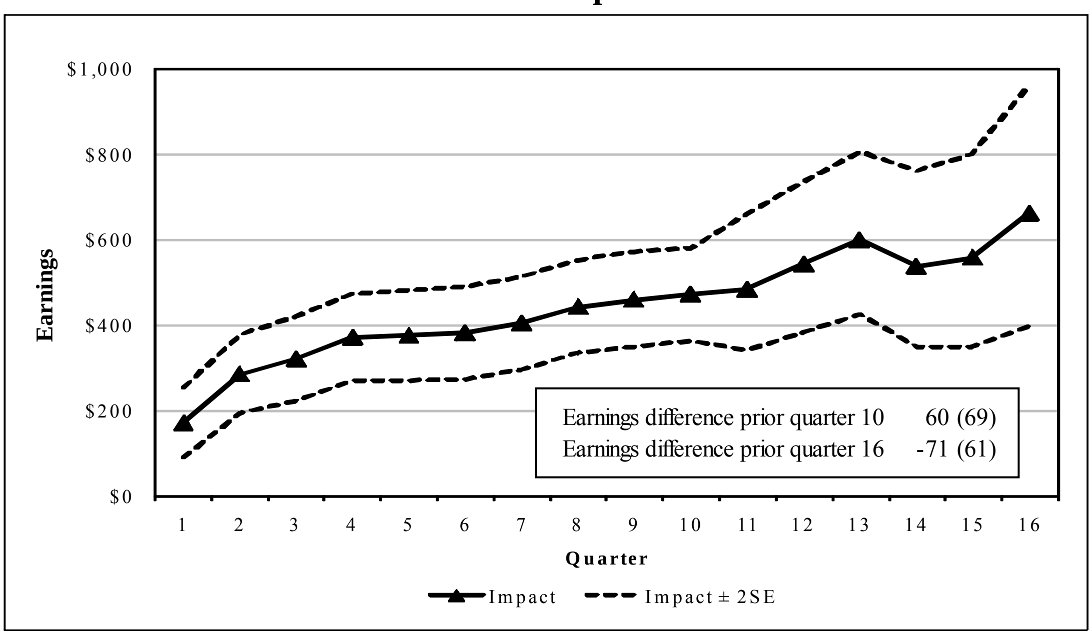


Figure V.8
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA versus ES Participants in 3 States



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. The overarching question of whether these estimates of program impact are valid is further addressed below.

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 are Core services, which are generally similar to the services available without restriction as part of the ES program. The next level of services are 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 in which 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. Once a client is accepted into the WIA Adult program, the staff in the agency is aware that client success in obtaining employment will be recorded upon exit from the system. 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,[40] 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, 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 more quickly than the earlier

[40] 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

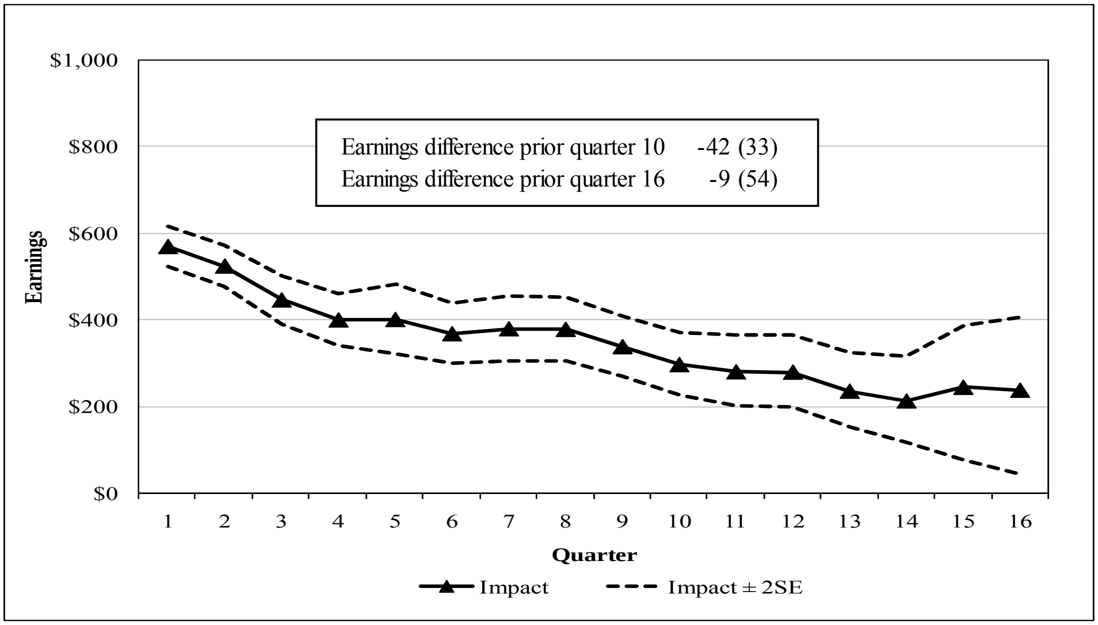
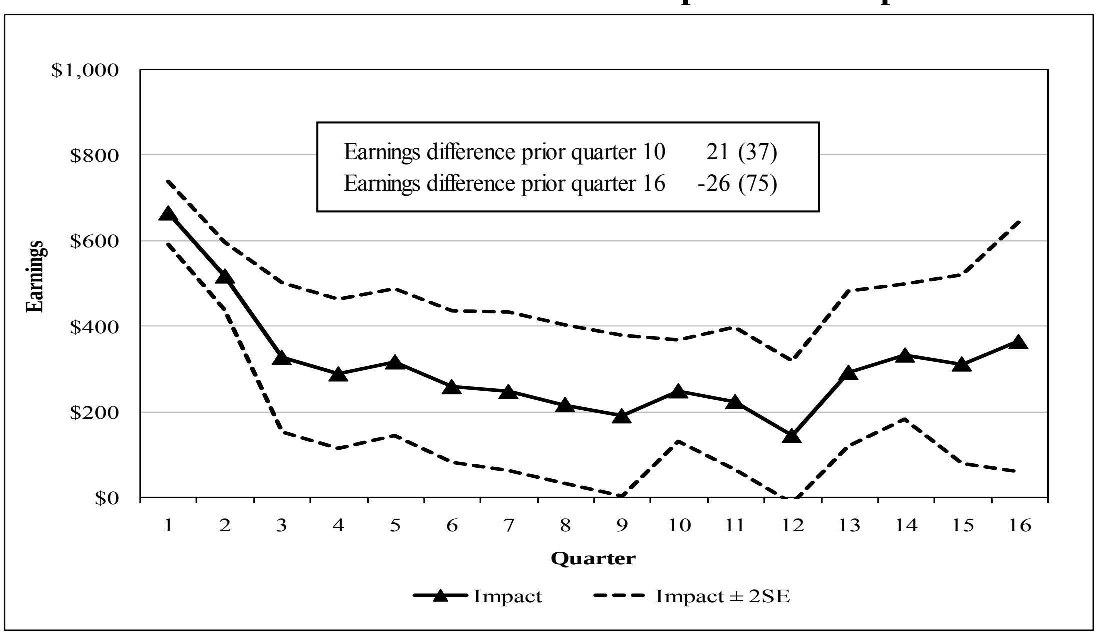


Figure V.10
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA Core/Intensive versus Comparison Group



results, 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. In this case, 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.

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

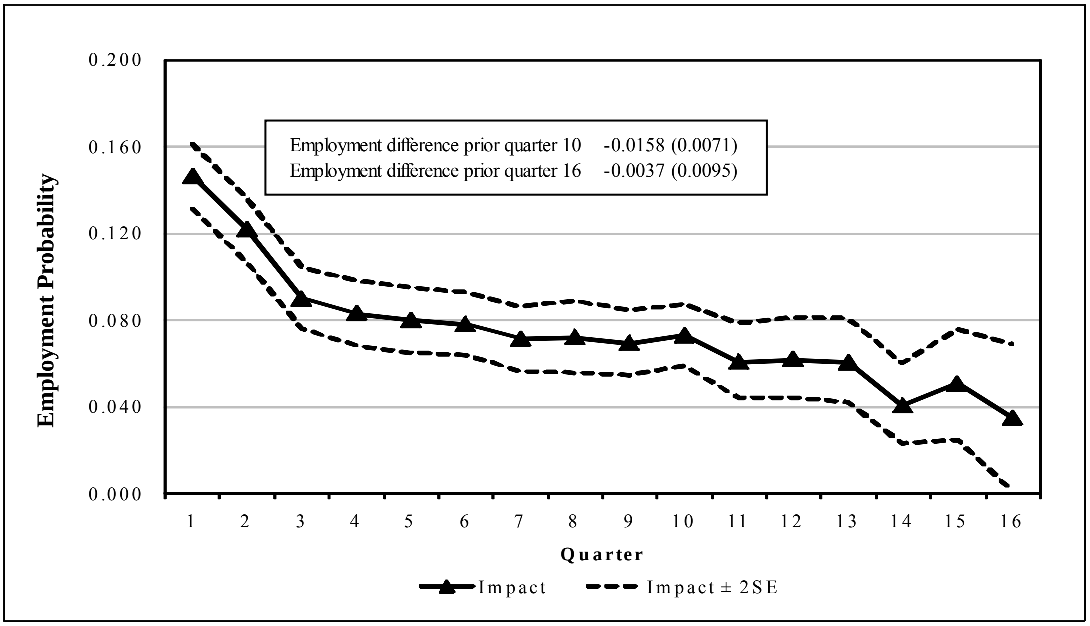
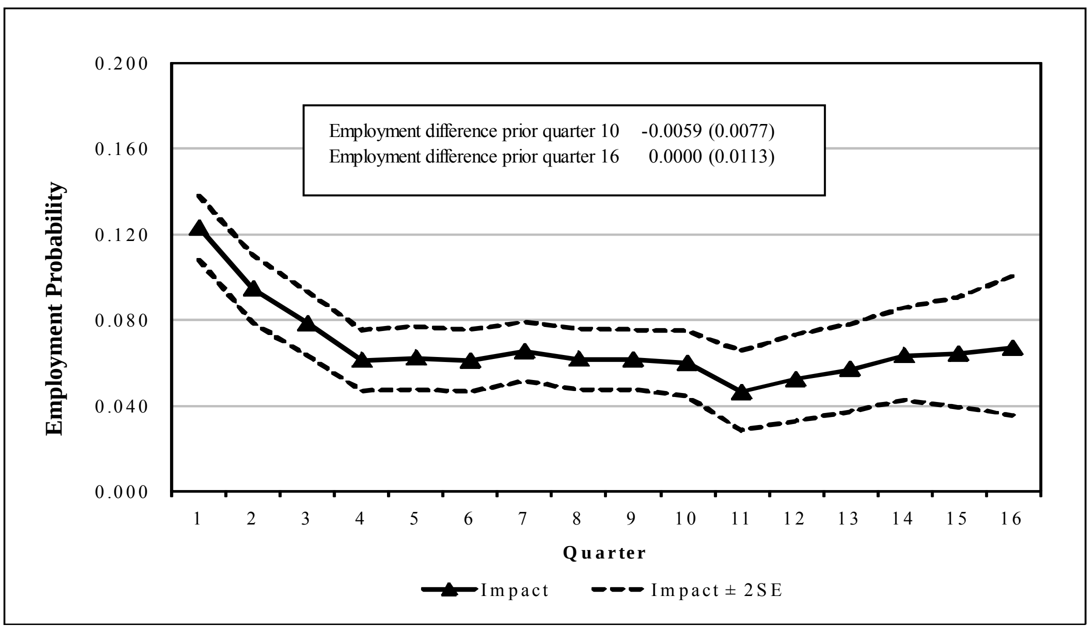


Figure V.12
Adult Program Treatment Effect on Quarterly Employment for Males, WIA Core/Intensive versus Comparison Group



The estimate of the "effect" of program participation for those who do not receive training on prior earnings 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 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.

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 to Adult participants may be sufficient to justify their fairly low costs.

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

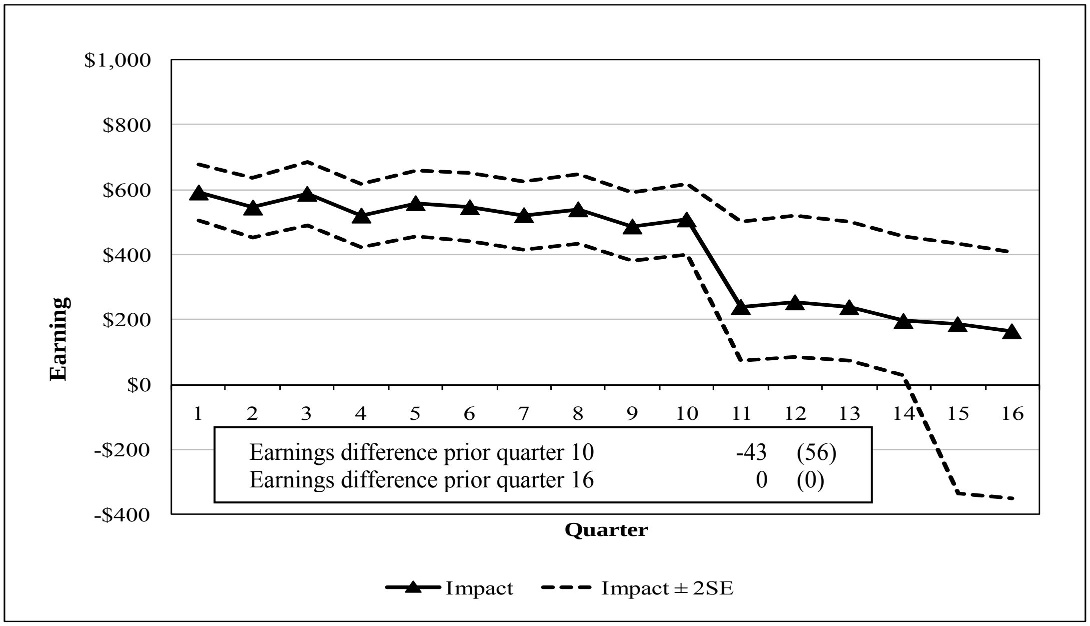
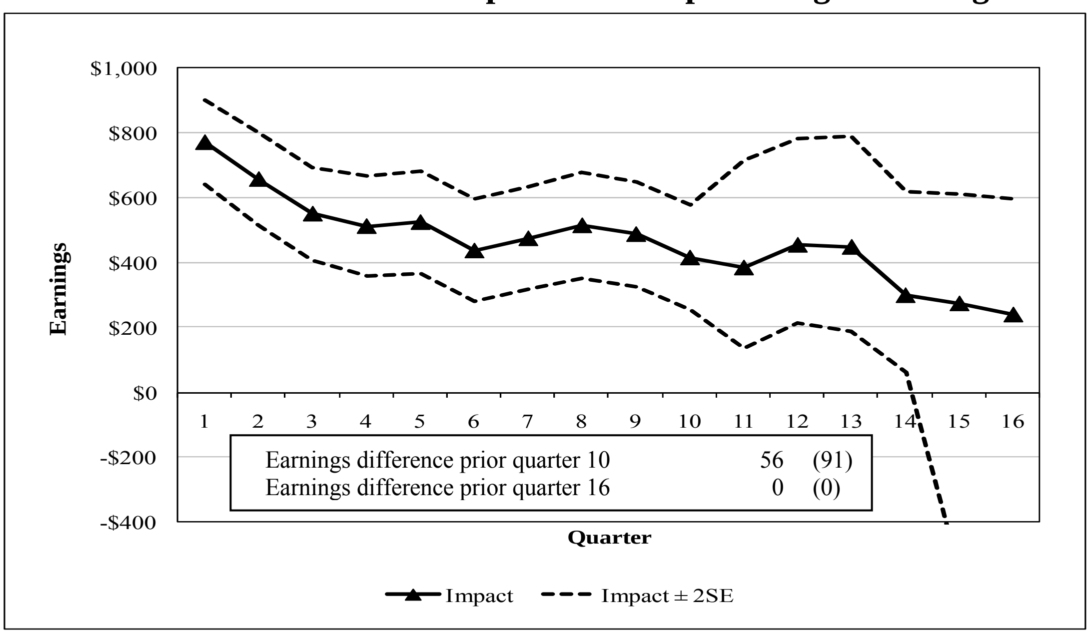


Figure V.14
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA Core/Intensive versus Comparison Group in 7 High Training States



3. Impacts of Training

The heart of WIA services is in the vocational skills training provided to individuals. Although a variety of training opportunities are widely available outside of WIA, some through the private sector and others subsidized at various levels by governmental entities, 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 importance, entry into the program would be expected to influence earnings or employment.[41]

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 higher earnings—in the range of $200 immediately after entry, with the increment remaining in the $500-600 range for the next 10 quarters.[42] 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 levels of employment are, however, more consistent with expectations. For females, employment is about 5 percentage points lower for those receiving training, and this 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 reported.)

For both males and females, there is some indication that selection on the basis of stable differences into training 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.

[41] 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. In this section, WIA participants who obtain training through the program are compared to WIA participants who do not receive WIA training. Some of this latter group will undoubtedly receive training through some other program.

[42] 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

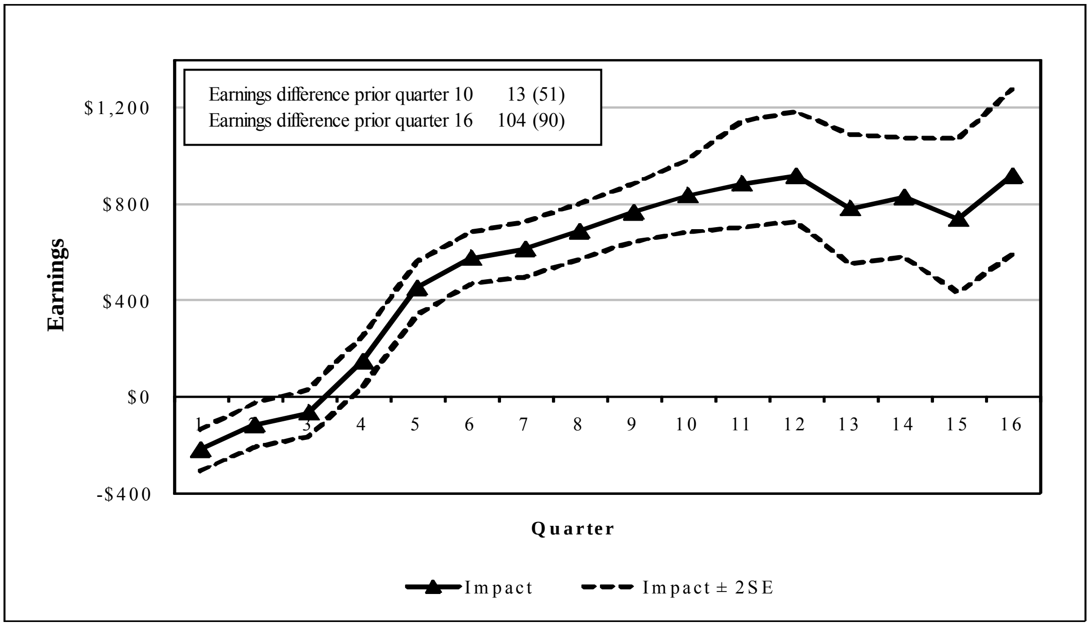


Figure V.16
Adult Program Treatment Effect on Quarterly Earnings for Males, WIA Training versus Comparison Group

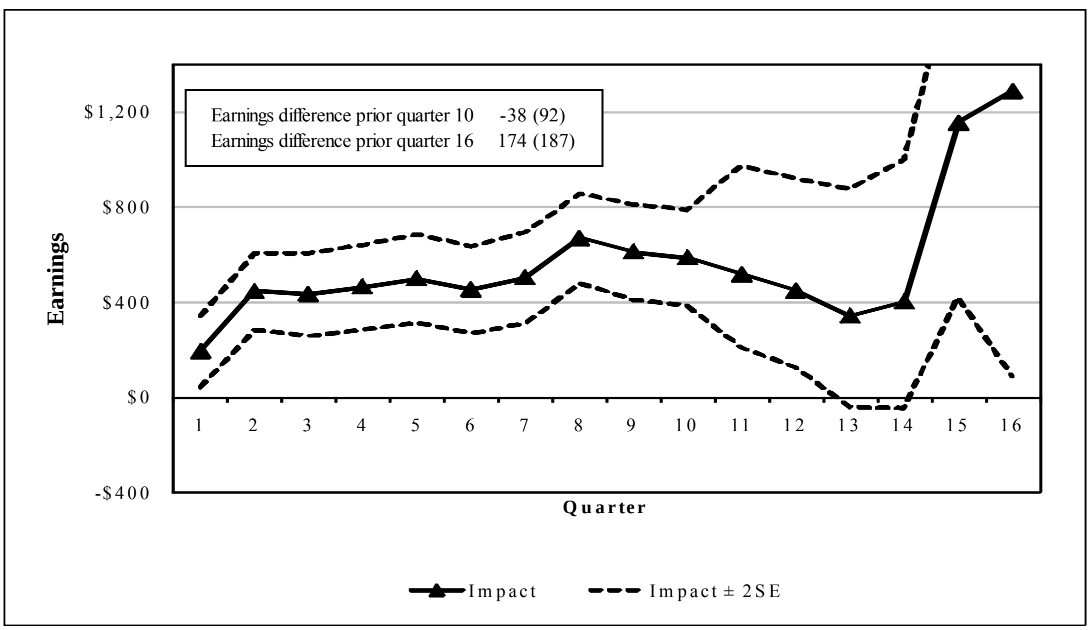


Figure V.17
Adult Program Treatment Effect on Quarterly Employment for Females, WIA Training versus Comparison Group

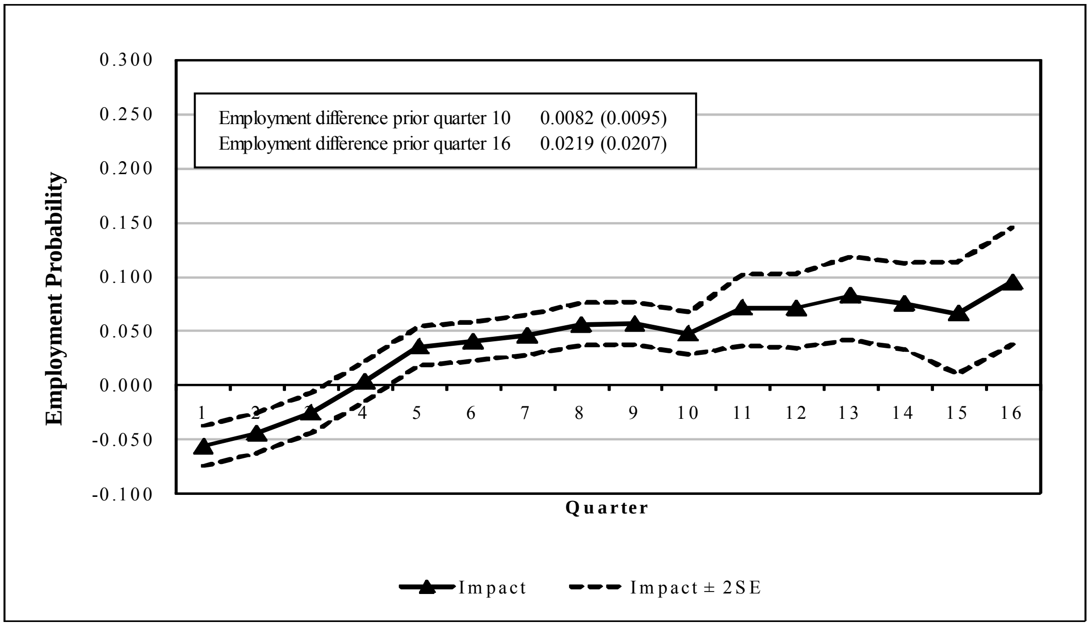
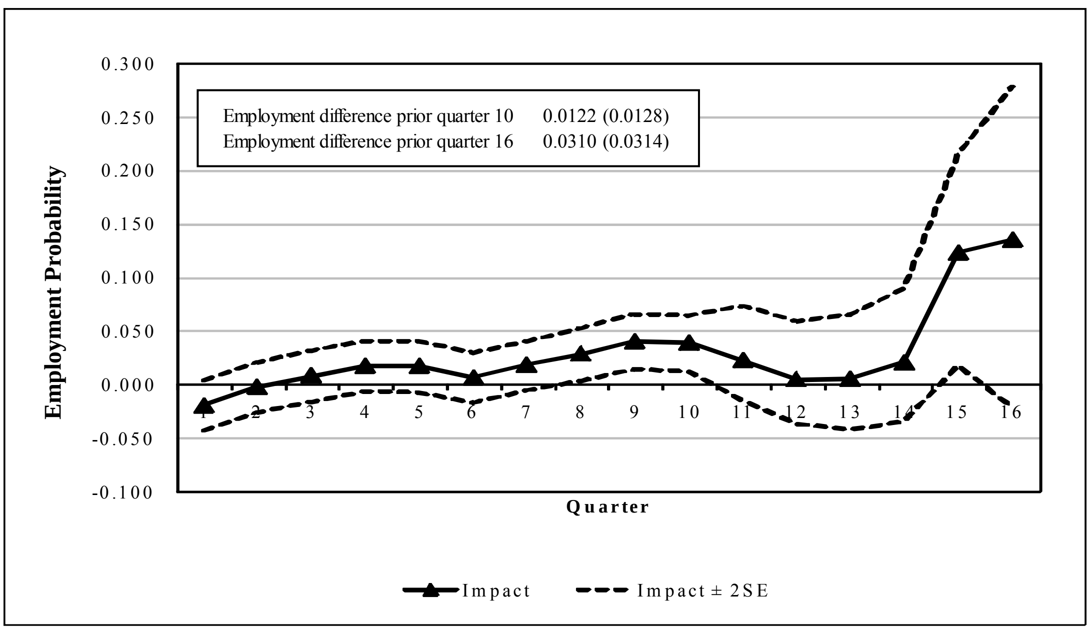


Figure V.18
Adult Program Treatment Effect on Quarterly Employment for Males, WIA Training versus Core/Intensive



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 benefits of over $400 per quarter. It is also important to point out, however that such an estimate is an average, and differences across states may be large.

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 participation 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.

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.

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, there is a positive impact—implying that individual participants in the Dislocated Worker program may be advantaged relative to the comparison group in ways not captured by control variables.[43] However, in examining individual states, sampling error is too large to allow this possibility to be investigated for individual states. It is therefore necessary to turn to estimates based on aggregating the state results.

