AN INDEPENDENT AUDIT OF THE WASHINGTON STATE EMPLOYMENT SECURITY DEPARTMENT EVALUATION OF THE TRAINING BENEFITS PROGRAM

Kevin Hollenbeck

November 2015

Submitted to

Washington Joint Legislative Audit and Review Committee
1300 Quince St. SE,
Olympia, WA 98504-0910

By

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Ave.
Kalamazoo, MI 49007
hollenbeck@upjohn.org

The purpose of this paper is to document the activities undertaken by Upjohn Institute staff members to support JLARC's audit and oversight of the Washington Employment Security Department's (ESD) net impact evaluation of the Training Benefits (TB) program. The activities essentially were twofold: a quantitative audit of the ESD findings and a review of the draft document written by ESD staff members. The conclusions that we have drawn are as follows:

- The methods employed by the ESD staff in producing their net impact estimates are state of the art.
- In an evaluation such as the one conducted by ESD, developing the estimates of net impact requires making many methodological and data editing decisions along the way. Our quantitative audit attempted to replicate the ESD estimates plus developed net impact estimates using other methods. The results were highly stable with respect to methodology implying that the ESD estimates are statistically accurate and are not the result of a particular methodological choice.
- The ESD estimates for the 2002 and 2003 cohorts suggest that the TB program is successful in increasing participants' employment and earnings outcomes. However, the net impact of the TB program on earnings and employment are not greater than zero with statistical significance for any of the other nine cohorts. The ESD draft report suggests that these lackluster outcomes may be due to a higher incidence of occupational change among the TB participants than their comparison group members combined with the timing of the Great Recession. That hypothesis may be correct, but we think that another simpler explanation comes from the "lock in" of training combined with the Great Recession.
- A caveat to keep in mind concerning the 2002 and 2003 cohorts is that a large share of the TB participants (approximately 40 percent) returned to their employer of record. No other cohort was even close to this percentage (2005 cohort had about 30 percent, but it was a relatively small cohort). It seems safe to assume that returning to one's employer of record engenders a significant wage/earnings advantage compared to starting a new job. The benefit-cost analysis in chapter five shows that even when the most advantage assumptions are used (figure 5-5, see appendix 1), the significant net present value of the TB program to participants is due almost completely to the participants who returned to their employer of record.
- An unfortunate choice of emphasis in our opinion in chapter five is to present the social net present value and internal rate of return first, followed by the private and then government or taxpayer perspectives. The legislative audience is likely most concerned about the impact on individuals and on the state's taxpayers. The social returns are simply a summing of these two.

The next section of the paper addresses the quantitative audit that we conducted followed by a section that briefly summarizes the technical assistance and review activities that we undertook.

QUANTITATIVE AUDIT

Specifications

The ESD evaluation was complex. It essentially was 22 different studies since it estimated the net impact of the TB program on employment, earnings, and unemployment insurance (UI) benefits separately for men and women for individuals who started a spell of UI benefit payments[1] in one of the years from 2002 to 2012. Each of the 22 year-sex groups is referred to as a cohort. In the early stages of the study, we decided to limit the Upjohn Institute audit to eight of the 22 cohorts: both sexes for the years 2002, 2005, 2007, and 2012. The justification for these four years is that 2002 will give us the longest follow-up period; whereas 2012 will be the most recent data. The 2005 cohort will give us a lengthy follow-up period of eight years; and 2007 was arbitrarily chosen to give us one more set of data points. JLARC staff persons concurred with limiting the audit to these cohorts.

The ESD analyses were conducted with a quasi-experimental methodology. Other methods of analysis are possible. With the ESD analyses, participation in the TB program is the "treatment" in a (quasi-) experiment. The employment, earnings, and unemployment insurance benefits for those participants are compared to outcomes for a group of individuals who did not participate in the TB program. The latter is referred to as the comparison group. As long as the individuals in the comparison group are similar to the TB participants, then it may be reasonably assumed that the differences in outcomes can be attributed to the TB program. The method that ESD used to get the comparison group to be similar to the treatment group is statistical matching.

Figures 1 and 2 show graphically the rationale for relying on statistical matching. These figures show the unadjusted net impact of the TB program on earnings and employment[2], by follow-up year, for the 2002 (figure 1) and 2005 (figure 2) male and female cohorts. The bottom (red) line compares the mean outcome for all UI recipients, except for those in the TB program, to the mean for the TB participants. The upper (green) line compares the outcomes for the matched comparison sample only (a subset of the all UI recipients, except for those in the TB program) to the mean for the TB participants. In all of the figures, the unadjusted net impact for all non-treated individuals (referred to as the full sample) lies beneath the unadjusted net impact for the (ESD) matched comparison sample. This means that if we just relied on a comparison of the individuals who participated in the TB program vis-à-vis those who did not, we would

[1] This report only provides information about the employment and earnings outcomes. ESD staff members altered the method for estimating the net impacts of the TB program on (future) UI benefits recently. They originally had used the propensity score matching methodology with annual UI benefits as the outcome variable (not using a difference-in-difference regression), but they have changed techniques to using something referred to as mediation analysis. Given an unfamiliarity with this type of analysis, we decided not to invest the time and energy that would be necessary to audit the ESD UI estimated impacts.

[2] Mean of the outcome variable for the treatment group minus mean for a comparison sample.