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 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.

Figures VI.3 and VI.4 show that employment levels are not nearly so much lower initially for program participants. For women, 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.

[43] As noted above, a specification test for the basic model is provided by "predicting" earnings prior to program entry. 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.

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

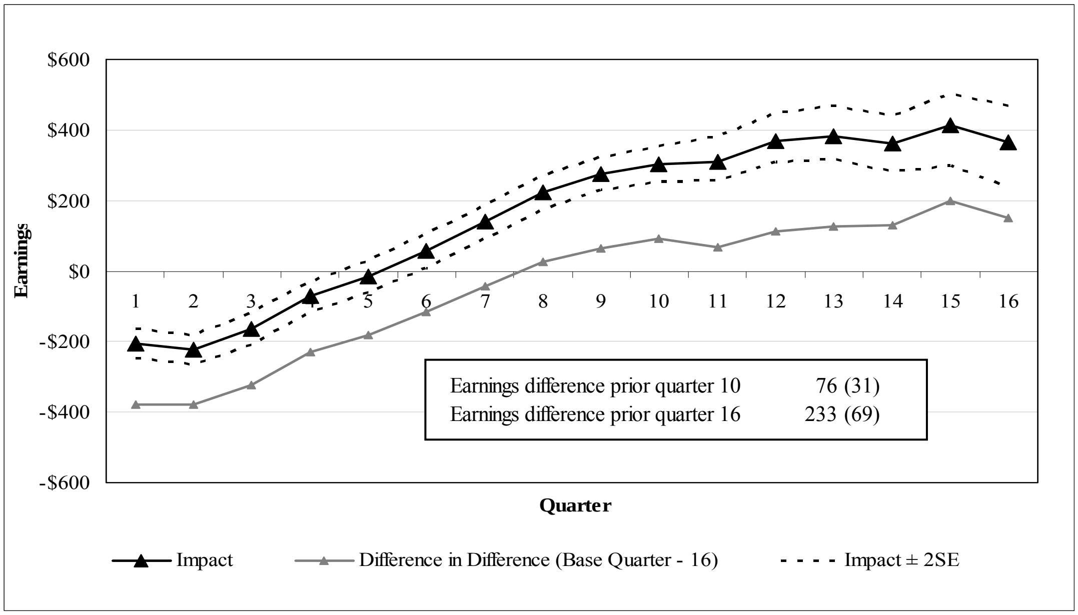


Figure VI.2
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA versus Comparison Group

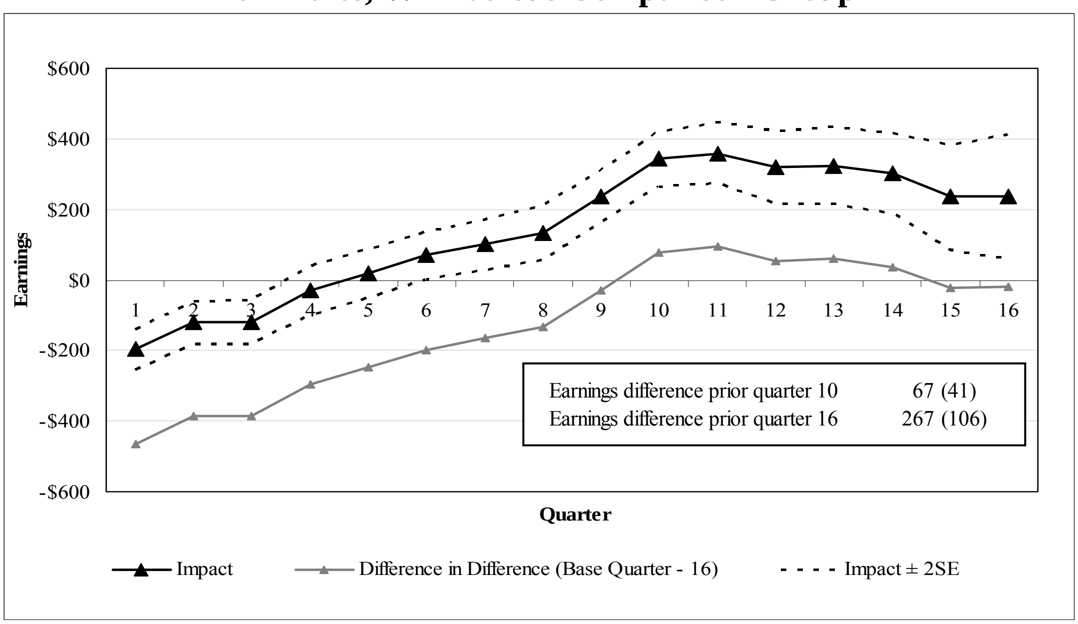


Figure VI.3
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA versus Comparison Group

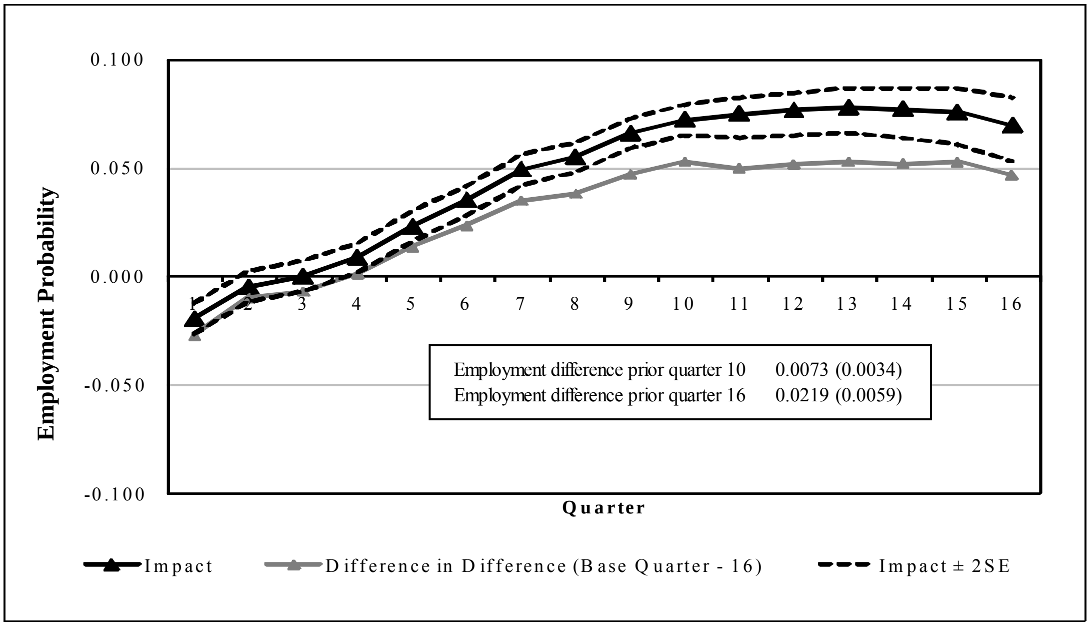
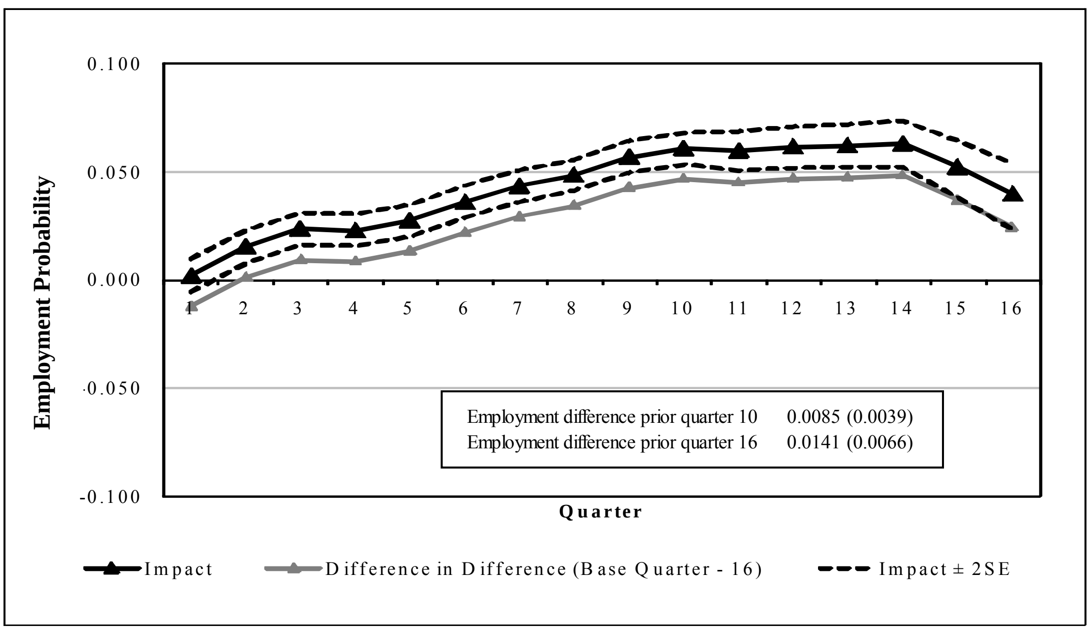


Figure VI.4
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA versus Comparison Group



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 participation, 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.

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. One way to test this assumption is to examine differences in earnings prior to program participation. Since controls are included for earnings in the eight quarters prior to program participation (matching assures there are no differences in that period), earnings are examined for 10 and 16 quarters prior to program entry. Any estimated "impact" of program participation on prior earnings identifies unmeasured differences; if prior earnings of treated individuals are greater, estimates of program impact for subsequent earnings are therefore suspect.

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.

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

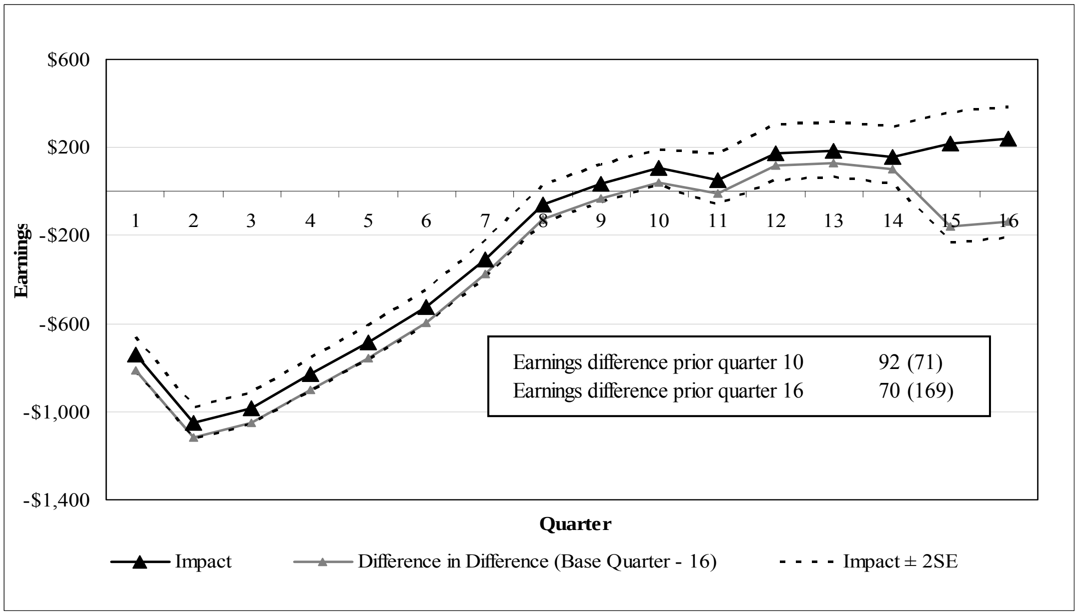


Figure VI.6
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA versus Comparison Group in 7 High Training States

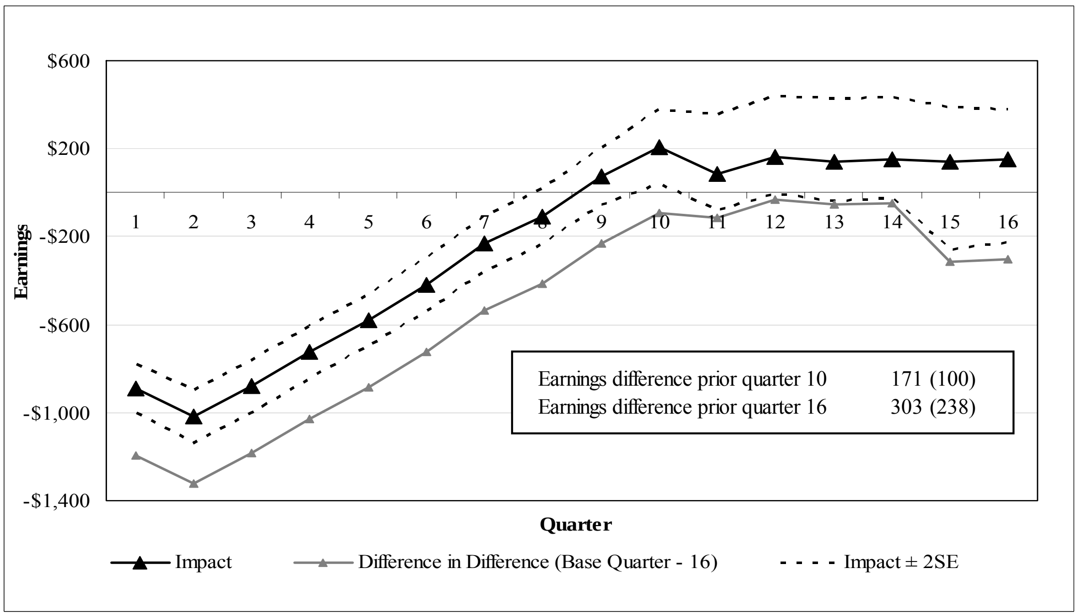


Figure VI.7
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA versus Comparison Group in 7 High Training States

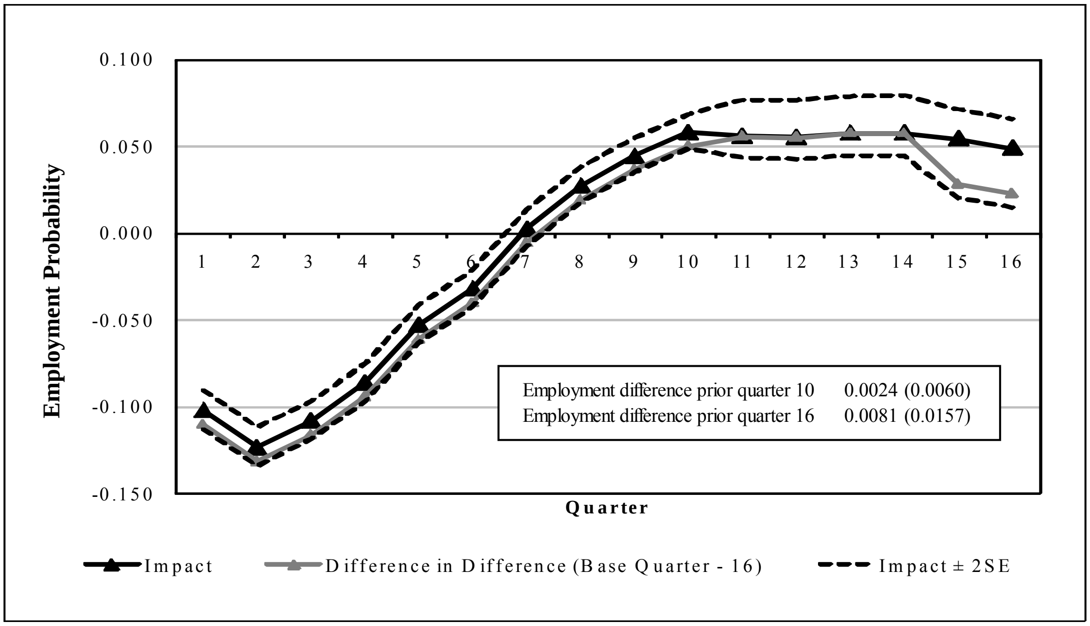
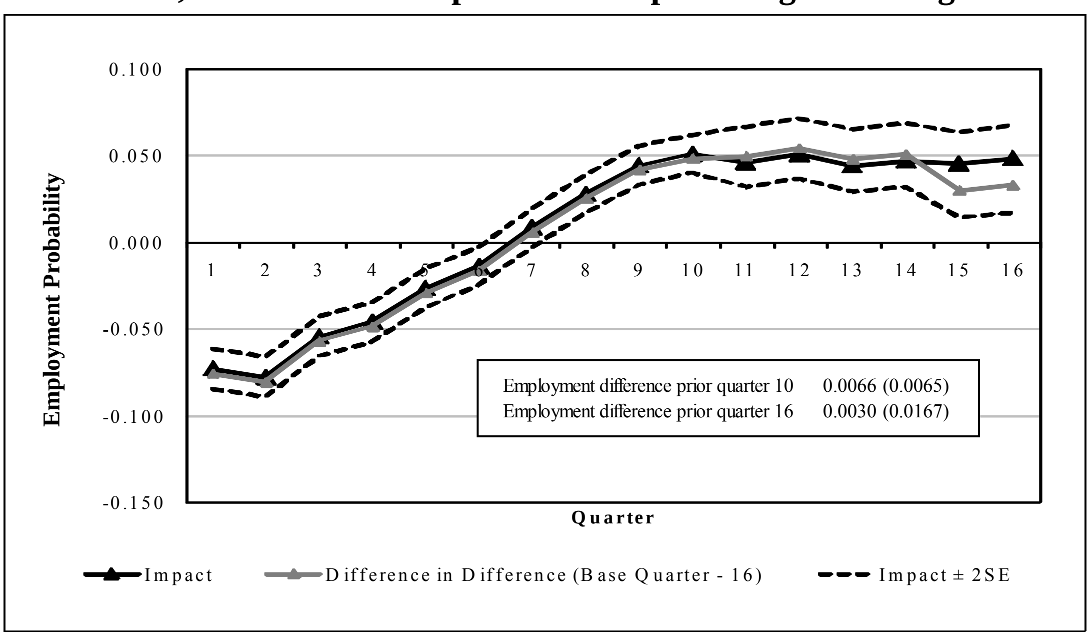


Figure VI.8
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA versus Comparison Group in 7 High Training States



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 (section V), the prior increment measures were of modest size, so the difference-in-difference estimates were not included in the figures. As here they appear to be substantial, these estimates are included in the figures in this section. These 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 others, and that these numbers increase over time.[44]

---

[44] 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

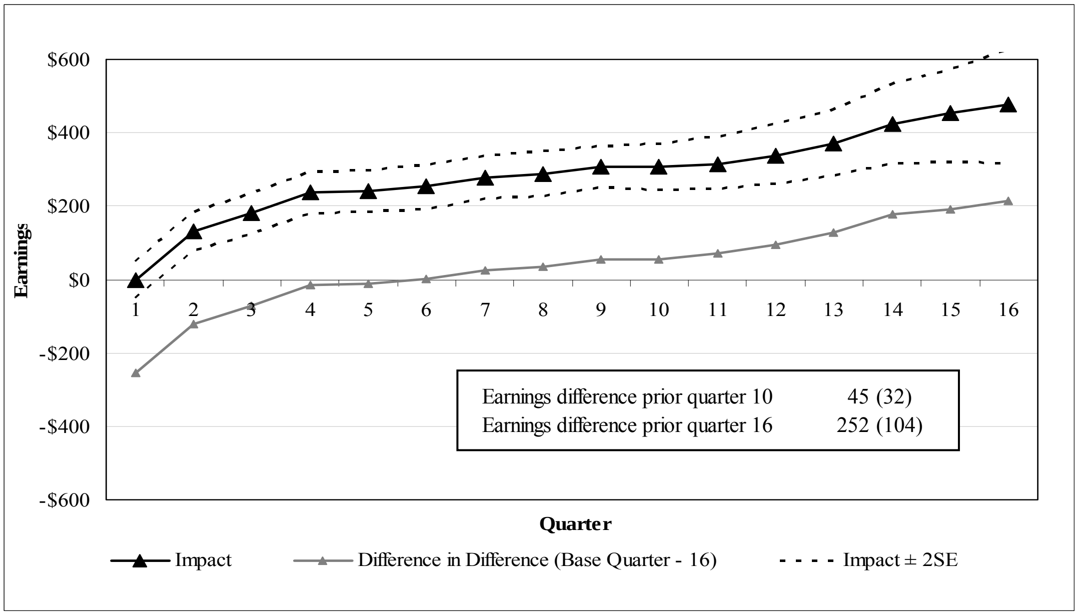


Figure VI.10
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA Core/Intensive versus Comparison Group

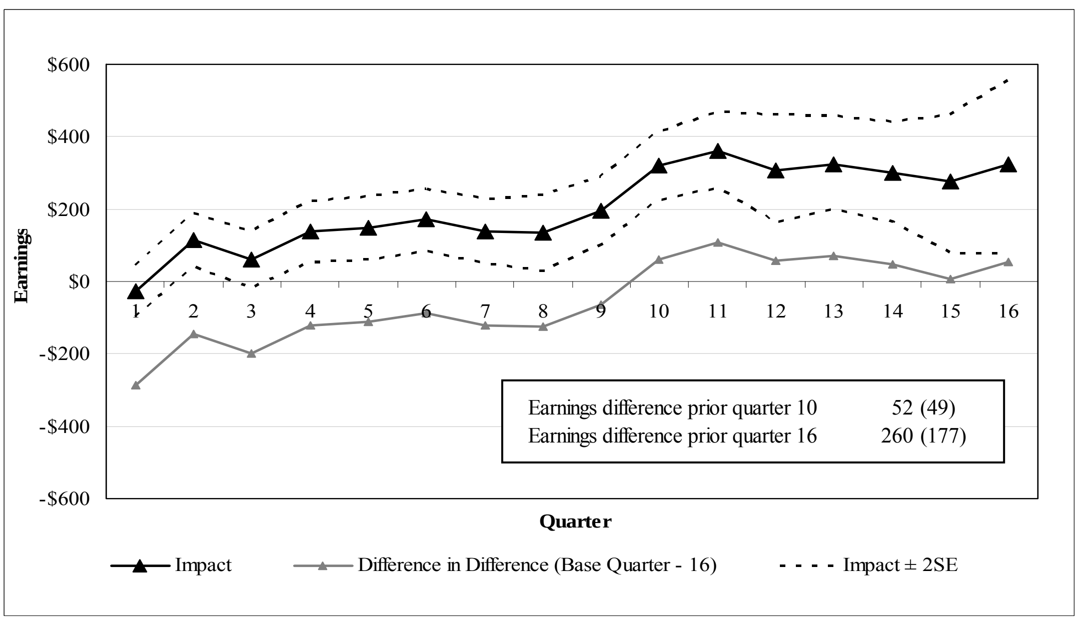


Figure VI.11
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA Core/Intensive versus Comparison Group

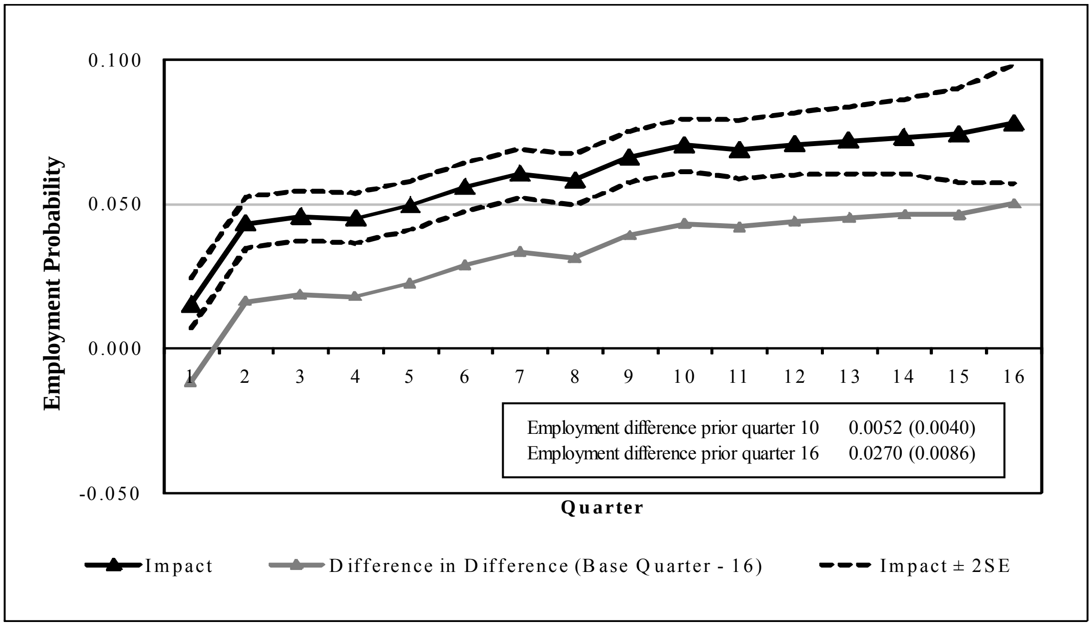
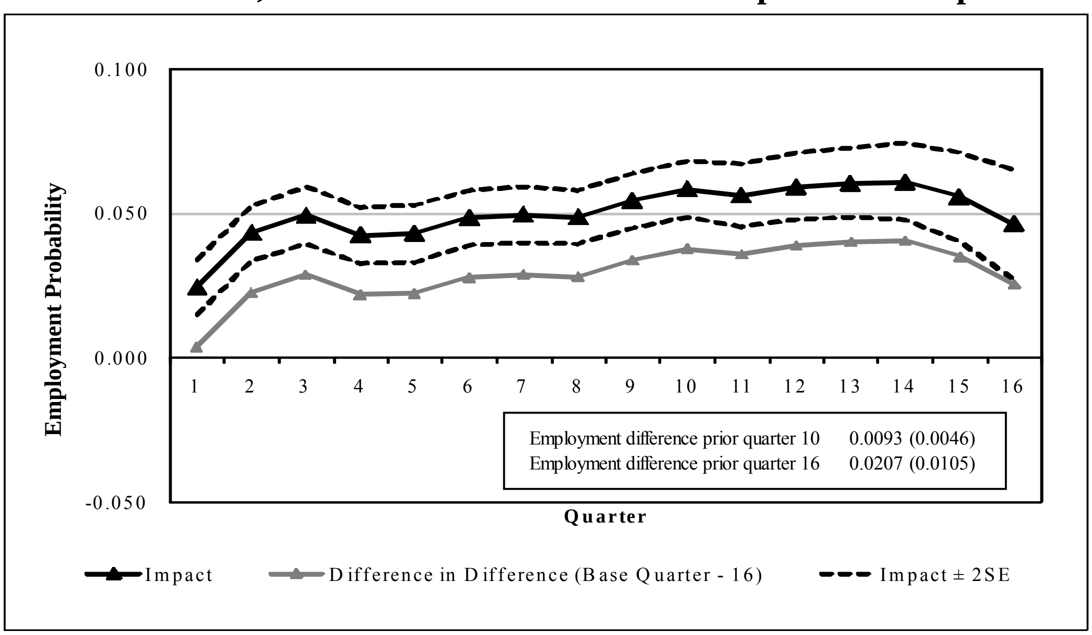


Figure VI.12
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA Core/Intensive versus Comparison Group



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. Intuitively, if participants have substantially greater earnings four years earlier, their subsequent earnings might well be higher as well, even if training has no effect. Difference-in-difference estimates provide an alternative measure of program impact. As noted above, such estimates will be valid if individuals are selected into the program on the basis of stable differences that reflect in earnings. 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.

3. Impacts of Training

The incremental impact of training is based on a comparison of WIA 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 than those who do not 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, or states where the comparison group is ES participants, 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

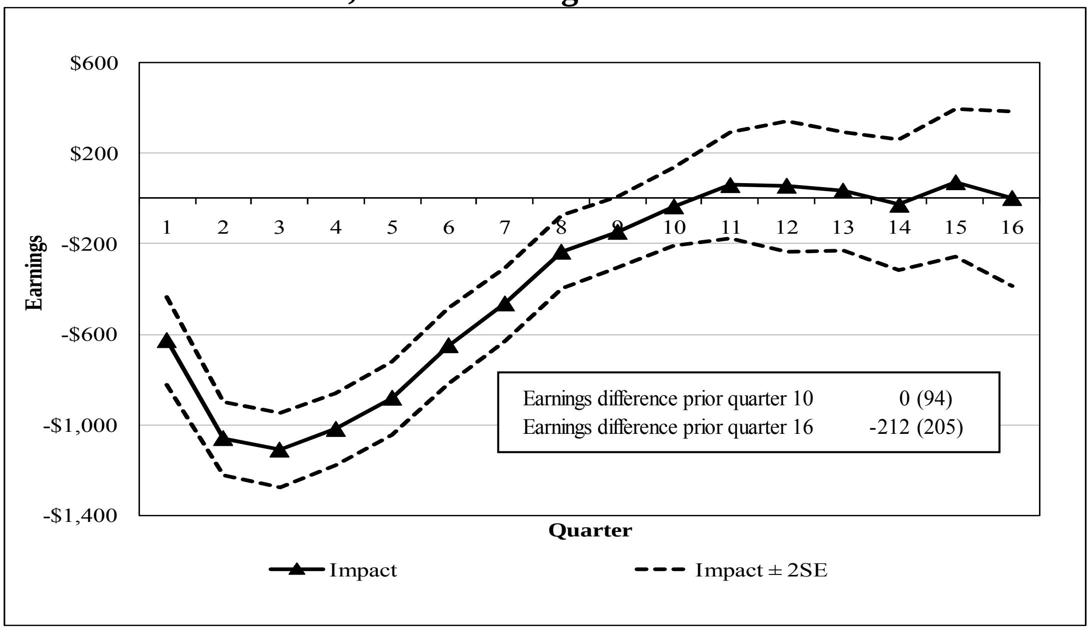


Figure VI.14
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, WIA Training versus Core/Intensive

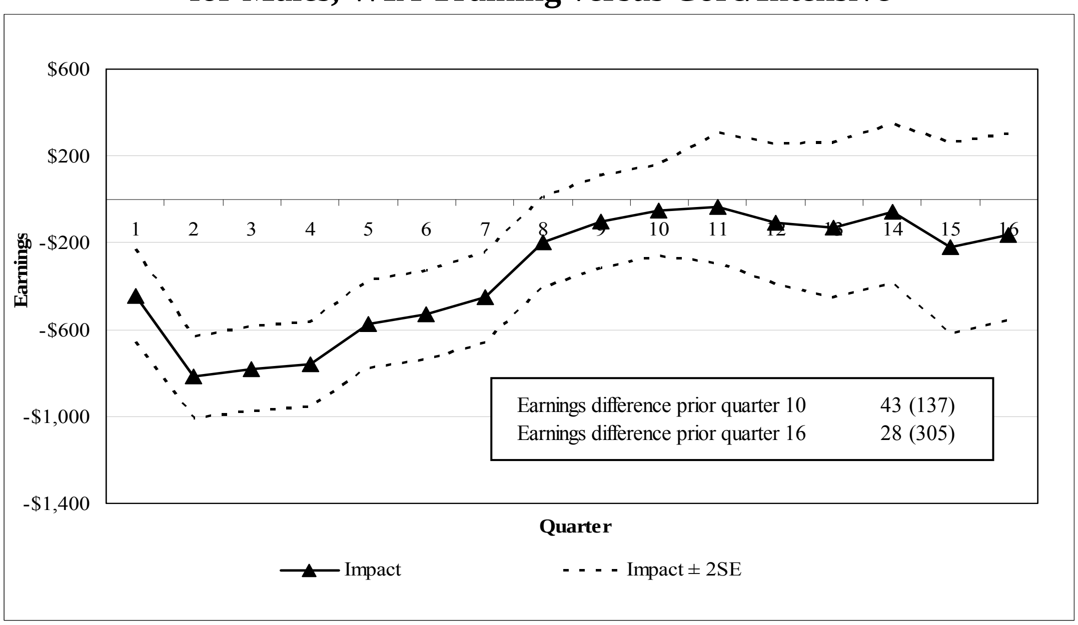


Figure VI.15
Dislocated Worker Program Treatment Effect on Quarterly Employment for Females, WIA Training versus Core/Intensive

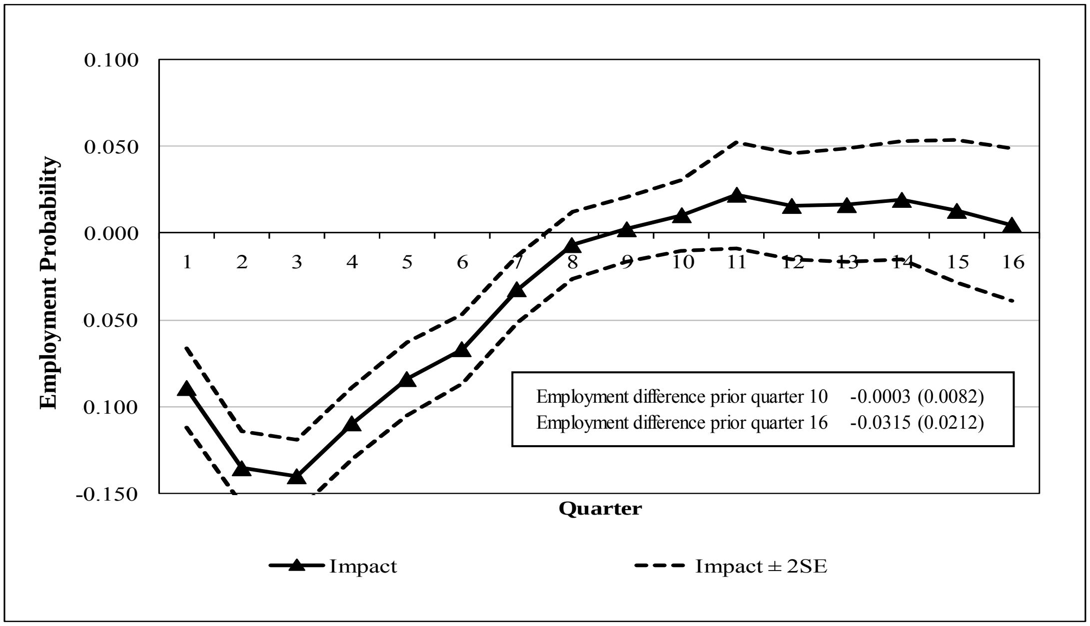
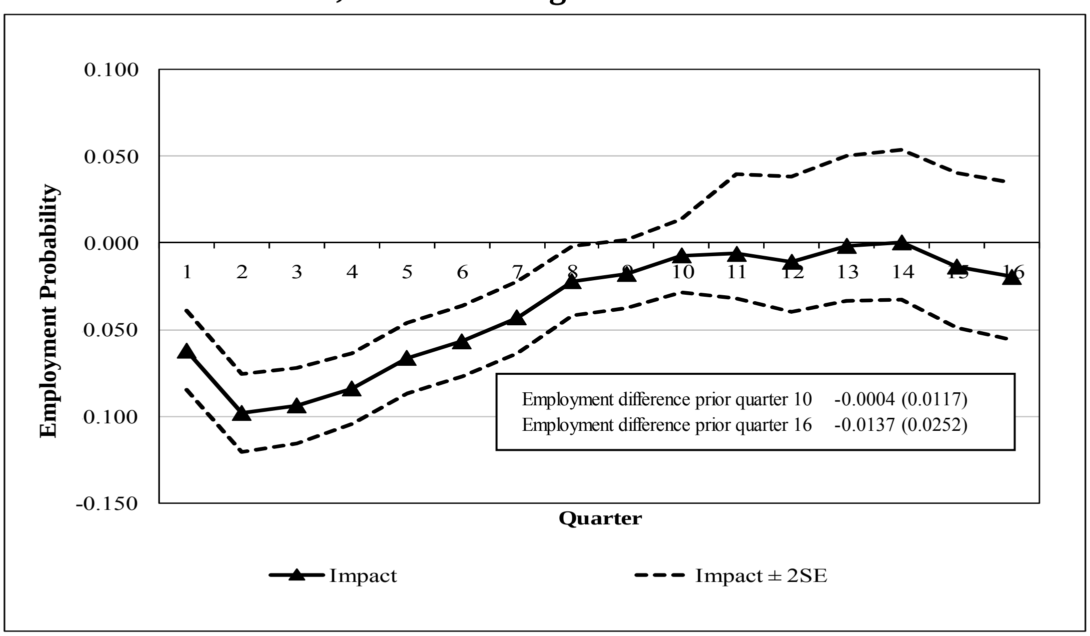


Figure VI.16
Dislocated Worker Program Treatment Effect on Quarterly Employment for Males, WIA Training versus Core/Intensive



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 all these results with caution. Whereas the results clearly imply that 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.

VII. WIA Program Impact Estimates for Subgroups

This section of the report considers results separately for various subgroups that are part of the larger population. The subgroups to be considered are: nonwhite non-Hispanics; Hispanics; individuals under 26 years of age at the time they entered the program; individuals 50 years of age or older at the time they entered the program; and veterans. There are several reasons for the focus on these subgroups. First, members of some of these subgroups, such as nonwhite minorities, tend to make up a larger share of participants in WIA than their share in the overall population. Second, it is believed that many of these groups face special challenges or barriers to working in the labor market, which may affect the impact of any training they receive. Previous research has shown that some of these groups have lower returns to education and training. For these reasons, analyzing these groups separately provides a more informative picture of the efficacy of the WIA program.

The procedure used in estimating the effects for each subgroup exactly replicates the procedures for estimating effects for the entire population that are described above.[45] The only difference is that the analysis is limited to individuals in the treatment and comparison samples who are members of the subgroup. Separate estimates for men and women within each subgroup are also provided, with the exception of veterans since there are simply too few women to generate reasonable estimates. Therefore, the analysis of veterans focuses exclusively on male veterans.[46] Separate estimates are provided for participants in the Adult program and participants in the Dislocated Worker program.

The discussion of results focus on the estimated effects on earnings and on estimated effects for all WIA participants in a given program—whether or not they received Core, Intensive or Training services. Additional results relating to the subgroup analysis are provided in Appendix 4. For the most part, the patterns of estimates for the subgroups mirror the results found for all participants. However, given the large standard errors associated with these estimates for the subgroups, it is not possible to determine with sufficient precision whether the size of the effects differs for the subgroups.

1. Nonwhites

The results for the nonwhite, non-Hispanic subgroup are shown in Figures VII.1-VII.4. This group will be referred to as nonwhites; however, it is important to keep in mind that nonwhite Hispanics are included in the Hispanic subgroup. The nonwhite non-Hispanic subgroup consists primarily of black participants, although this group does include a few Native Americans, Asians and Pacific Islanders.

Figures VII.1 and VII.2 present results for nonwhite female and male participants in the Adult program. Comparing Figure VII.1 to Figure V.1, which presents impact estimates for all female participants in the Adult program, it is clear that estimated impacts for female nonwhite participants largely mirror those for all females. Both figures show that estimated program effects are around $500 in the first quarter after starting the program, although the nonwhite effects decline somewhat over 10 quarters following program entry.

[45] Since the subgroups are smaller, the number of cases with information for a full 16 quarters of data was not sufficient to support estimates. Subgroups estimates are therefore limited to a follow-up period of 10 quarters (30 months) following program entry.

[46] Since the analysis is performed separately by state, the results from any state that includes too few members of a group to generate meaningful estimates are omitted.

Figure VII.1
Adult Program Treatment Effect on Quarterly Earnings for Non-White Females, WIA versus Comparison Group

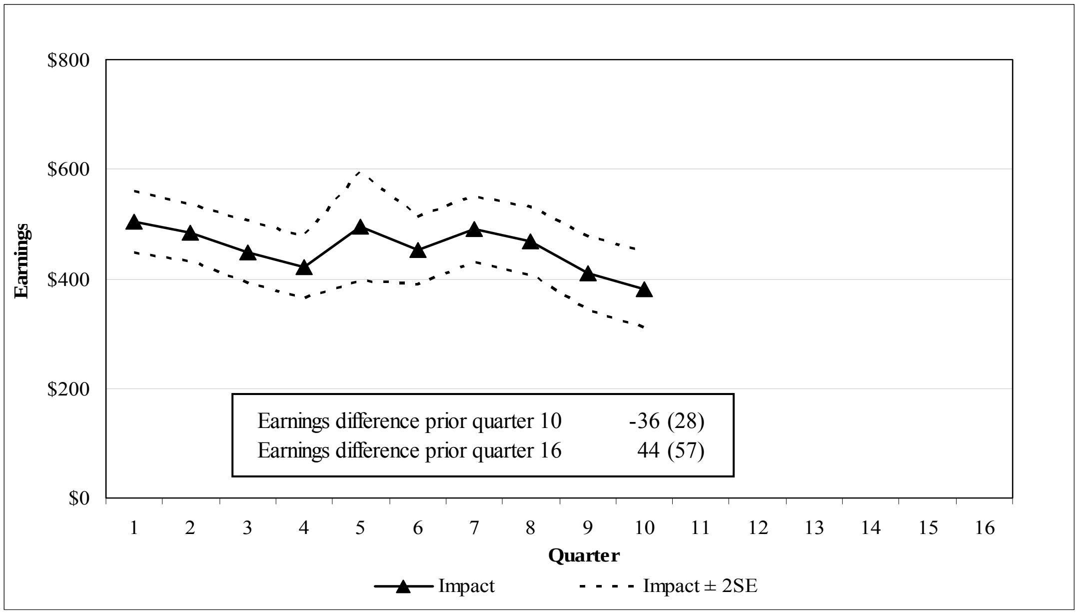
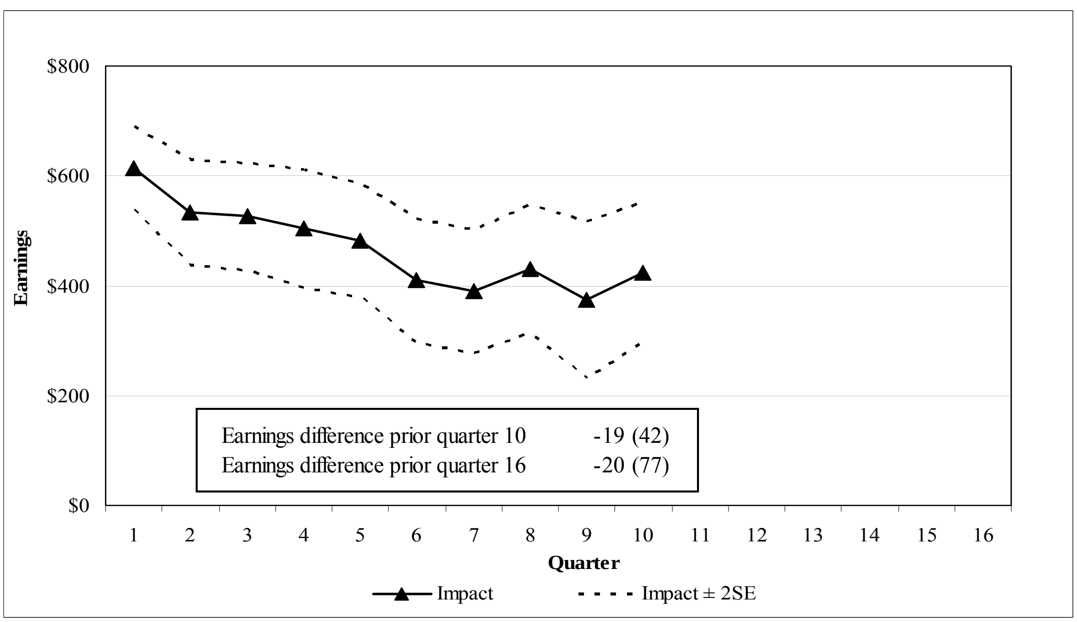


Figure VII.2
Adult Program Treatment Effect on Quarterly Earnings for Non-White Males, WIA versus Comparison Group



Comparing Figure VII.2 to Figure V.2, which presents impact estimates for the full population of male participants in the Adult program, it is clear that the estimated effect for nonwhite males is initially similar to the estimated effect for all males but does not decline quite as much as the impact for all male participants in the Adult program. Based on these results, it appears that male nonwhite participants in the Adult program face slightly higher estimated impacts than all male participants, although, given the large standard errors associated with these estimates, most differences are not statistically significant.

Figure VII.3 presents estimated effects for nonwhite females in the Dislocated Worker program while Figure VII.4 presents results for nonwhite males. Comparing these figures to the comparable figures for all participants (Figures VI.1 and VI.2) shows that both nonwhite male and female Dislocated Workers experience impacts that increase at a faster rate in the quarters immediately after entering the program. Impacts reach a peak at a similar level as for all participants and tend to remain around this peak for the time period under consideration.

2. Hispanics

Results for Hispanics are presented in Figures VII.5-VII.8. Female Hispanic participants have average estimated effects that are fairly large (over $700), with the impacts exhibiting a slight increasing trend for ten quarters after entering the program (Figure VII.5). For male Hispanics, the estimated effect is even larger immediately after entering the program (over $1,000) but declines fairly rapidly before leveling off around three to four quarters after entering the program (Figure VII.6). Comparing these figures to the comparable figures for all participants (Figures V.1 and V.2) shows that the estimated impact is larger than the estimate for all participants, although the large standard errors associated with the Hispanic estimates make it clear that the differences are not statistically significant.

Figure VII.3
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Non-White Females, WIA versus Comparison Group

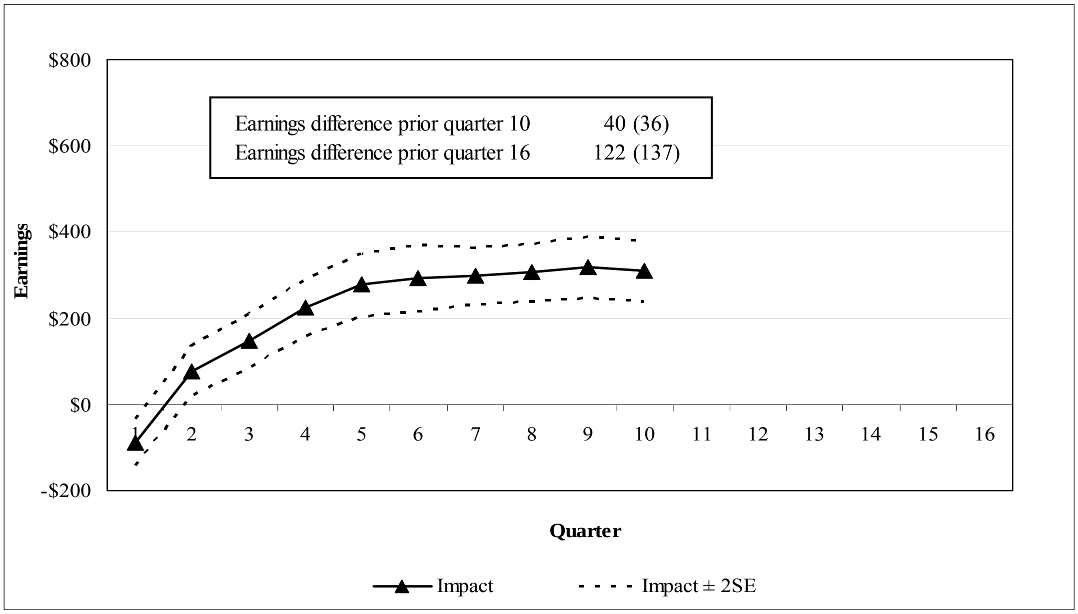


Figure VII.4
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Non-White Males, WIA versus Comparison Group

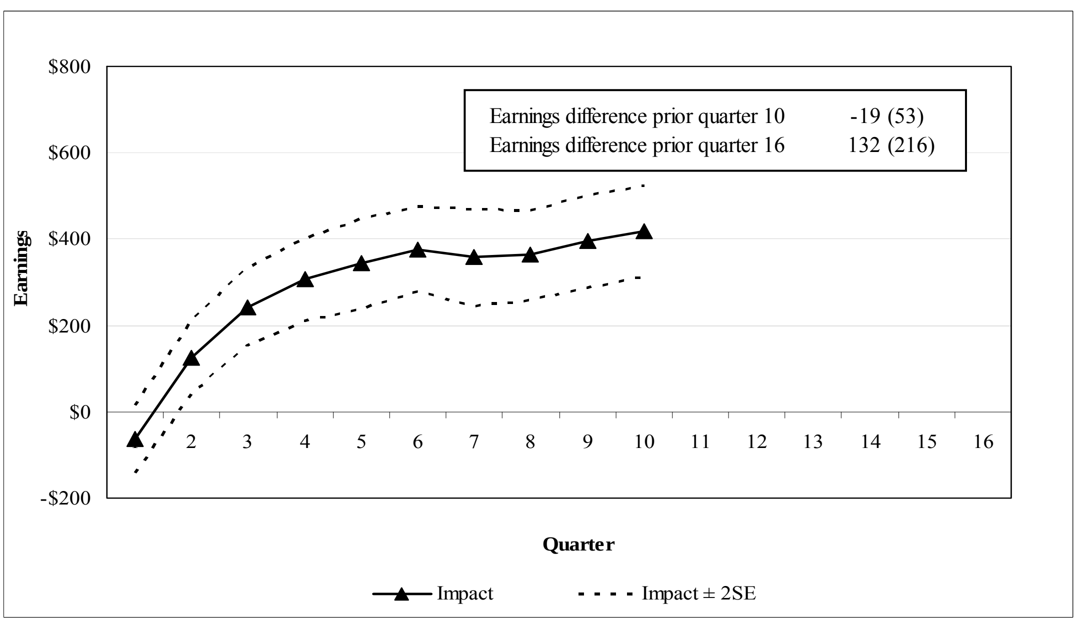


Figure VII.5
Adult Program Treatment Effect on Quarterly Earnings for Hispanic Females, WIA versus Comparison Group

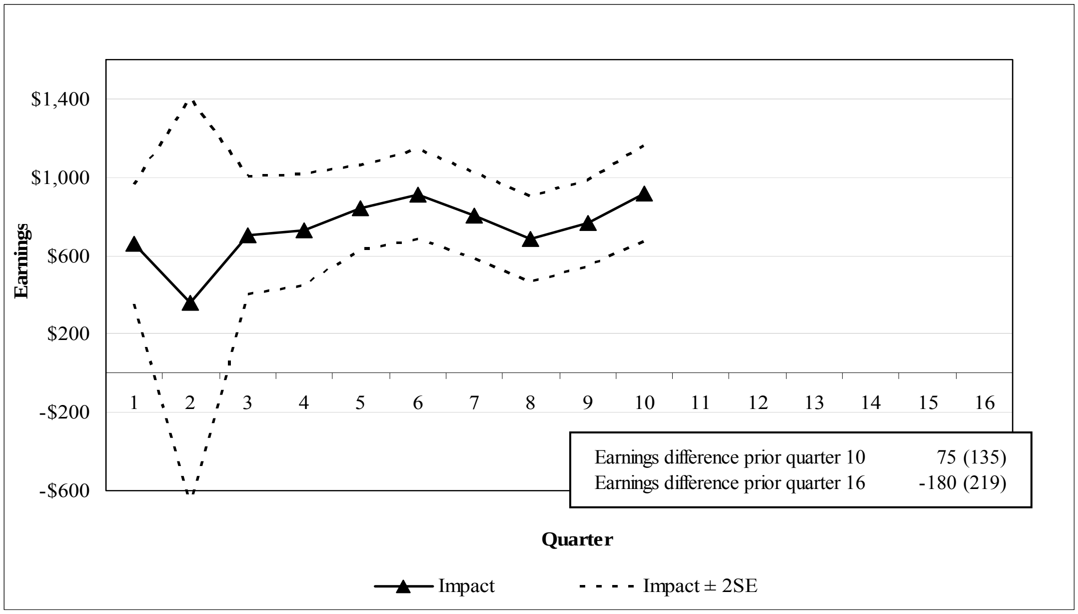
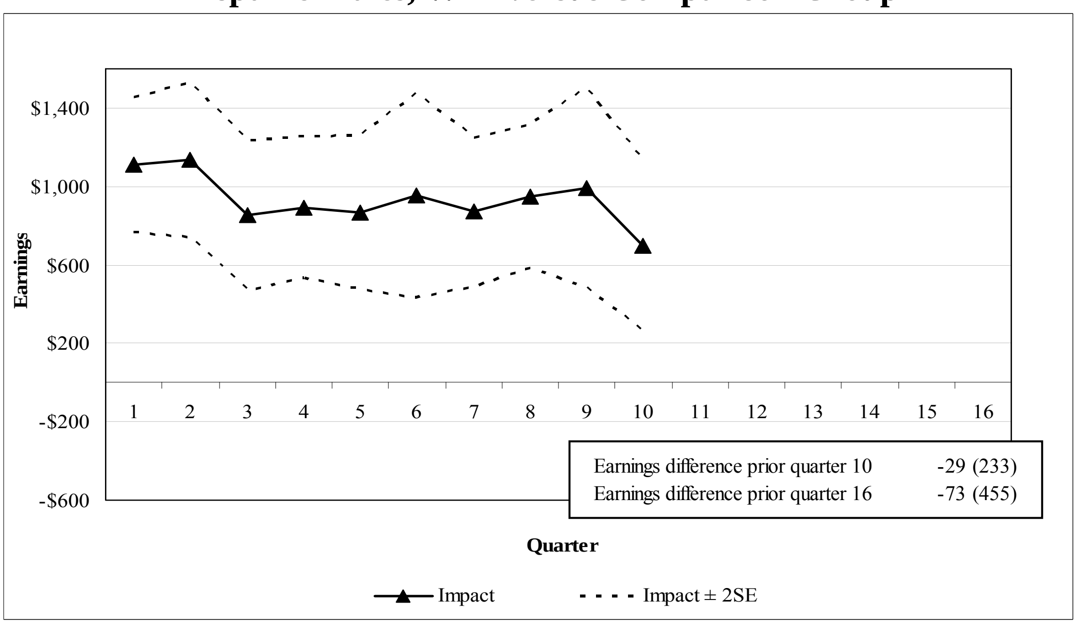


Figure VII.6
Adult Program Treatment Effect on Quarterly Earnings for Hispanic Males, WIA versus Comparison Group



Figures VII.7 and VII.8 present estimates for female and male Hispanic participants in the Dislocated Worker program, respectively. These figures display similar patterns to the figures for all participants in the Dislocated Worker program (Figures VI.1 and VI.2), although in the first two or three quarters Hispanic earnings appear to be substantially lower. Again, sampling error is large, so that most differences are not statistically significant.

3. Participants under Age 26

This analysis focuses on estimates for young participants—that is, participants who enter the program prior to their twenty-sixth birthday—in the Adult program. The estimates for young female participants in the Adult program are presented in Figure VII.9 and the estimates for young male participants in the Adult program are presented in Figure VII.10. Focusing first on the results for young female participants, the estimated effects average around $500, although there is growth through quarter eight followed by a decline. Allowing for sampling error, this is essentially the same as the estimated effects for all female participants in the Adult program. The estimates for young males correspond quite closely as well to those for the population of all males.

The estimated impacts for young female and male participants in the Dislocated Worker program, seen in Figures VII.11 and VII.12, parallel the pattern of estimated effects for the typical female and male participant in the Dislocated Worker program, but they are appreciably higher. Prior earnings (16 quarters prior to entry) are about $200 higher for participants than the matched comparison group, suggesting that the outcome difference may reflect differences on stable unmeasured personal attributes. However, the large sampling error underscores the difficulty of identifying differences with any certainty.

4. Participants Age 50 and Over

Figures VII.13-VII.16 present results for older female and male participants in the Adult and Dislocated Worker programs, respectively. Older participants refer to participants who were 50 years old or older when they entered the program. While there appear to be some minor differences in the patterns of effects over time compared to the effects for the typical female and male participants in these programs, none of the differences are statistically significant, due to the large standard errors associated with the estimates for older workers. The most reasonable conclusion is that the estimated impact of these programs on older participants is largely consistent with the estimated effects observed for all participants.

Figure VII.7
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Hispanic Females, WIA versus Comparison Group

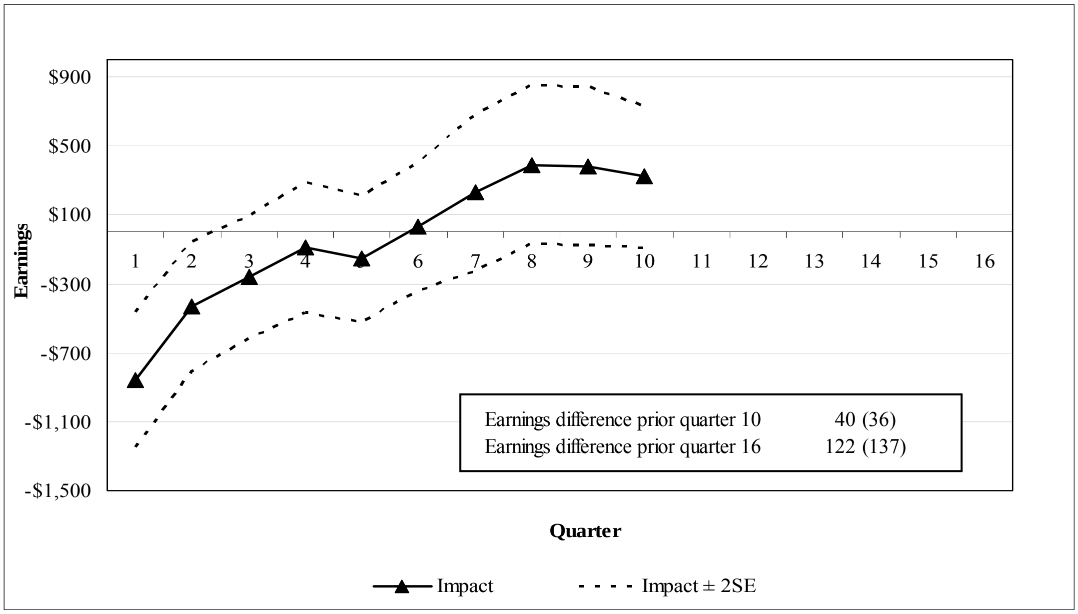


Figure VII.8
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Hispanic Males, WIA versus Comparison Group

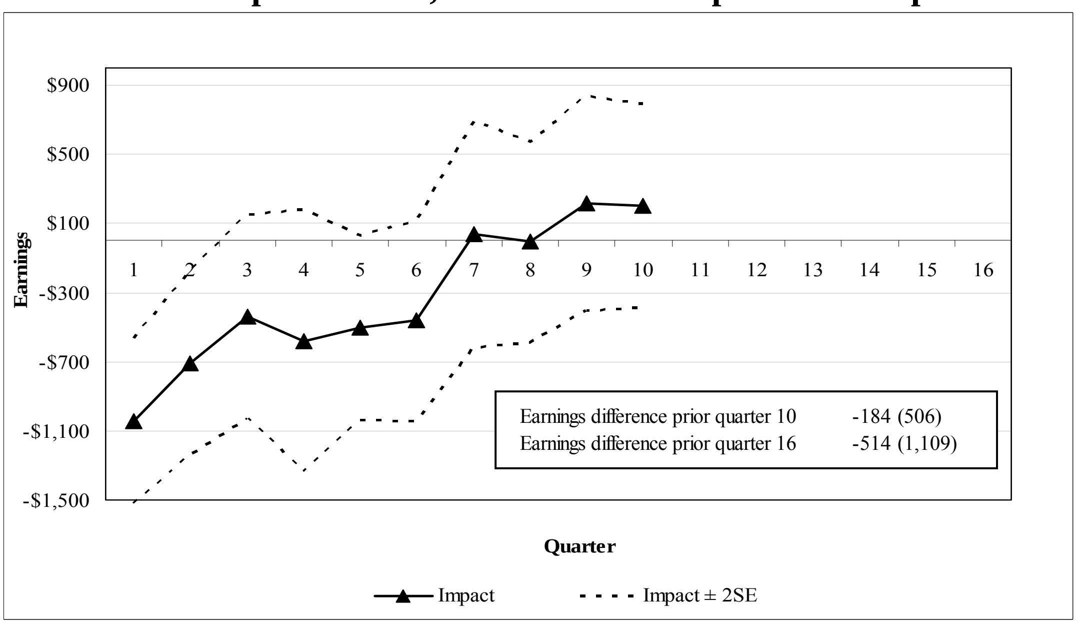


Figure VII.9
Adult Program Treatment Effect on Quarterly Earnings for Females under Age 26, WIA versus Comparison Group

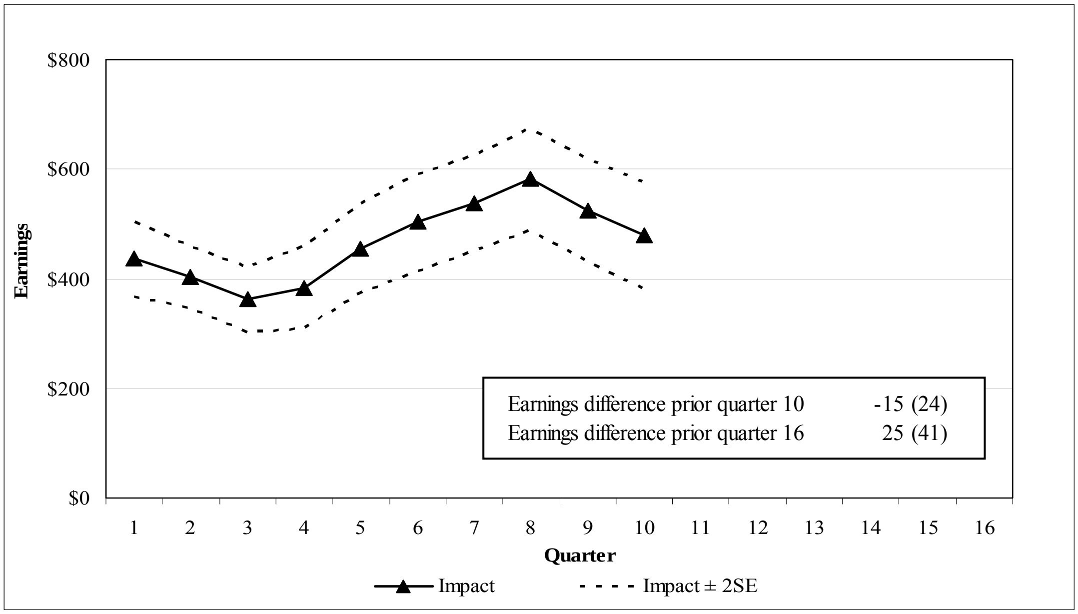


Figure VII.10
Adult Program Treatment Effect on Quarterly Earnings for Males under Age 26, WIA versus Comparison Group

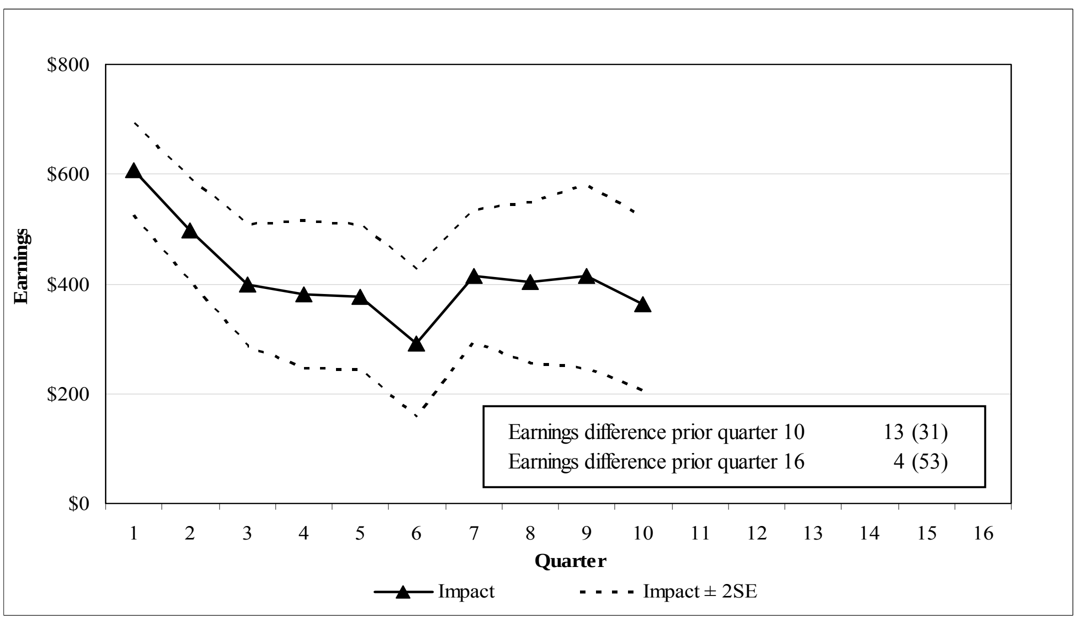


Figure VII.11
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Females under Age 26, WIA versus Comparison Group

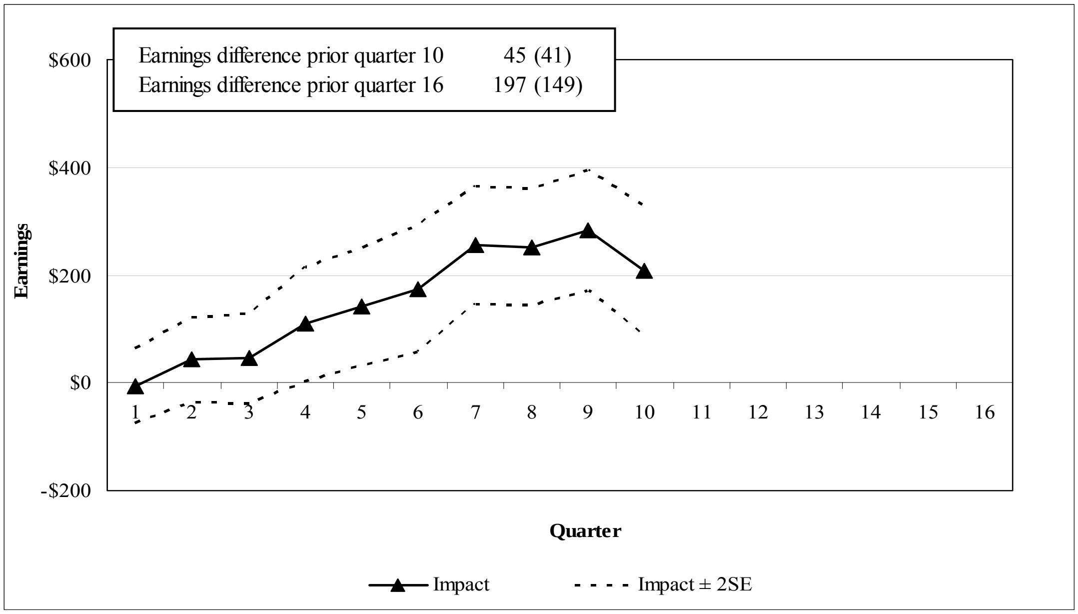


Figure VII.12
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males under Age 26, WIA versus Comparison Group

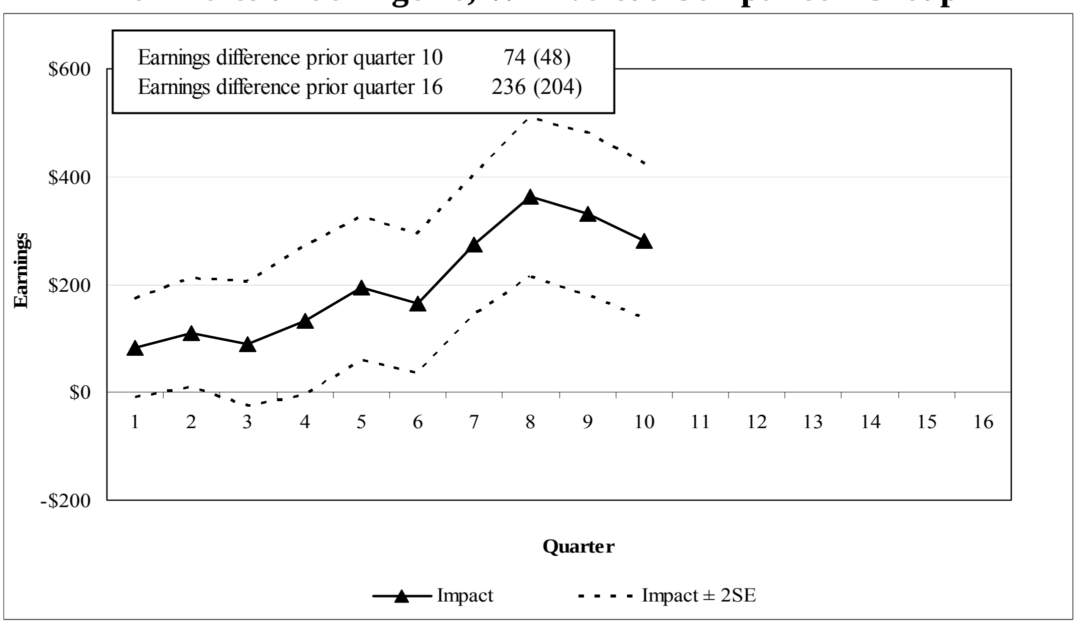


Figure VII.13
Adult Program Treatment Effect on Quarterly Earnings for Females Age 50+, WIA versus Comparison Group

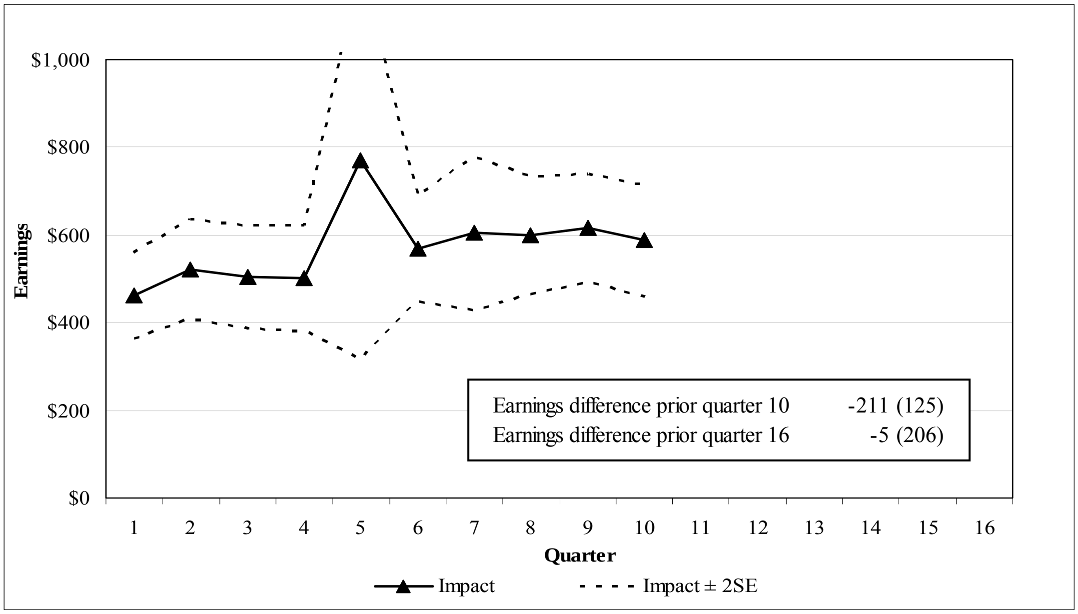


Figure VII.14
Adult Program Treatment Effect on Quarterly Earnings for Males Age 50+, WIA versus Comparison Group

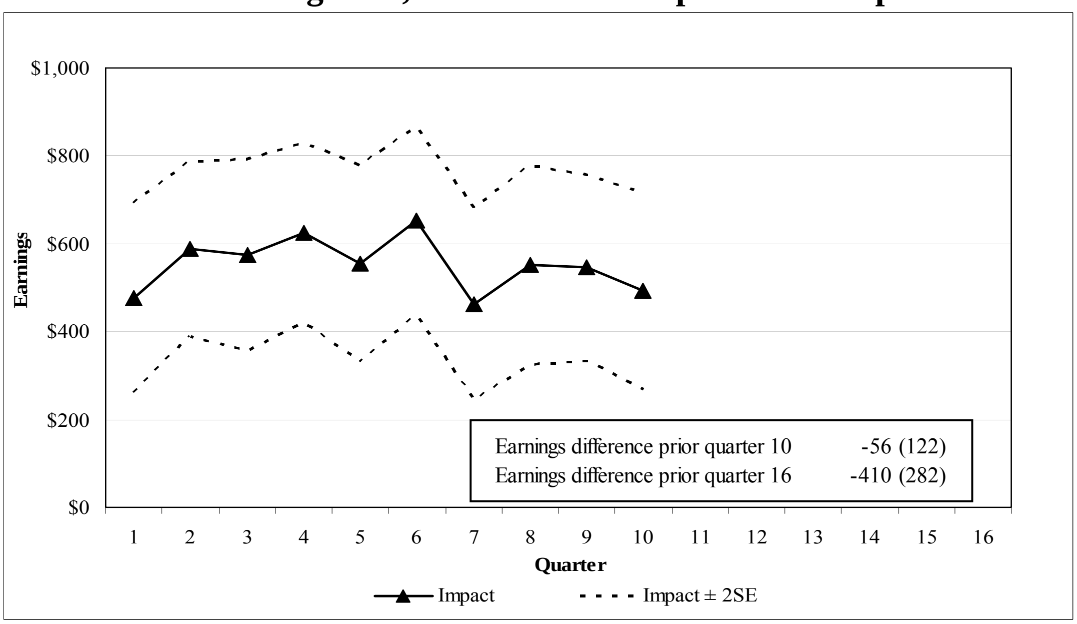


Figure VII.15
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Females Age 50+, WIA versus Comparison Group

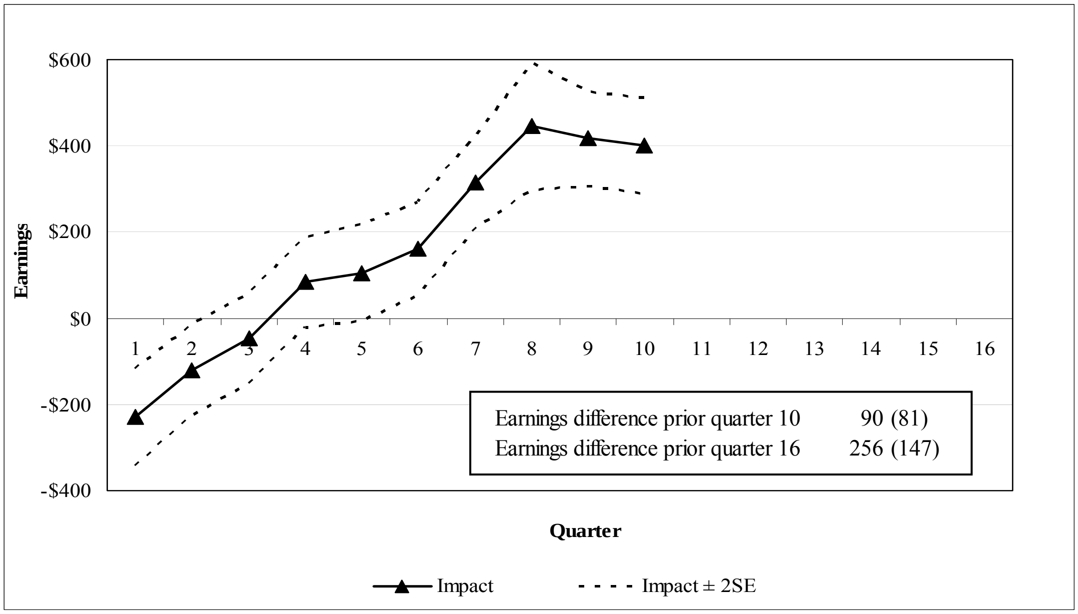


Figure VII.16
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Males, Age 50+, WIA versus Comparison Group

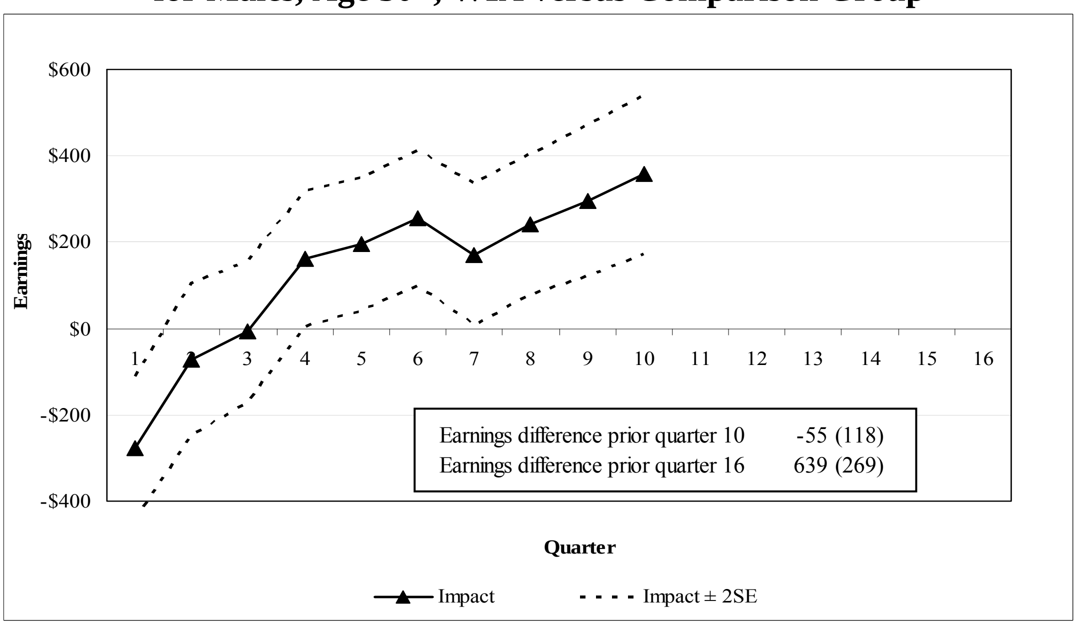


Figure VII.17
Adult Program Treatment Effect on Quarterly Earnings for Male Veterans, WIA versus Comparison Group

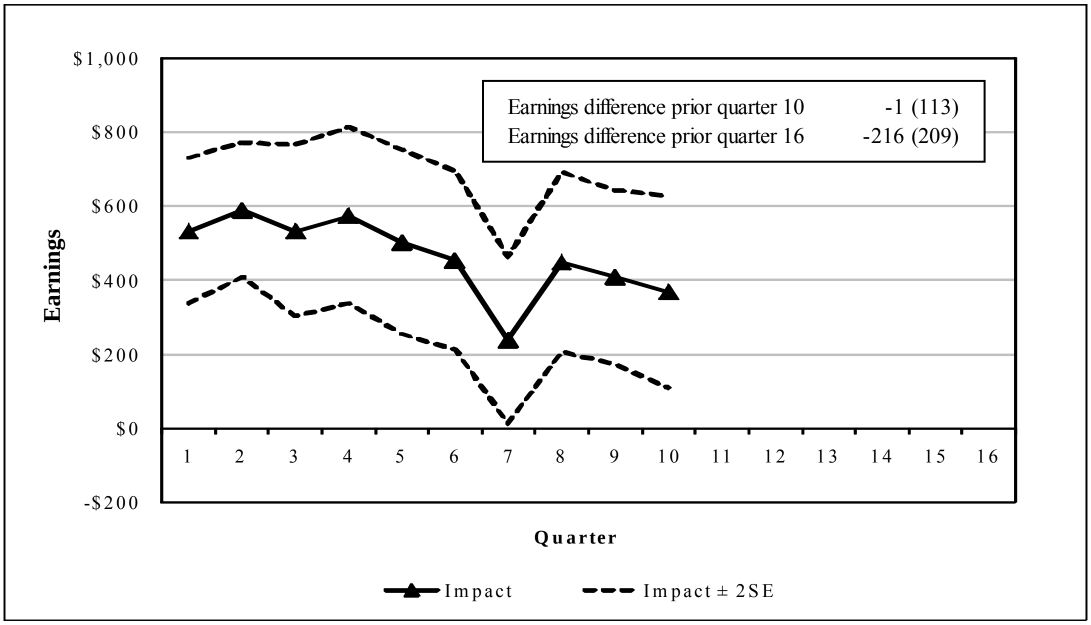
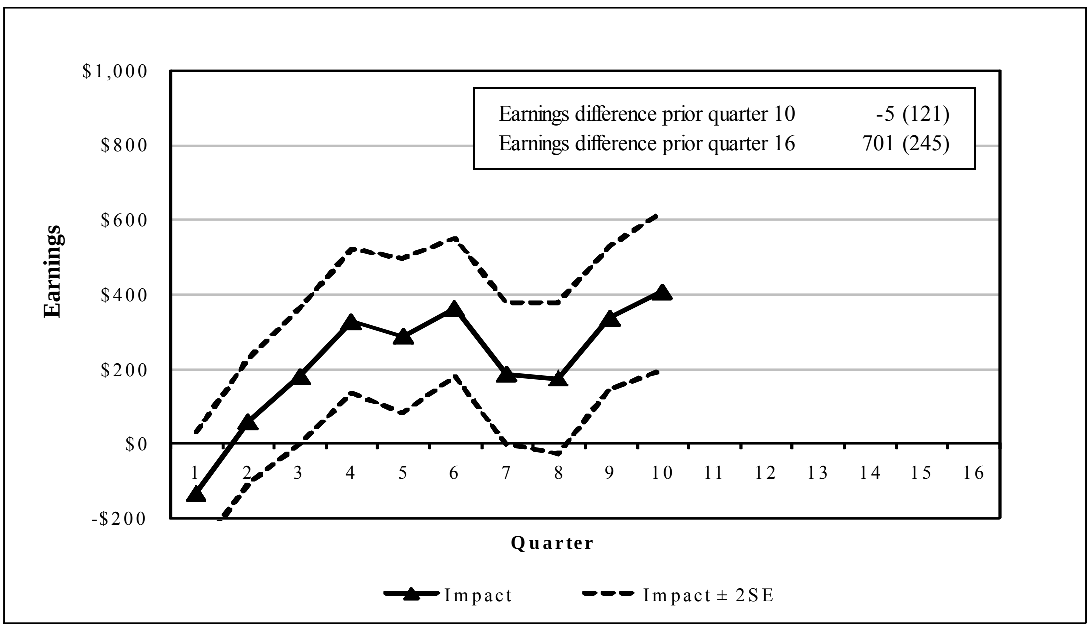


Figure VII.18
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Male Veterans, WIA versus Comparison Group



5. Veterans

The estimated impacts of the Adult program on earnings for male veterans who participated in the Adult program are presented in Figure VII.17. These estimated impacts are substantially smaller than the estimated impacts for the typical male participant in the Adult program shown in Figure V.2. However, the standard errors associated with these impacts are fairly large and any estimated difference is unlikely to be statistically significant. The estimated impacts on veterans participating in the Dislocated Worker program are shown in Figure VII.18. The initial earnings are slightly greater than that for male workers as a group (i.e., the decrement relative to the comparison group is smaller) but generally the effects are very similar to the estimated effects for the average male participant (compare Figure VI.2).

VIII. Summary and Implications

The estimates of WIA program impact presented in this report 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. The study employed state-of-the-art nonexperimental methods to compare WIA program participants with matched comparison groups of individuals who are observationally equivalent across a range of demographic characteristics, social welfare benefit receipt and labor market experiences, but who either did not receive WIA services or did not receive WIA training.

Given variation in the way that states have implemented WIA, it was anticipated that estimates of the WIA program impact would differ across states. For example, in some states, a much larger proportion of individuals receive training than in others. The response to training was also expected to differ across individuals, with some obtaining valuable skills that produce returns in the labor market and others gaining less from their participation. 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 applications 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. It is important to reiterate that these estimates are averages, and differences across states are substantial.

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. At the same time, as the Adult and Dislocated Worker programs draw from very different populations, there is a chance that observed differences between the programs may be spurious.

Impacts were estimated separately for various sub-groups (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, including: nonwhite non-Hispanics; Hispanics; individuals under 26 years old at the time they entered the program; those 50 years old or older; and veterans. For the most part, the estimated effects for the 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 subgroup. On the other hand, sampling error for these groups is quite large, so some differences could exit.

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 nonexperimental evaluation. 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.

IX. Next Steps

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. For programs in each of the omitted states, the model would then predict how state characteristics and program attributes influence expected WIA impacts, permitting the construction of a simulated measure of impact in all states. Not only would this allow inferences about overall program effect, but it would also allow estimates regarding the uncertainty associated with this estimate. A variety of assumptions would need to be made and clearly justified for a convincing analysis.

The results of such an analysis for the WIA program could then form the first element in undertaking a comprehensive benefit-cost analysis. A credible cost-benefit analysis also requires accurate information on the costs of the program and the translation of all important program benefits into monetary terms. The estimated quarterly earnings impact would be used to calculate the lifetime benefits for participants. To the extent that participants may have obtained training elsewhere, the costs avoided due to substitution of WIA services for such training would need to be included as one of the program's benefits as well.

Cost data would need to be collected, based on either survey or administrative data sources. Although federal funding makes up the bulk of the WIA budget, state funds and local workforce investment areas contribute in various indirect ways. One-Stop center cost-sharing arrangements would require that interactions with other state and local programs be considered as well. Local accounting practices, which typically do not require per-person accounting of all program expenditures, make reliance on administrative data alone for cost estimation less desirable.

It is also typical for cost-benefit analyses to calculate the costs and benefits that accrue to the general public or society as well as to individuals. For example, WIA participants who gain access to employment through the program may be less likely to engage in drug use or criminal activity, which would bring about societal benefits. One way to assess these types of benefits is to compare actual arrests or other measures of illicit behavior to predicted outcomes based on the baseline information available for individuals in the study. One can also calculate the distributional consequences of the program by separately calculating the net present value to program participants and to the rest of society, taking into account outcomes such as tax payments, receipt of social welfare support, and other transfers between participants and society.

Like any program of this size operating at a national level, it is also necessary to consider the implications of general equilibrium effects that are likely to result. That is, programs like WIA may have indirect effects on both participants and non-participants to the extent that they change the equilibrium of the labor market, influencing individual training returns and the training choices of others. For example, a resulting increase in the number of individuals entering the labor market and securing employment may have the general equilibrium effect of increasing labor supply and reducing wages. This could, in turn, have effects on other individuals' choices to enter the labor market or continue their education. On the one hand, the number of individuals participating in the WIA program is small relative to all educational and training investments made by young people. Still, because WIA programs are targeted to disadvantaged and dislocated workers, some impacts may be more localized, making these types of effects in local labor markets of consequence.

Existing research, reviewed by Lise, Seitz and Smith (2005), suggests that the equilibrium effects of large-scale policies may be substantial and can lead to very different conclusions regarding the cost-benefit performance of labor market policies than standard (partial-equilibrium) analyses. Lise, Seitz and Smith develop a framework for general equilibrium program evaluations that explicitly considers the effects of changes in financial incentives introduced by programs on the intensity with which individuals search for jobs and on the process by which wages are determined in the labor market, and that also takes into account key features of social welfare and unemployment programs that are likely to have important feedback effects on the labor market.

Each of the supplementary analyses described above—meta-evaluation, cost-benefit analysis, and analysis of general equilibrium effects—would require substantial additional data collection and programming, model development and estimation, and sensitivity testing to produce broadly generalizable information on WIA impacts. At the same time, the insights produced from a more comprehensive study of this nature would undoubtedly generate more valuable information for all WIA program managers and policymakers to use in improving day-to-day operations, program performance over time, and of course, the labor market outcomes of low-skilled and disadvantaged workers. Ultimately, policymakers want to know how effective the WIA program is in helping U.S. citizens to become more employable and productive, and understanding both individual behavioral responses as well as broader labor market and societal effects is essential. The current study represents a critical first step toward this larger evaluation goal.

References Cited

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).

Angrist, Joshua D., and Jinyong Hahn, "When to Control for Covariates? Panel-Asymptotic Results for Estimates of Treatment Effects," NBER Technical Working Paper No. 241 (1999).

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

Barnow, Burt S., Glenn G. Cain, and Arthur S. Goldberger, "Issues in the Analysis of Selectivity Bias," in E. Stromsdorfer and G. Farkas (Eds) *Evaluation Studies*, Vol. 5 (Beverly Hills, CA: Sage Publications, 1980).

Black, Daniel A., and Jeffrey A. Smith, "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching," *Journal of Econometrics* 121 (July-August 2004), 99-124.

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.

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.

Doolittle, Fred, Bell, Steve, Bloom, Howard, Cave, George, Kemple, James, Orr, Larry, Traeger, Linda and John Wallace. 1993. The Design and Implementation of the National JTPA Study. A Summary. Manpower Demonstration Research Corp., New York, NY.

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.

Friedlander, Daniel, and Philip K. Robins, "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods," *American Economic Review* 85:4 (September 1995), 923-937.

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

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

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies* 65 (April 1998), 261-294.

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, "Assessing the Case for Social Experiments," *Journal of Economic Perspectives* 9 (Spring 1995), 85-110.

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.

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., "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrica* 87 (2000), 706-710.

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.

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).

Lise, Jeremy, Shannon Seitz and Jeffrey Smith. 2005. "Evaluating Search and Matching Models Using Experimental Data." Working paper, University of Michigan.

Manski, Charles F., "Learning About Treatment Effects from Experiments with Random Assignment of Treatments," *Journal of Human Resources* 31 (Fall 1996), 709-33.

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).

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., and Donald B. Rubin, “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score,” *The American Statistician* 39 (February 1985), 33-38.

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).

Zhao, Zhong, “Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence,” *Review of Economics and Statistics* 86:1 (February 2004), 91-10

Workforce Investment Act Non-Experimental Net Impact Evaluation

FINAL REPORT
APPENDIX

December 2008

Lead Investigators

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

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

Kenneth R. Troske
University of Kentucky

Project Director

Jacob M. Benus
IMPAQ International, LLC

APPENDIX 1

Variable Coding

It is necessary to identify the most important factors that influence program entry and ultimate earnings. In addition to demographic measures, employment experience over the prior 18 months is clearly critical. Extensive controls for prior program participation are also included. In order to undertake the matching, these measures are included in a logit that predicts participation in the treated group for the combined sample of treated and comparison cases. What follows is a description of the variables included in these analyses and an explication of the coding strategy.

Table A1 first lists standard demographic measures. In many cases, continuous terms are combined with dummy variables in order to assure that matching captures all differences. In controlling for age, the specification includes a linear and square term in age as well as dummies to capture differences that might not be fully captured in a polynomial. Very few WIA participants in the Adult or Dislocated Workers program are under 18, and such individuals were omitted from the analysis.

Table A1
Coding of Measures Used in Matching

Variable	***Coding***
Demographic Measures	
Gender	Separate analyses
Age	Linear term Squared term Dummies for the categories: 18 =< age < 26 26=< age < 36 36=< age < 50 50=< age
Race	Dummies for the following exclusive categories: White only (omitted category) African American (including African American combined with other racial category/ies) Other, including other multiple race No racial category identified Categories simplified as necessary to assure consistency between WIA and comparison program participants
Hispanic	Hispanic coded regardless of race if available (missing coded as not Hispanic)

Education	Linear term for years of education Also dummies for the following categories: Less than 12—not finished HS 12—high school degree High school degree equivalency 13-15—some college 16 bachelor's degree 17-20—masters doctorate Education not reported
Disability	Dummy
Veteran	Dummy
Geographical identifier	Generally coded into WIB areas WIB areas combined or divided up as necessary to capture local labor markets
Date of program entry	Matching undertaken within quarter except where sample size precludes it. Where individuals entering in different quarters are combined, dummies identify quarter Date of entry coded as days since January 1, 2000
Prior Employment	
Prior employment transitions	Based on earnings data for quarters in prior five quarters and current quarter, coded into four exclusive categories, captured by dummies: Employed-employed if quarters -5 to 0 all have positive earnings Employed-not employed if quarter 0 has zero earnings and any quarter from -5 to -1 has positive earnings Not employed-employed if quarter 0 has positive earnings and any quarter from -5 to -1 has zero earnings Not employed-not employed if no quarter -5 to 0 has positive earnings
Prior earnings and employment	Separate measures for each of the previous eight quarters: Linear earnings Square of earnings Dummy positive earnings
Employment interactions based on history	Interactions identifying: Positive earnings in both immediately prior quarters Positive earnings in all three immediately prior quarters Positive earnings in all four immediately prior quarters No earnings in both immediately prior quarters No earnings in all three immediately prior quarters No earnings in all four immediately prior quarters

Prior Program Experience	
WIA prior experience	Dummy variables identifying WIA participation in immediately prior year WIA participation two years prior
Unemployment Insurance prior experience	For prior quarter, dummies identifying the following exclusive categories: No UI experience Claim and receive UI benefits Receive UI benefits but no claim in quarter Claim but did not receive any UI benefits in quarter Missing information For second quarter prior, dummies identifying categories as above For third and fourth prior quarters (combined) dummies identifying quarters as above For second year prior, dummies identifying categories as above For third year prior, dummies identifying categories as above For fourth year prior, dummies identifying categories as above Interactions for experience across time periods: Dummy for any experience in prior two quarters Dummy for any experience in prior year Dummy for any experience in prior two years
U.S. Employment Service (ES) prior experience	For prior quarter, coded into the following nonexclusive dummies: Received 1 to 5 service events Received more than 5 service events Number of months received service events Received counseling services Missing information For second quarter prior, dummies identifying categories as above For third and fourth prior quarters (combined) dummies identifying quarters as above For second year prior, dummies identifying categories as above For third year prior, dummies identifying categories as above For fourth year prior, dummies identifying categories as above Interactions for experience across time periods: Dummy for any experience in prior two quarters Dummy for any experience in prior year Dummy for any experience in prior two years

Prior TANF experience (females only)	Dummy for TANF receipt in current quarter Dummy for TANF receipt in prior quarter Dummy for TANF receipt in third and fourth quarters prior Dummy for TANF receipt in second year prior
Current Program Experience (Comparison 3 only)	
Current experience in alternative program	Coded same as prior experience
Time since previous job or layoff	Coded in days only for those with current UI experience Dummy for those with no current UI experience

Race was classified into four categories. This categorization allows for the fact that, in some states, a substantial proportion of recipients do not identify a racial category, and if this is the case for both WIA participants and comparison group members, these are permitted to match. Also included is a Hispanic variable, which is potentially independent of race, although in some cases state coding practices clearly coded Hispanics into the "no race" category. In many of the states, racial and ethnic categorization differed for WIA and comparison program participants. In some states, Hispanic was treated as an alternative racial category, whereas in other cases Hispanics could identify in any racial category. No attempt was made to classify race and ethnicity in a comparable way across all states—a task of little value, given the differences in racial composition between states. Efforts focused on assuring that the coding was comparable for WIA and comparison program participants.

Education is captured as a linear term in years of attained schooling, as well as up to seven dummy variables for ranges of schooling. In most states, high school equivalency is not separately identified, and so such individuals are coded with high school graduates. In several of the states, a large portion of individuals were coded as having zero years of education in the original data. In cases where this number was clearly implausible, they were recoded as missing. WIA participants with missing education were not generally omitted. Where the coding was clearly the same for WIA and the comparison group participants, missing education was identified with a dummy, which would allow those missing education to match for WIA and the comparison group. In some states, the coding system for education was clearly different for WIA participants and comparison individuals, so variables were recoded to make them at least roughly comparable.[47]

[47] In some cases, comparison cases with missing education were omitted. In other cases, missing values were recoded to 12 years of education where this was the best way to obtain comparability.

A disability measure was not available consistently for WIA and the comparison program in any state, but, in five states, the disability measure was used for WIA participants in comparison 3 analysis. A dummy identifying veterans was available in seven states.[48] For both of these variables, cases coded as missing were recoded to zero where the frequencies suggested that this was appropriate. Cases with missing values on these variables were not discarded.

In many states, WIA participants were only identified by their WIB areas. Fortunately, in most cases, these corresponded at least roughly to local labor market areas, or they aggregated areas with similar social and economic structure. Where these areas appeared inappropriate because they combined very different regions, they were divided into smaller areas if information was available. Generally, the geographic coding for the comparison group differed from that for WIA participants, but in most cases counties could be identified, allowing comparison cases to be placed into areas consistent with the WIA data. In three states, geographic area was not available for the comparison group, and geographic identifiers on wage record data were used to allow matching by area.[49]

In addition to controlling for quarter of entry (usually requiring an exact match), the actual date of program entry within a quarter was also controlled, in order to assure that systematic differences in date of entry within a quarter were taken into account.

Extensive measures are available to control for prior employment experience. Following Heckman et al. (1999), dummies were coded identifying transitions into or out of employment based on six quarters of wage record data. These attempts to capture transitions are limited in several important respects. Rather than identifying all relevant employment transitions, it is only possible to determine if an individual is employed during the course of a full quarter. Some transitions undoubtedly are missed because of the coarseness of this measure. A second problem is that unemployment cannot be distinguished from time out of the labor force. (Below measures of participation in the UI system or U.S. Employment Service system are described, which capture some of these differences.)

Earnings in the eight quarters prior to program participation are each controlled using three measures, a linear term, the square of earnings, and a dummy that identifies whether earnings were greater than zero. Also controlled were measures indicating, whether individuals were

[48] In one state, the veteran measure was available only for WIA participants in comparison 3 analysis.

[49] Geographic identifiers on the wage record data generally identify place of employment—or in some cases, the location of a central office that processes paperwork—and so may not be suitable for matching with WIA location data. For this reason, wage record data was used to obtain geographic information for *both* WIA and comparison group participants when geographic information was unavailable for either WIA or the comparison sample. For up to 10 percent of WIA participants, no wage record information was available. In these cases, the geographic information on the WIA file was used to assign these individuals to one of the areas identified in the wage record data. In these states, the very small number of comparison cases for which no wage record data were available (less than 3 percent of the sample) were omitted.

continuously employed, or not at all employed, in both immediately prior quarters, in all three immediately prior quarters, and in all four immediately prior quarters. As a result of these detailed controls, detailed work history over the prior two years correspond closely for treated and comparison cases.

Prior program experience is captured, first, by dummy variables identifying participation in WIA in each of the two previous years. This assures that where a treated case was also in WIA in a previous period, it is matched with a comparison case that also participated in WIA. (Note that a comparison case may involve an individual who previously participated in WIA; however, those who subsequently participate in WIA are omitted from the comparison pool.)

Experience in the comparison program in prior quarters is captured by a series of dummy variables. In some states, there are UI claimants who are denied benefits but are available on the files provided by the state, and these are separately identified. For most claimants, benefits begin within a few weeks of the claim, and so this group is identified as well. Finally, a third group consists of individuals who, during some quarters, receive continuing benefits after making a claim in the prior quarter. These dummy variables identify experience in the prior quarter, as well as in five earlier time periods. Also included are interactions that identify experience across more than one time period. These dummies assure that WIA participants who have prior UI experience are matched with comparison group members with similar patterns of experience.

In the case of U.S. Employment Service (ES), for the same periods the extent and intensity of service are captured with four dummy variables. In those states where it is available, for women, participation in Temporary Assistance for Needy Families (TANF) in the current quarter and prior periods is identified.

For comparison 3, matching was performed on comparable information for whether the WIA participant was involved in the alternative program (UI or ES) in the current quarter.

Cases omitted due to missing data. As detailed above, coding schemes were chosen to avoid omitting treated cases (i.e., WIA participants) from the analysis. Of course, any case missing program entry date or not identified by program could not be considered, since there was no way to determine whether it was in the universe. In addition, cases missing data on gender and age, or those whose birthdays indicated an age of less than 18 at the time of program entry, were omitted. In addition, in some cases a variable was not consistently coded in WIA and the comparison group, and, as a result, cases where matching appeared problematic were omitted. The total number of cases omitted for these reason was 1 percent or less in 11 states. In one state, inconsistency in the coding of several variables required the omission of slightly under 6 percent of the WIA cases in the universe.

Interaction variables used for balancing tests. The measures listed above identify the basic variables included in the logit equation for propensity score estimation. Balancing tests focused on the difference in the means for these variables, squares of the continuous variables, and interactions involving the following variables: any UI experience in the prior eight quarters, age, black race, years of education, the employment transition variables, employment in prior two quarters, employment in the third through eighth prior quarters, total earnings in the prior two quarters, total earnings in the third through eighth prior quarters, and dummies for geographic units. Although not all the listed variables are included in the regression as coded above, component variables that capture the same information are included in the regression. Hence, an interaction is only included in a specification that controls for the linear terms in the interaction. Depending on the number of regions, the number of resulting interaction terms ranged from 80 to over 200.

Balancing tests count the number of differences that are statistically significant at the 0.05 level, considering a matching attempt to have failed the balancing test if substantially more than 5 percent are significant. As described in the text, when the balance test failed, interaction terms were entered in groups, depending on sample size and the pattern of interactions. In those cases, for example, where significant interactions involved a particular variable, all interactions involving that variable were included. Where the sample size was large, many interactions were included, whereas in smaller samples interactions were included in more selectively.

APPENDIX 2

Standard Errors

Abadie and Imbens (2006) provide an approach for estimating the unconditional standard error of the matching estimator for the impact of the treatment on the treated. The variance of the estimator (i.e., the square of the standard error) may be written as:

$$SE^2 = V^t / N_1, \tag{2-1}$$

where

$$\begin{aligned} \hat{V}^t &= \frac{1}{N_1}\sum_{W_i=1}(Y_i(1) - \hat{Y}_i(0) - \hat{\tau}_M^t)^2 \\ &+ \frac{1}{N_1}\sum_{i=1}^{N}(1-W_i)\left[\left(\sum_{j=1}^{N}\frac{K(i,j)}{M(j)}\right)^2 - \left(\sum_{j=1}^{N}\frac{K(i,j)}{[M(j)]^2}\right)\right]\hat{\sigma}^2(X_i,W_i) \end{aligned} \tag{2-2}$$

$K(i,j)$ is an indicator of whether comparison case i is matched with treated case j (1 if matched, 0 otherwise) (set to 0 if j is a comparison case rather than a treated case),
$M(j)$ is the number of comparison cases matched with treated case j (arbitrarily set to 1 if j is a comparison case),
N_1 is the number of treated cases,
N is the total number of cases,
$Y_i(1)$ and $\hat{Y}_i(0)$ are the observed outcomes for treated case i and the estimate of the untreated outcome for case i (based on cases matched to i),
W_i is an indicator of whether case i is treated (1) or comparison (0),
$\hat{\tau}_M^t$ is the estimated average effect of the treatment on the treated, and
$\hat{\sigma}^2(X_i,W_i)$ is an estimate of the variance of the dependent variable conditional on particular values of X, for treated (W_i=1) or comparison (W_i=0) cases.

Each of the measures in (2-2) is readily available, except for $\hat{\sigma}^2(X_i,W_i)$, which must be estimated. It is estimated using a matching procedure, but where treated cases are matched with treated cases and comparison cases are matched with comparison cases. The formula is

$$\hat{\sigma}^2(X_i,W_i) = \frac{J(i)}{J(i)+1}\left(Y_i - \frac{1}{J(i)}\sum_{m=1}^{J}Y_{l_m(i)}\right)^2 \tag{2-3}$$

where

$Y_{l_m(i)}$ is the outcome for the the m^{th} closest case to case i,
$J(i)$ is the number of cases within the radius distance of case i.

However, Imbens and Wooldridge (2008) suggest a conceptually simpler estimate identifying the *conditional* standard error:

$$SE^2 = \sum_{i=1}^{N} \lambda_i^2 \hat{\sigma}^2 (X_i, W_i)$$

where λ_i is the weight associated with case i. In the case of radius matching and for estimates of the treatment on the treated, $\lambda_i = 1/N_1$ for treated case i, and

$\lambda_i = \dfrac{1}{N_1} \displaystyle\sum_{j=1}^{N_1} \dfrac{K(i,j)}{M(j)}$, for a comparison case i.

The expression can then be written as

$$SE^2 = \frac{1}{N_1^2} \sum_{i=1}^{N} \left(W_i - (1 - W_i) \sum_{j=1}^{N_1} \frac{K(i,j)}{M(j)} \right)^2 \hat{\sigma}^2 (X_i, W_i). \qquad \text{(2-4)}$$

Both the unconditional standard error estimate (2-1) and the conditional standard error estimate (2-4) assume that matching is undertaken on the basis of the independent variable matrix *X*, not on the basis of a propensity score. However, if *X* is replaced with the true propensity score in each formula, the proofs all follow. The approach in this study uses the predicted (i.e., estimated) propensity score in place of *X*. Imbens (2008) notes that there do not appear to be any formal proofs that the predicted propensity score can be used in the above formulas, but he nonetheless recommends this approach, conjecturing that results will likely be conservative, i.e., estimates of standard errors will be larger than true values.

Imbens (2008) also points out that substituting the propensity score for the vector of independent variables requires that the conditional estimate of the standard error be subject to reinterpretation. Rather than representing the standard error of an estimate conditional on the particular values of *X*, it represents the standard error conditional on a given propensity score. The conditional standard error based on the propensity score will be greater than one based on the values of the independent variables. The findings in this study that the conditional and unconditional standard errors are not substantially different may be partly a result of use of the propensity score.

APPENDIX 3

Radius Choice

Galdo, Smith and Black (2008) provide an extended discussion of optimal radius choice. The study here undertook a preliminary analysis for two states using the methods suggested by them, focusing on earnings in the fifth quarter after WIA entry and earnings in the twelfth quarter after WIA entry.

Galdo et al. argue that the optimal matching radius is one that minimizes the mean integrated standard error (MISE), which is estimated as

$$MISE = \frac{1}{N_C}\sum_{j=1}^{N_C}(Y_{0j} - \hat{m}_{-j}(\rho_j, h))^2 W_j$$

where

N_C is the number of comparison cases (the summation is over comparison cases),
Y_{0j} is the outcome variable for comparison case j,
$\hat{m}_{-j}(\rho_j, h)$ is the mean outcome for those cases matched to j (comparison cases other than j),
ρ_j is the propensity score for case j,
h is the radius, and
W_j is a weight.

The basic idea is to estimate the extent of the deviation of actual values from "estimated" values. The weight represents the treated case distribution. Galdo et al. suggest several alternative ways to estimate the weight. They don't find important differences. The preliminary work here uses

$$W_j = \frac{\rho_j}{1-\rho_j}\left[\sum_{l=1}^{N}\frac{\rho_l}{1-\rho_l}\right]^{-1},$$

where the summation is over all comparison cases. The benefit of this calculation of the weight is that the weights are directly calculated from the estimated propensity scores. In contrast, the other methods they propose require that matching be undertaken, and the appropriate radius value for the matching is itself uncertain.

In estimating these values, the preliminary analyses undertaken here found that a substantial number of the comparison cases did not match with any other comparison case. Naturally, the number of such unmatched cases increased with smaller radius values. The above formula needs to be modified to take account of cases that do not match, so that MISE is estimated for matched cases. In particular, where W_j is defined as above, and the summation is over all comparison cases *whether or not they match,* the weighted sum of cases that do not match may be written as

$$PNM = \sum_{i \in \{unmatched\}} W_i \ .$$

Noting that $\sum_i W_i = 1$, it is clear that the proportion of cases that match is (1-*PNM*). In order for *MISE* to apply to the cases that do match, one can write

$$MISE = \frac{1}{1-PNM}\frac{1}{N_C}\sum_{j \in \{matched\}} (Y_{0j} - \hat{m}_{-j}(\rho_j, h))^2 W_j = \frac{1}{N_C}\sum_{j \in \{matched\}} (Y_{0j} - \hat{m}_{-j}(\rho_j, h))^2 \frac{W_j}{\sum_{j \in \{matched\}} W_j}$$

As noted in the text, this study undertook the calculation of the MISE for alternative radius values. In some cases, the number of matched cases declined substantially as the radius value was reduced, and so it was possible to reject these radius values on this basis alone. However, for the most part, there was no meaningful trade-off between number of cases matched and the MISE value for cases that did match, because larger radius values increased the proportion of matched cases *and* reduced MISE for cases that did match.

APPENDIX 4

Subgroup Analysis

This appendix provides estimates for subgroups, supplementing the discussion in section VII of the report. Although the subgroup analysis applies the same methods as used for estimates reported in sections V and VI, the small sample sizes complicate the implementation. Table A2 provides information about the size of the available samples for each of the subgroups. Numbers in italics identify comparisons where subgroup sizes were too small to permit meaningful analyses, either because the absolute numbers were too small, or too large a proportion of treated cases had to be dropped.

There were sufficient numbers of nonwhites available to fit all the models considered for the full sample. However, in the case of Hispanics, only comparison 1 could be fitted. The reason is that only a small number of states had substantial numbers of Hispanics in them, and these were generally small states. They tended also to be states providing training for a large portion of their participants, so there were relatively few participants who had obtained only Core/Intensive training, which served as the treatment group in comparison 2 and the comparison group in comparison 3. The Hispanic sample is also smaller than it would otherwise be because in some large states with very small proportions of Hispanics, it appeared that the variable identifying Hispanics was not comparable for the WIA and comparison groups, and so the state was omitted from the Hispanic analysis.

Because the number of cases was small and spread evenly across states (leading to difficulties in fitting models in each state), both the young and old subgroup analysis based on comparison 3 for the Dislocated Worker program were omitted. Finally, as noted in the text, too few female veterans were available to undertake meaningful analysis for them. For male veterans, too few cases were available to undertake comparison 3.

With the exceptions noted, the figures that follow provide subgroup analyses corresponding to those undertaken for the full population. For ease of reference, the designation of the table identifies the subgroup.

Table A2
Matched Treated Cases Available for Estimation: Subgroups

	Comparison	Nonwhites		Hispanics		Age less than 26		Age 50 or Over		Veterans	
		Treated Cases	Cases Analyzed	Treated Cases	Cases Analyzed	Treated Cases	Cases Analyzed	Treated Cases	Cases Analyzed	Treated Cases	Cases Analyzed
Adult Program											
Males	1. WIA vs. Comparison Group	18,781	15,869	1,065	786	14,597	13,048	3,960	3,170	5,687	4,743
	2. WIA Core/Intenstive vs. Comparison Group	15,950	13,278	*356*	*143*	11,701	9,879	3,170	2,421	4,856	3,605
	3. WIA Training vs. WIA Core/Intensive	2,831	1,532	*709*	*0*	2,896	1,733	*790*	*0*	*831*	*306*
Females	1. WIA vs. Comparison Group	25,503	22,035	1,735	1,271	19,582	17,080	4,619	4,063		
	2. WIA Core/Intenstive vs. Comparison Group	20,527	15,636	*610*	*329*	13,941	11,392	3,643	3,133		
	3. WIA Training vs. WIA Core/Intensive	4,976	3,327	*1,125*	*0*	5,641	2,103	*976*	*202*		
Dislocated Worker Program											
Males	1. WIA vs. Comparison Group	9,805	8,773	732	496	4,629	4,036	7,255	5,548	5,479	4,361
	2. WIA Core/Intenstive vs. Comparision Group	7,874	7,052	*283*	*131*	3,510	3,099	5,204	4,168	4,118	3,305
	3. WIA Training vs. WIA Core/Intensive	1,931	1,077	*449*	*0*	*1,119*	*352*	*2,051*	*627*	*1,361*	*315*
Females	1. WIA vs. Comparison Group	12,186	10,941	695	420	4,187	3,539	7,394	6,163		
	2. WIA Core/Intenstive vs. Comparison Group	9,697	8,698	*266*	*87*	3,079	2,661	4,947	4,398		
	3. WIA Training vs. WIA Core/Intensive	2,489	1,418	*429*	*0*	*1,108*	*293*	*2,447*	*749*		

Note: Numbers in italics identify comparisons for which results are not reported due to small sample size.

Figure YNG-A5
Adult Program Treatment Effect on Quarterly Earnings for Age less than 26 Females, WIA Core/Intensive versus Comparison Group

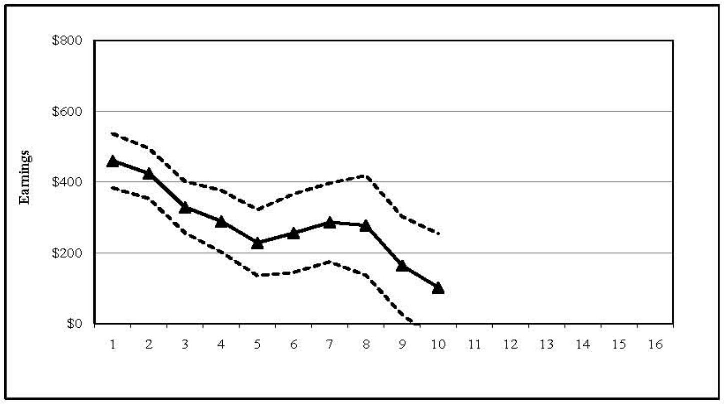


Figure YNG-A7
Adult Program Treatment Effect on Quarterly Employment for Age less than 26 Females, WIA Core/Intensive versus Comparison Group

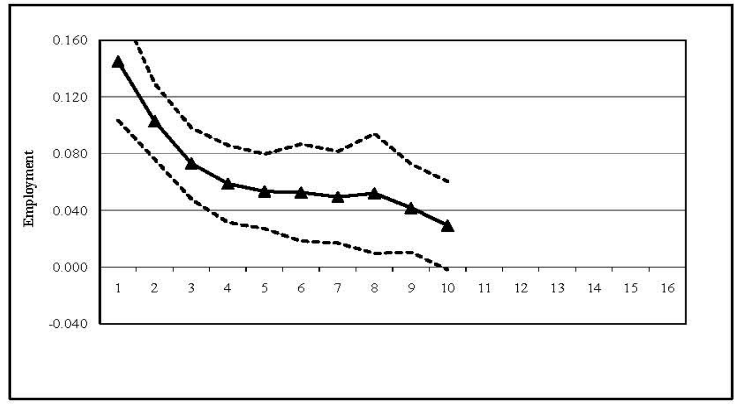


Figure YNG-A6
Adult Program Treatment Effect on Quarterly Earnings for Age less than 26 Males, WIA Core/Intensive versus Comparison Group

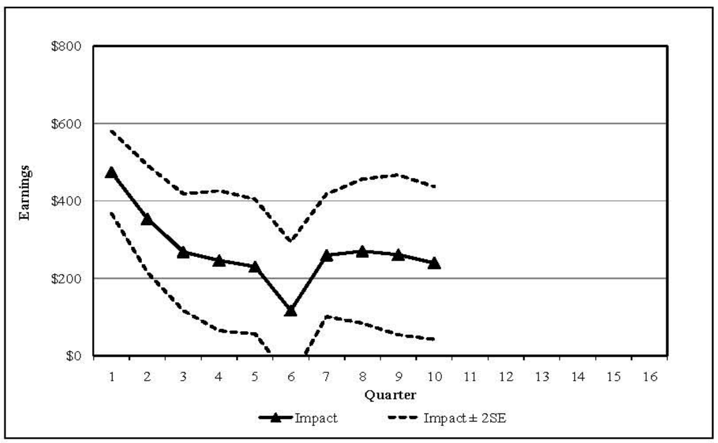


Figure YNG-A8
Adult Program Treatment Effect on Quarterly Employment for Age less than 26 Males, WIA Core/Intensive versus Comparison Group

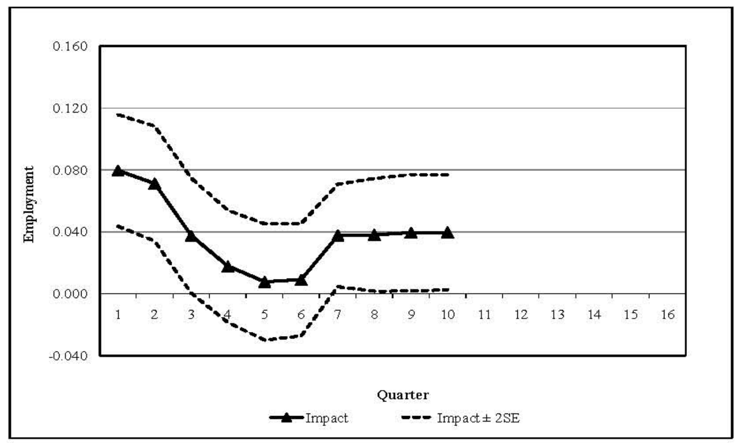


Figure YNG-A9
Adult Program Treatment Effect on Quarterly Earnings for Age less than 26
Females, WIA Training versus Comparison Group

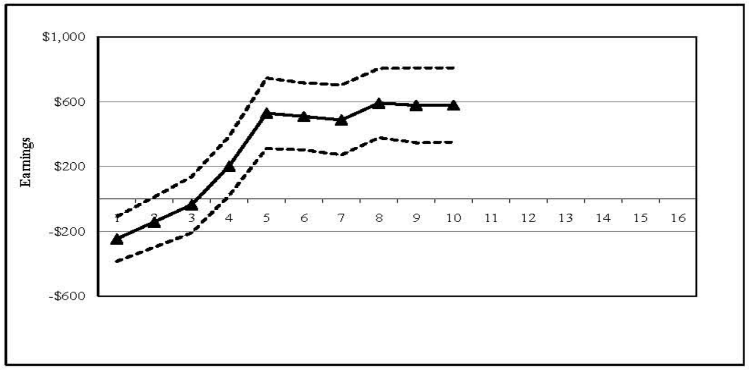


Figure YNG-A11
Adult Program Treatment Effect on Quarterly Employment for Age less than 26
Females, WIA Training versus Comparison Group

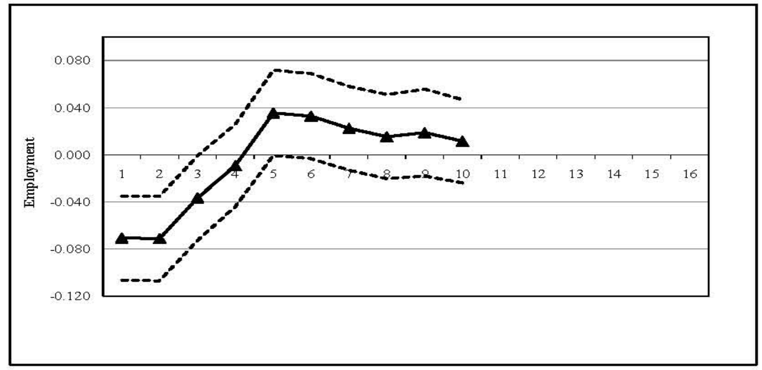


Figure YNG-A10
Adult Program Treatment Effect on Quarterly Earnings for Age less than 26
Males, WIA Training versus Comparison Group

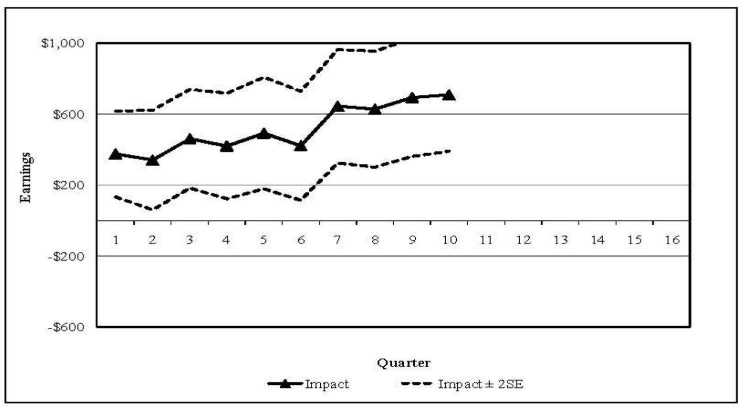


Figure YNG-A12
Adult Program Treatment Effect on Quarterly Employment for Age less than 26
Males, WIA Training versus Comparison Group

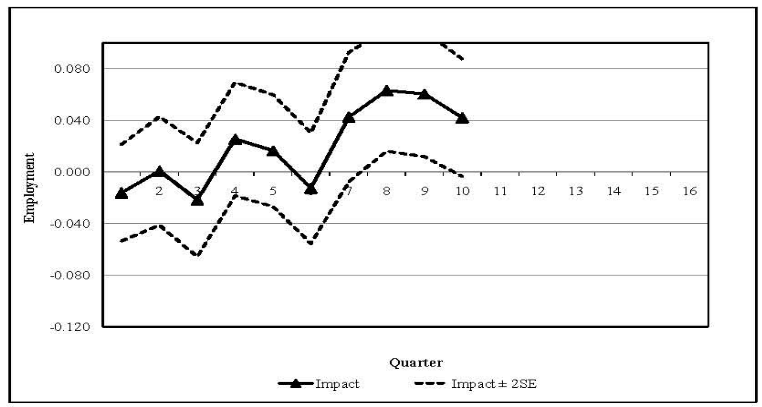


Figure YNG-A13
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age less than 26 Females, WIA versus Comparison Group

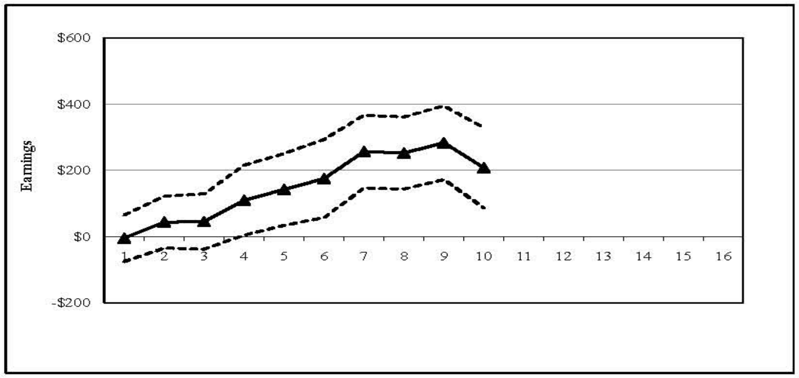


Figure YNG-A15
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age less than 26 Females, WIA versus Comparison Group

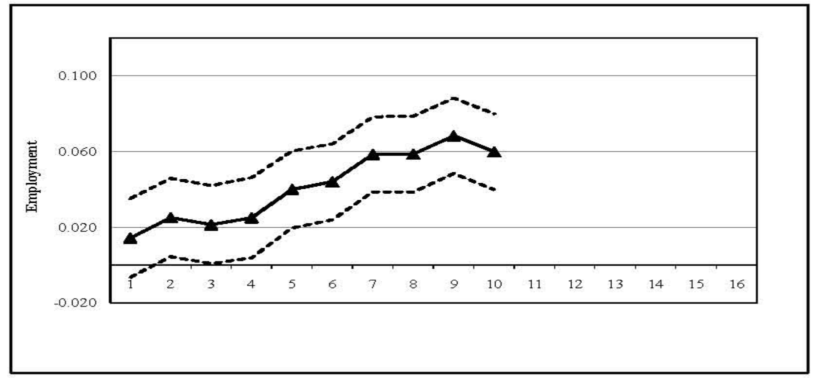


Figure YNG-A14
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age less than 26 Males, WIA versus Comparison Group

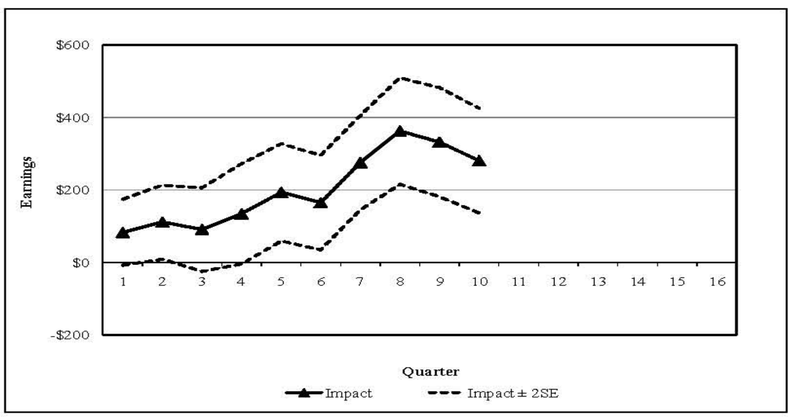


Figure YNG-A16
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age less than 26 Males, WIA versus Comparison Group

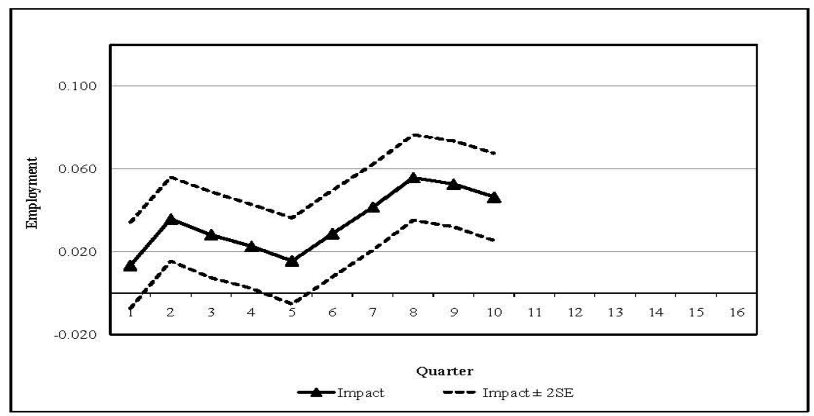


Figure YNG-A17
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age less than 26 Females, WIA Core/Intensive versus Comparison Group

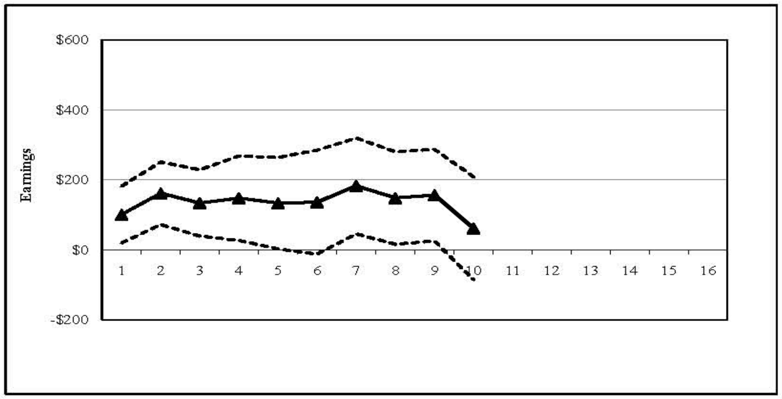


Figure YNG-A19
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age less than 26 Females, WIA Core/Intensive versus Comparison Group

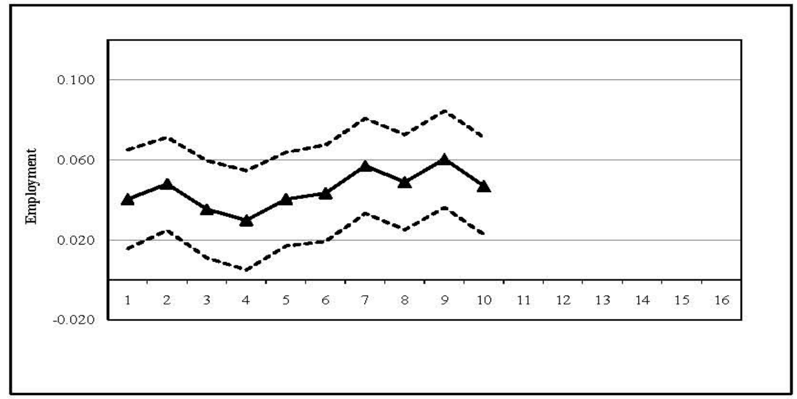


Figure YNG-A18
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age less than 26 Males, WIA Core/Intensive versus Comparison Group

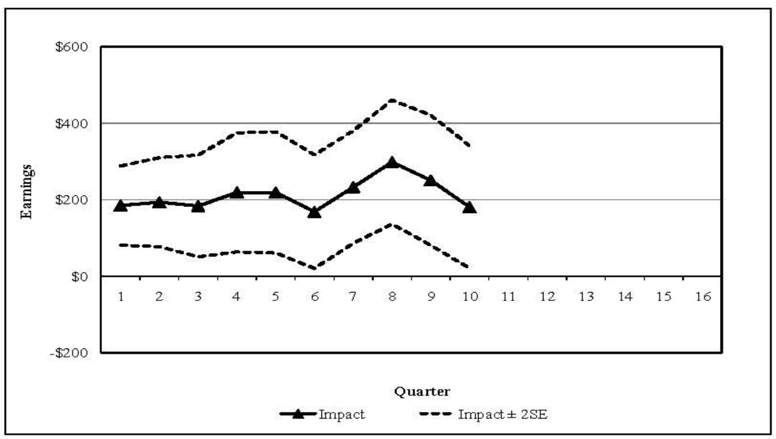


Figure YNG-A20
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age less than 26 Males, WIA Core/Intensive versus Comparison Group

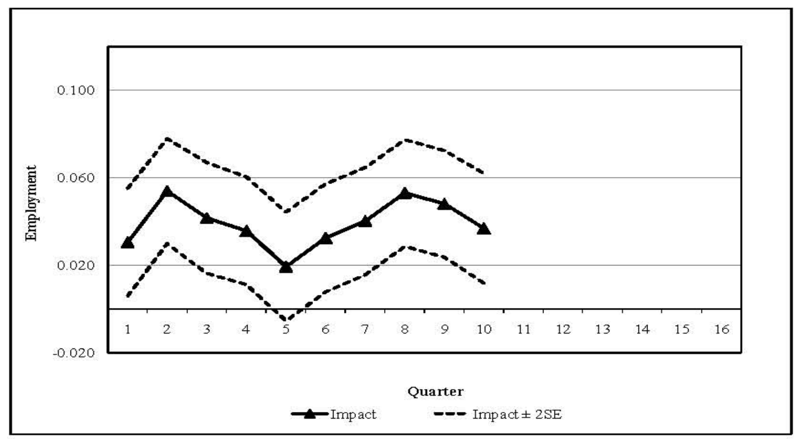


Figure ELD-A1
Adult Program Treatment Effect on Quarterly Earnings for Age 50 or Over Females, WIA versus Comparison Group

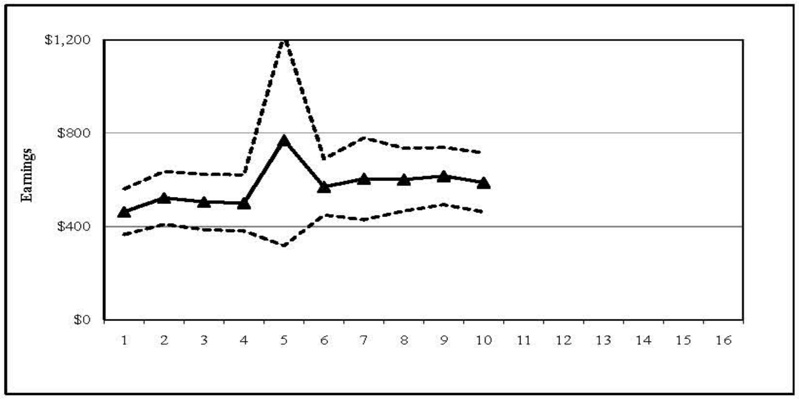


Figure ELD-A3
Adult Program Treatment Effect on Quarterly Employment for Age 50 or Over Females, WIA versus Comparison Group

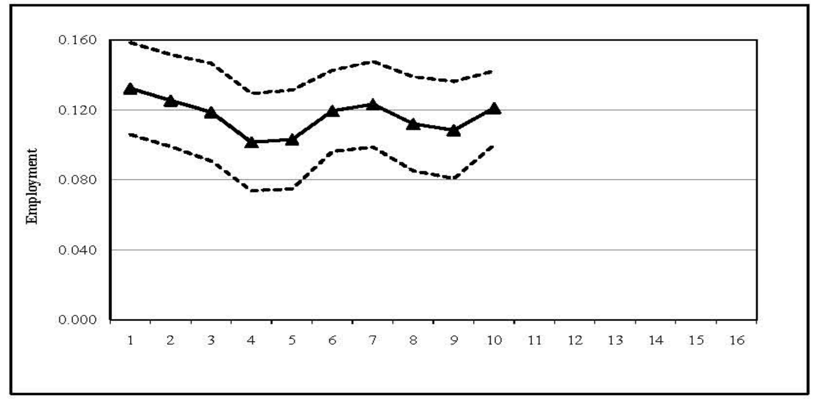


Figure ELD-A2
Adult Program Treatment Effect on Quarterly Earnings for Age 50 or Over Males, WIA versus Comparison Group

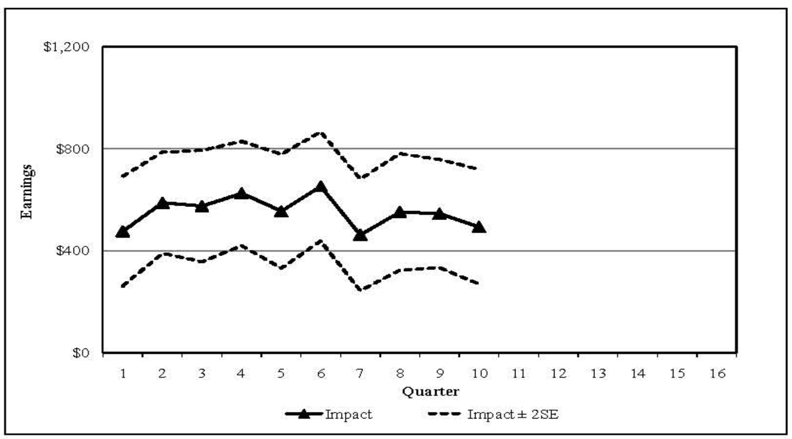


Figure ELD-A4
Adult Program Treatment Effect on Quarterly Employment for Age 50 or Over Males, WIA versus Comparison Group

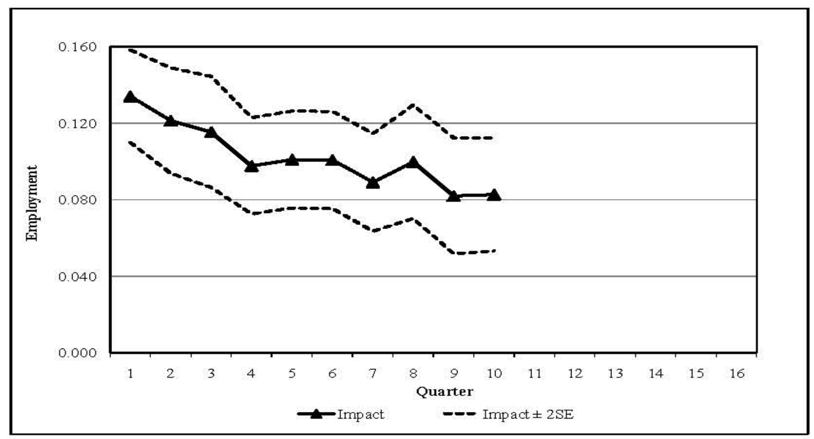


Figure ELD-A5
Adult Program Treatment Effect on Quarterly Earnings for Age 50 or Over Females, WIA Core/Intensive versus Comparison Group

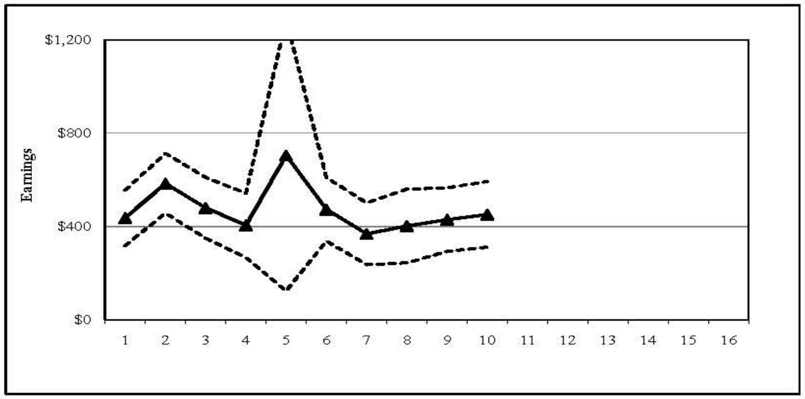


Figure ELD-A7
Adult Program Treatment Effect on Quarterly Employment for Age 50 or Over Females, WIA Core/Intensive versus Comparison Group

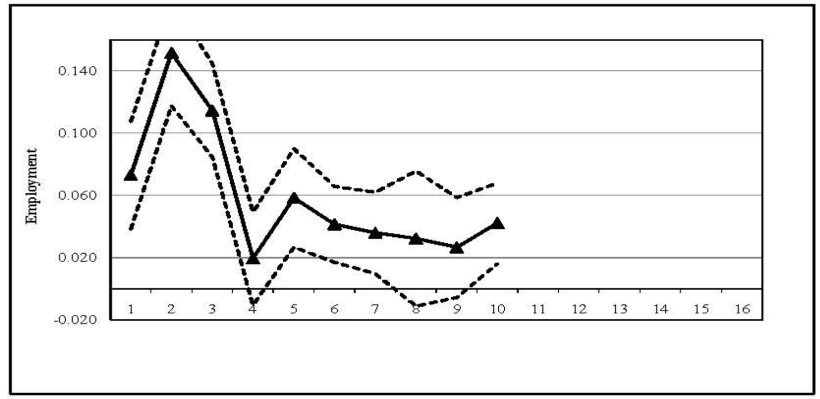


Figure ELD-A6
Adult Program Treatment Effect on Quarterly Earnings for Age 50 or Over Males, WIA Core/Intensive versus Comparison Group

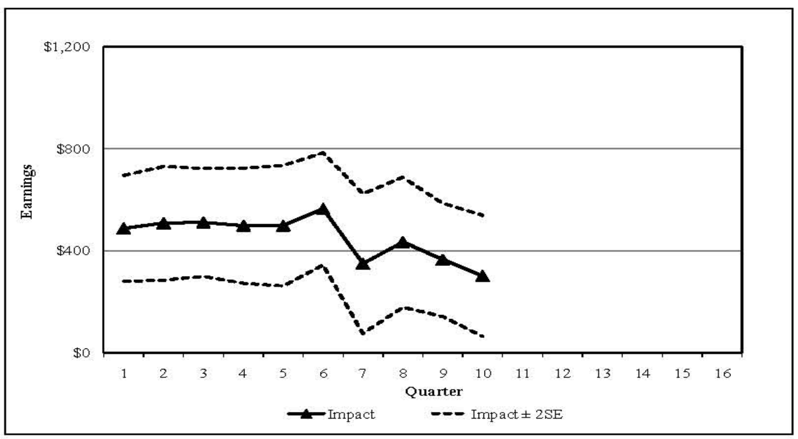


Figure ELD-A8
Adult Program Treatment Effect on Quarterly Employment for Age 50 or Over Males, WIA Core/Intensive versus Comparison Group

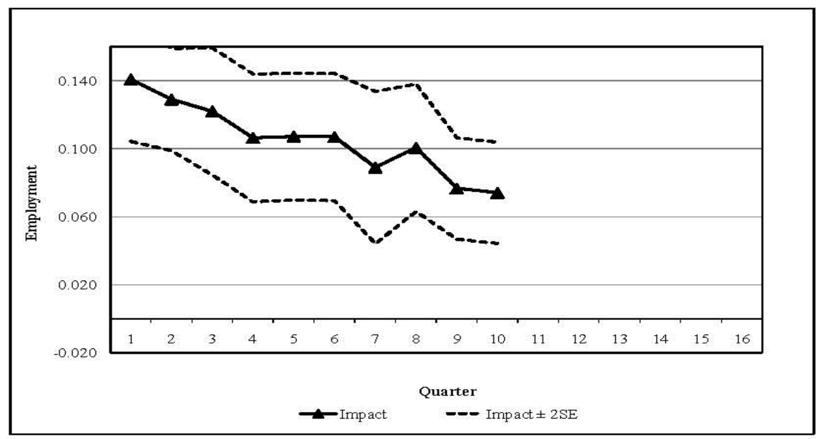


Figure ELD-A9
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age 50 or Over Females, WIA versus Comparison Group

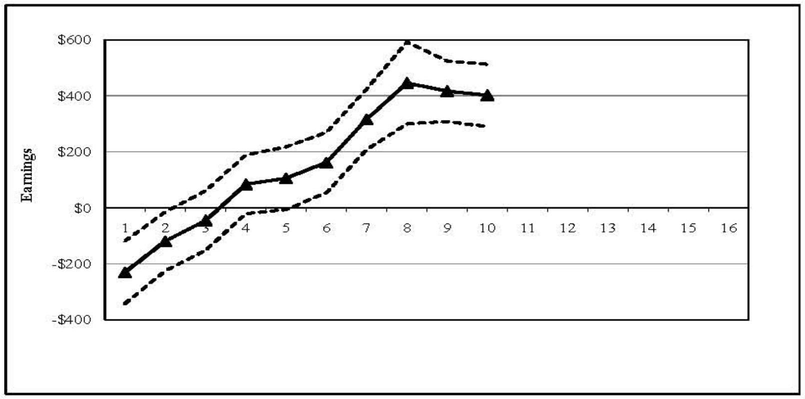


Figure ELD-A11
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age 50 or Over Females, WIA versus Comparison Group

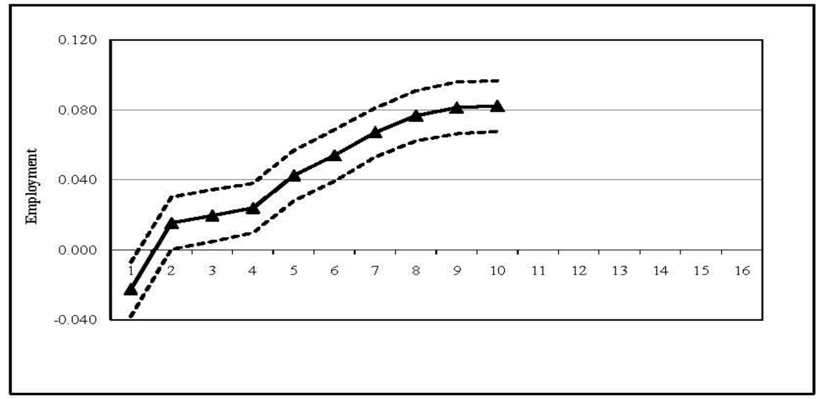


Figure ELD-A10
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age 50 or Over Males, WIA versus Comparison Group

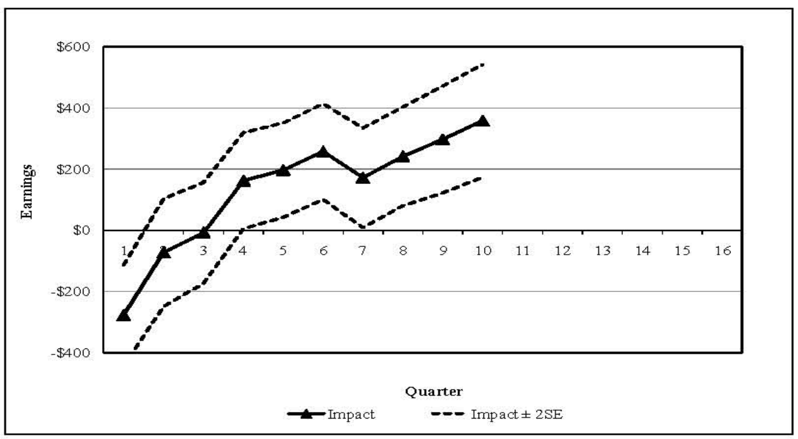


Figure ELD-A12
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age 50 or Over Males, WIA versus Comparison Group

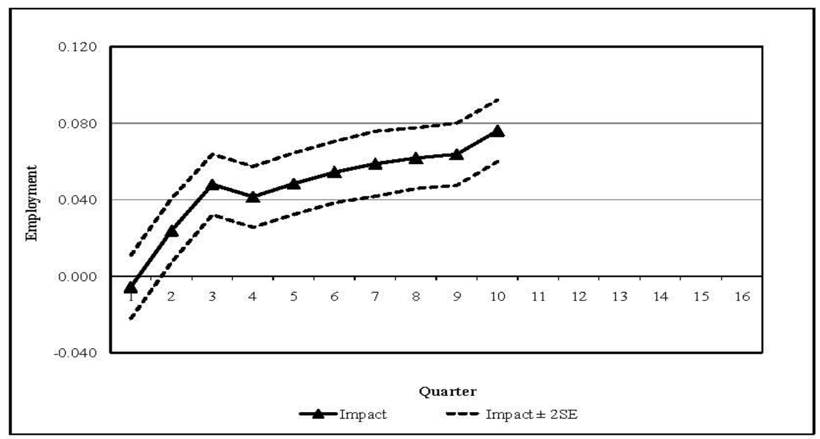


Figure ELD-A13
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age 50 or Over Females, WIA Core/Intensive versus Comparison Group

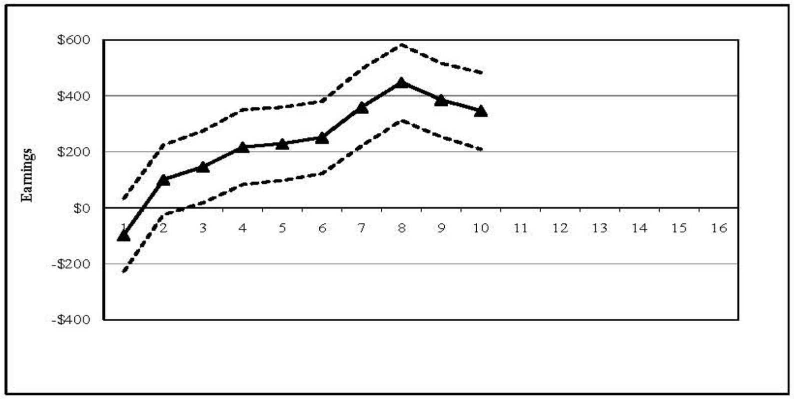


Figure ELD-A15
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age 50 or Over Females, WIA Core/Intensive versus Comparison Group

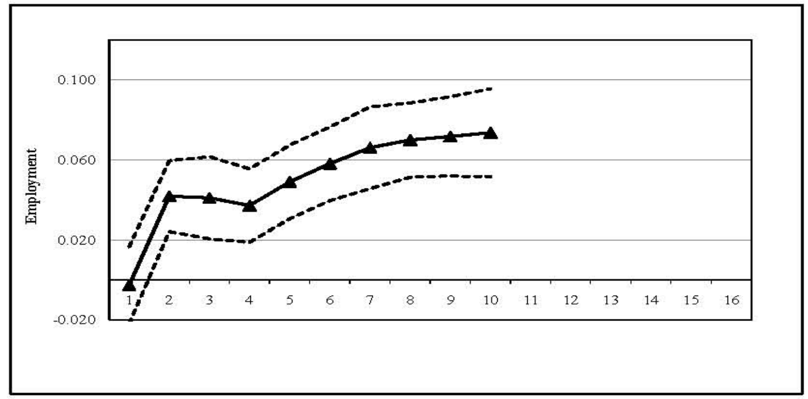


Figure ELD-A14
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Age 50 or Over Males, WIA Core/Intensive versus Comparison Group

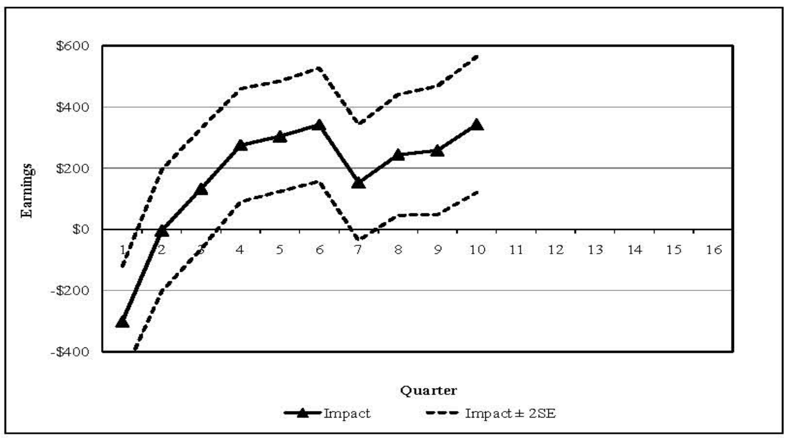


Figure ELD-A16
Dislocated Worker Program Treatment Effect on Quarterly Employment for Age 50 or Over Males, WIA Core/Intensive versus Comparison Group

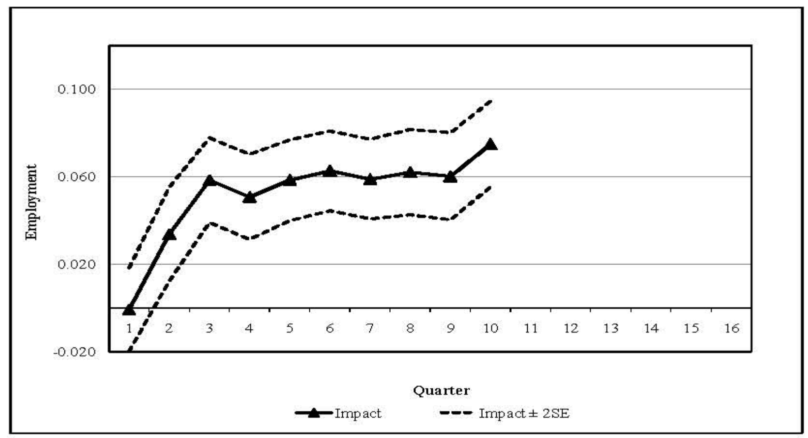


Figure V-A1
Adult Program Treatment Effect on Quarterly Earnings for Veterans Males, WIA versus Comparison Group

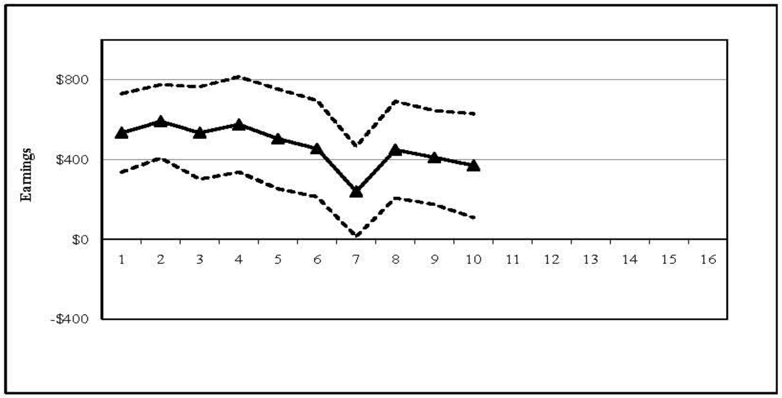


Figure V-A3
Adult Program Treatment Effect on Quarterly Employment for Veterans Males, WIA versus Comparison Group

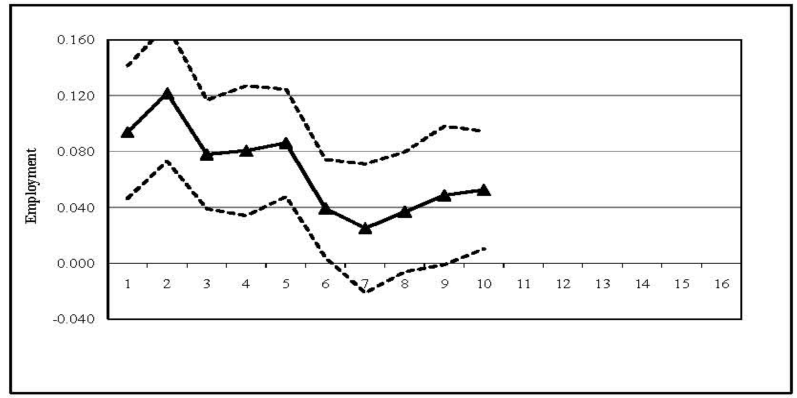


Figure V-A2
Adult Program Treatment Effect on Quarterly Earnings for Veterans Males, WIA Core/Intensive versus Comparison Group

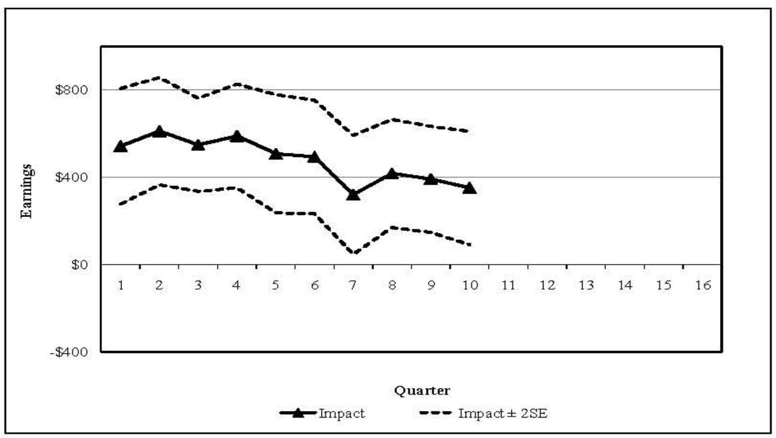


Figure V-A4
Adult Program Treatment Effect on Quarterly Employment for Veterans Males, WIA Core/Intensive versus Comparison Group

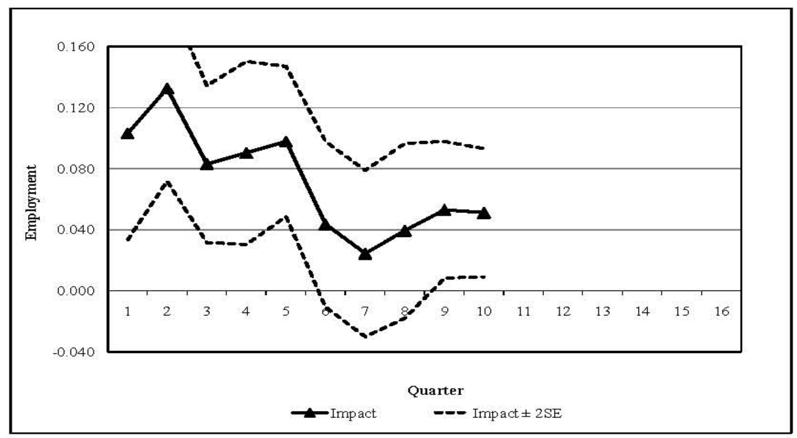


Figure V-A5
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Veterans Males, WIA versus Comparison Group

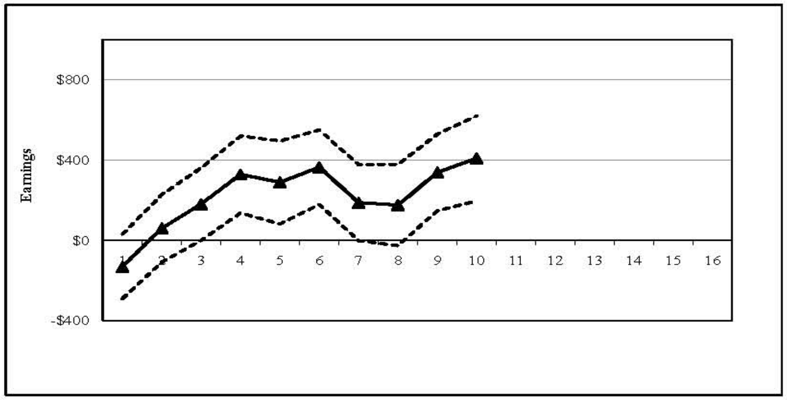


Figure V-A7
Dislocated Worker Program Treatment Effect on Quarterly Employment for Veterans Males, WIA versus Comparison Group

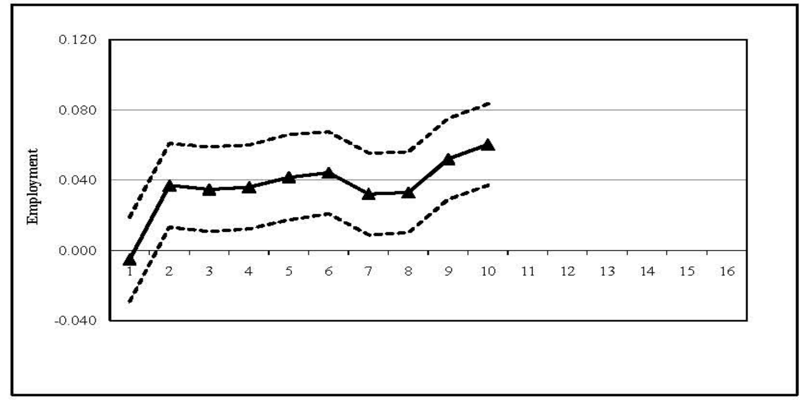


Figure V-A6
Dislocated Worker Program Treatment Effect on Quarterly Earnings for Veterans Males, WIA Core/Intensive versus Comparison Group

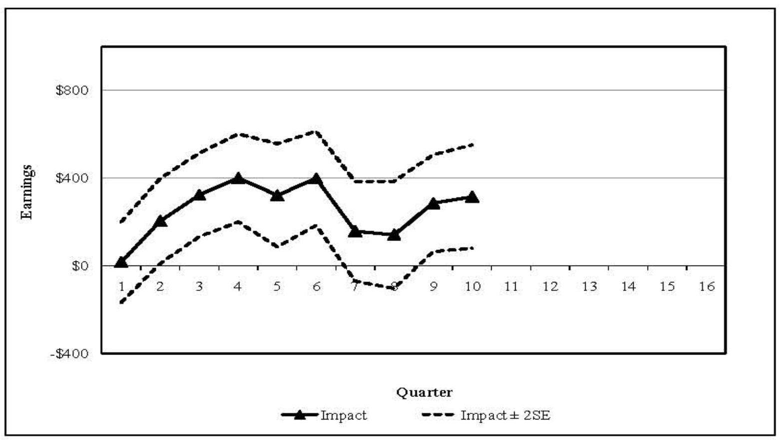
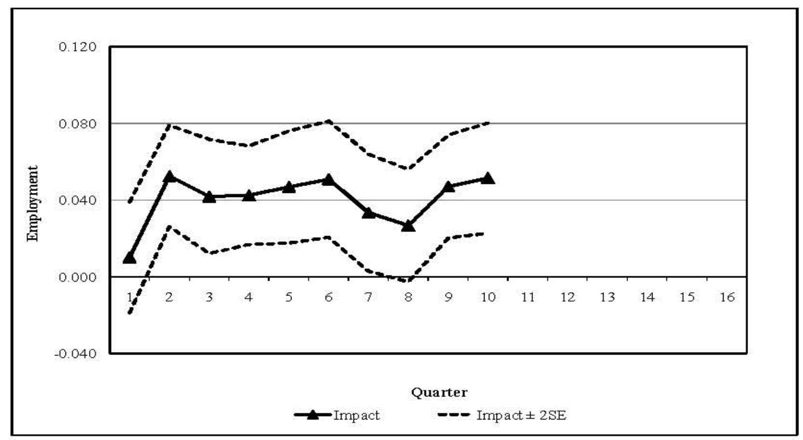


Figure V-A8
Dislocated Worker Program Treatment Effect on Quarterly Employment for Veterans Males, WIA Core/Intensive versus Comparison Group



APPENDIX 5

Earnings Patterns for WIA Program Participants

The following tables provide mean earnings for male and female participants in the Adult and Dislocated Worker programs, disaggregated by service level. The horizontal axis identifies earnings relative to the date of entry into the program. For each series, the sample is limited to program participants with complete data on earnings eight quarters prior to program entry and 10 quarters following program entry. The confidence intervals are also presented, identifying the point two standard errors above the mean and two standard errors below. In contrast to impact estimates, the means are very precisely estimated.

Figure ADULT EARN – A1
Mean Earnings for Adult Program Participants: Females

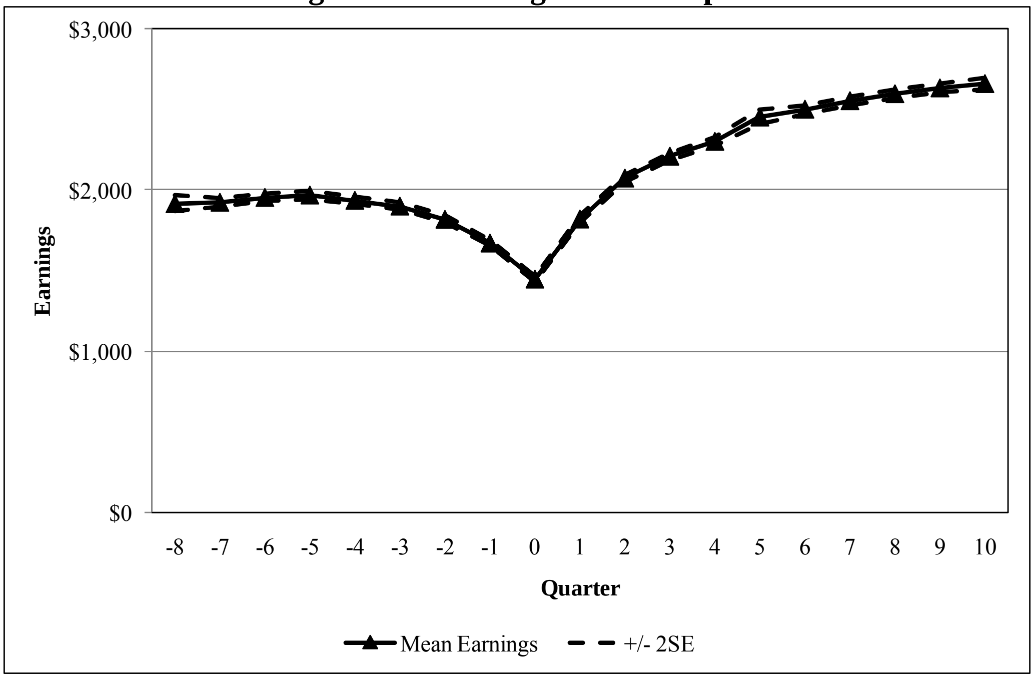


Figure ADULT EARN – A2
Mean Earnings for Adult Program Participants: Males

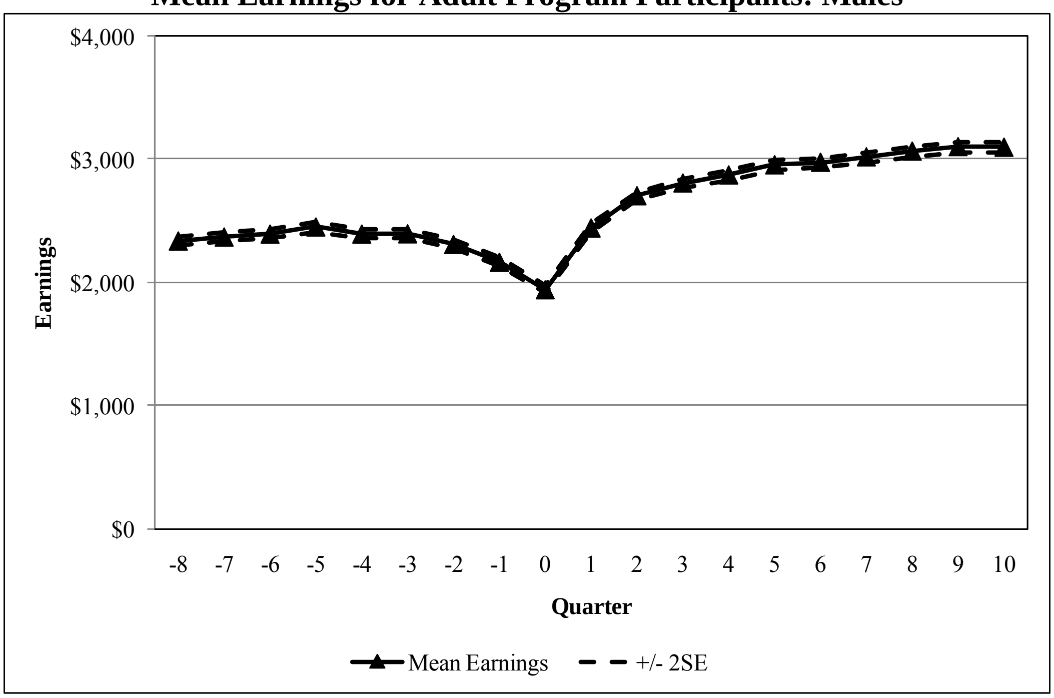


Figure ADULT EARN – A3
Mean Earnings for Adult Program Participants Receiving Core/Intensive Services: Females

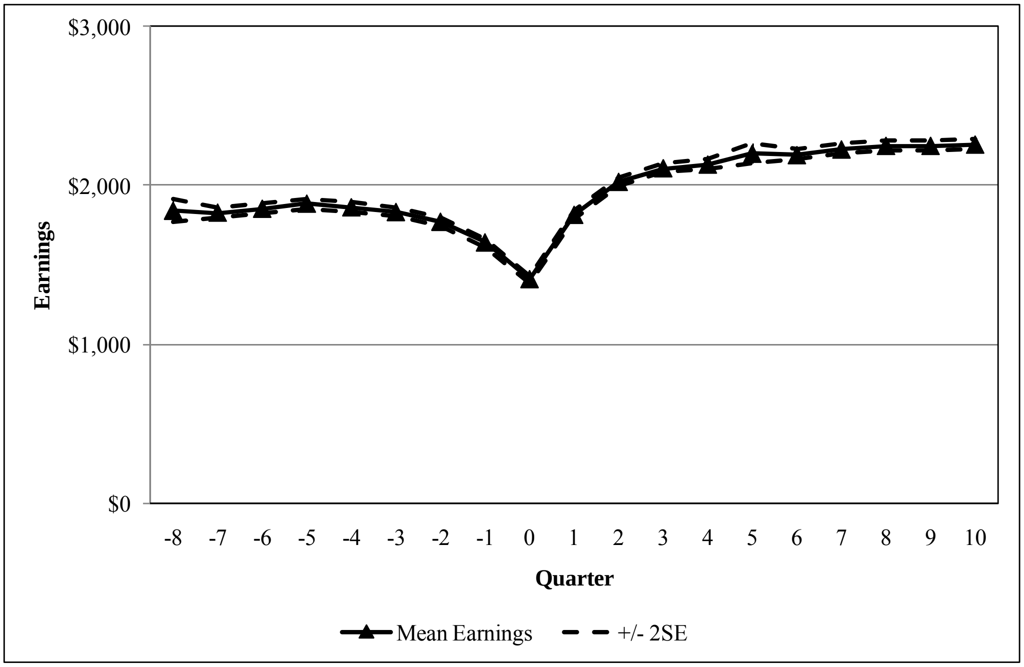


Figure ADULT EARN – A4
Mean Earnings for Adult Program Participants Receiving Core/Intensive Services: Males

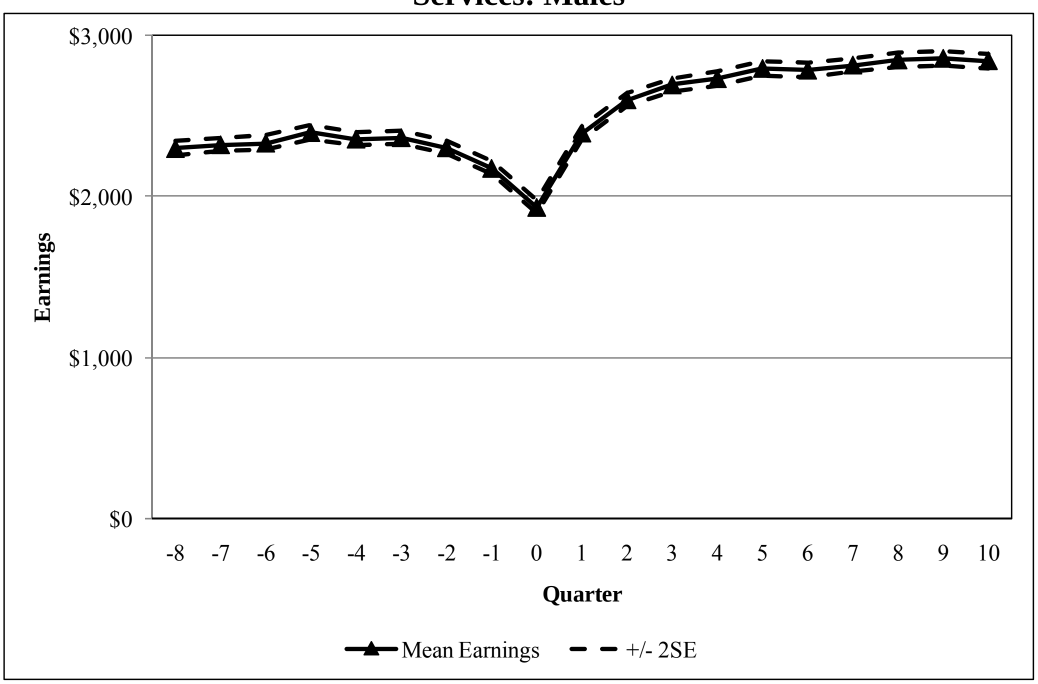


Figure ADULT EARN – A5
Mean Earnings for Adult Program Participants Receiving Training: Females

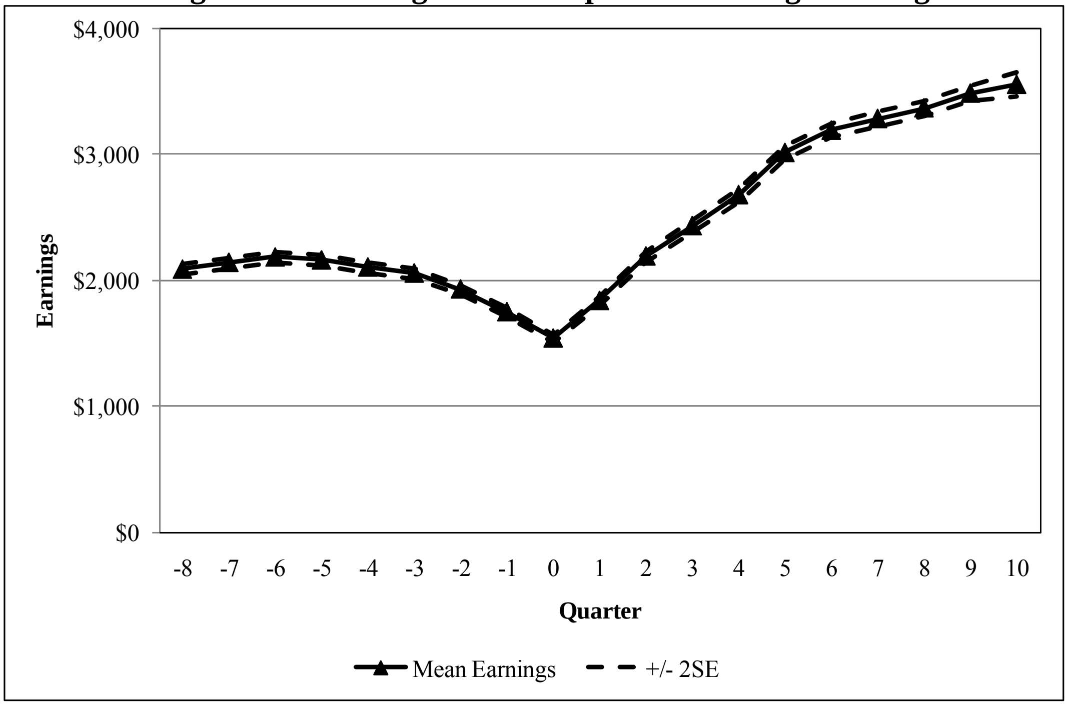


Figure ADULT EARN – A6
Mean Earnings for Adult Program Participants Receiving Training: Males

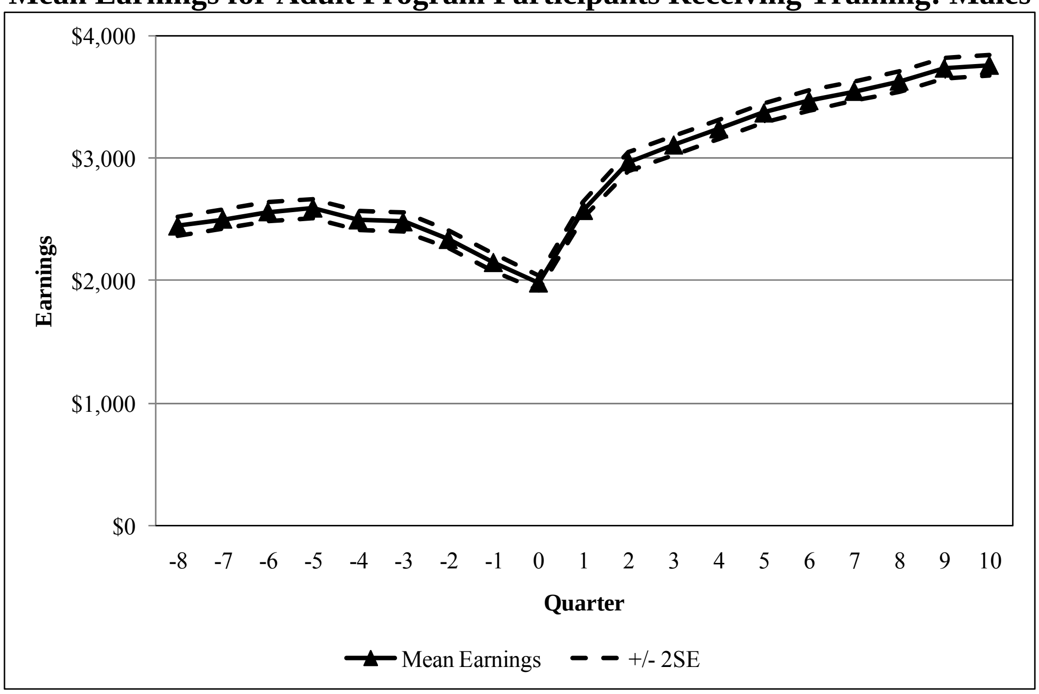


Figure DW EARN – A1
Mean Earnings for Dislocated Worker Program Participants: Females

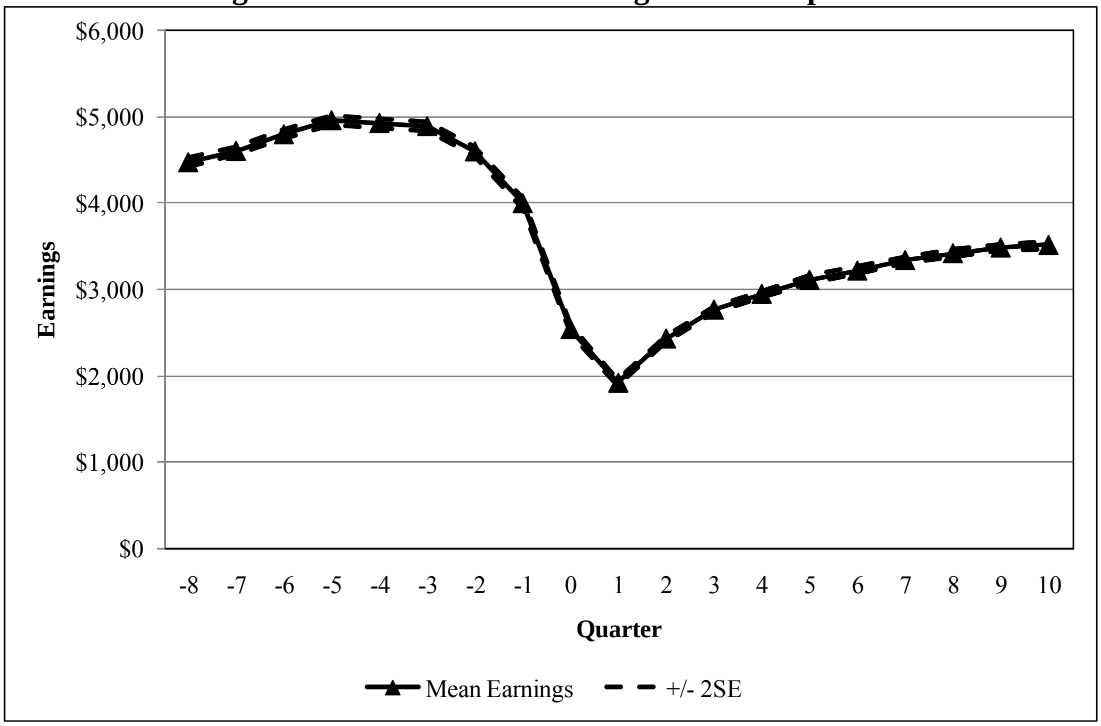


Figure DW EARN – A2
Mean Earnings for Dislocated Worker Program Participants: Males

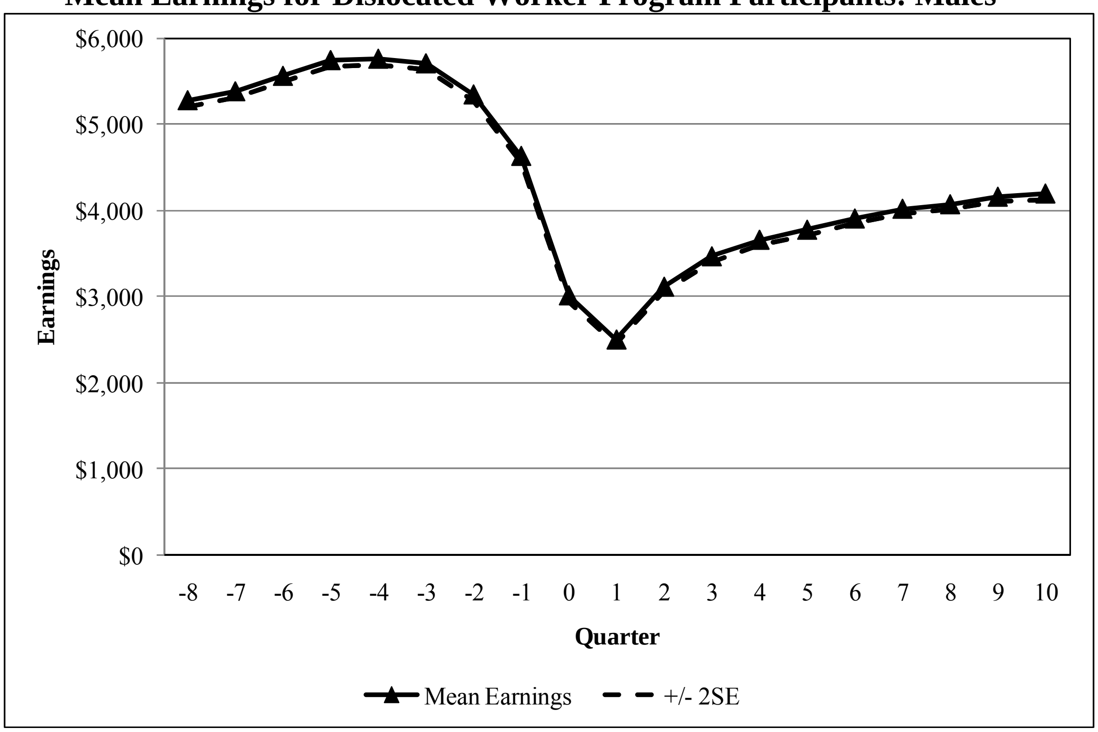


Figure DW EARN – A3
Mean Earnings for Dislocated Worker Program Participants Receiving Core/Intensive Services Only: Females

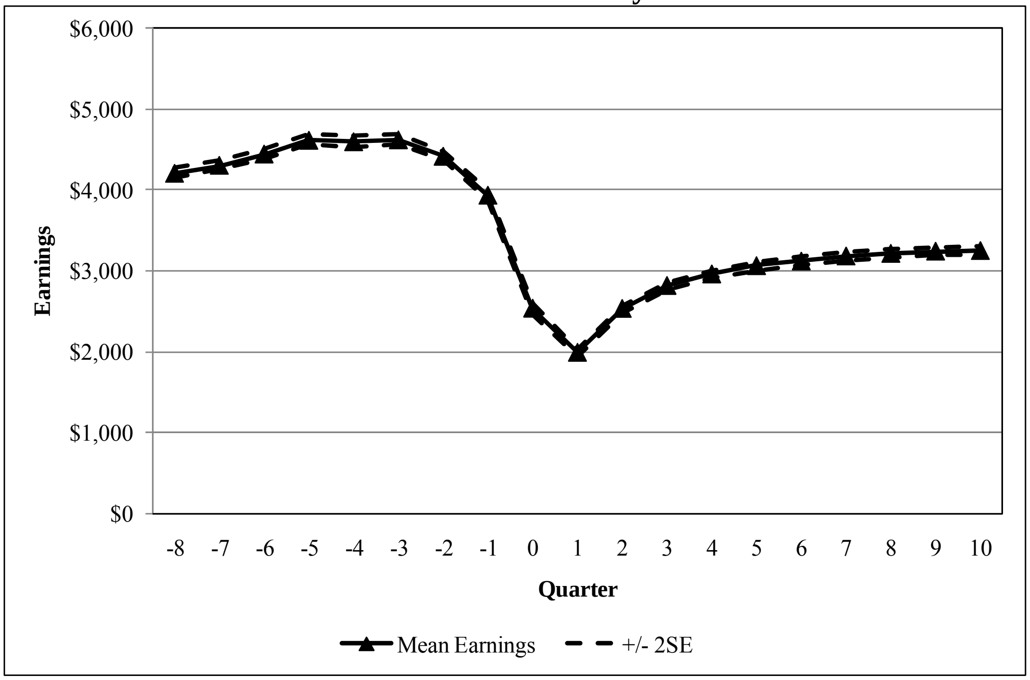


Figure DW EARN – A4
Mean Earnings for Dislocated Worker Program Participants Receiving Core/Intensive Services Only: Males

Figure DW EARN – A5
Mean Earnings for Dislocated Worker Program Participants Receiving Training: Females

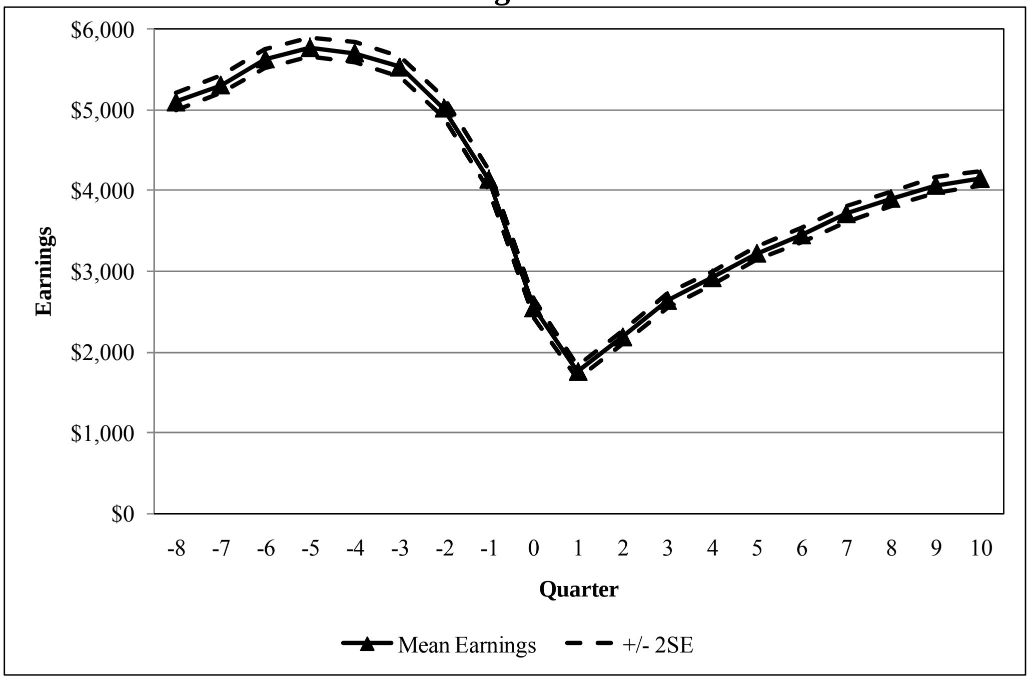


Figure DW EARN – A6
Mean Earnings for Dislocated Worker Program Participants Receiving Training: Males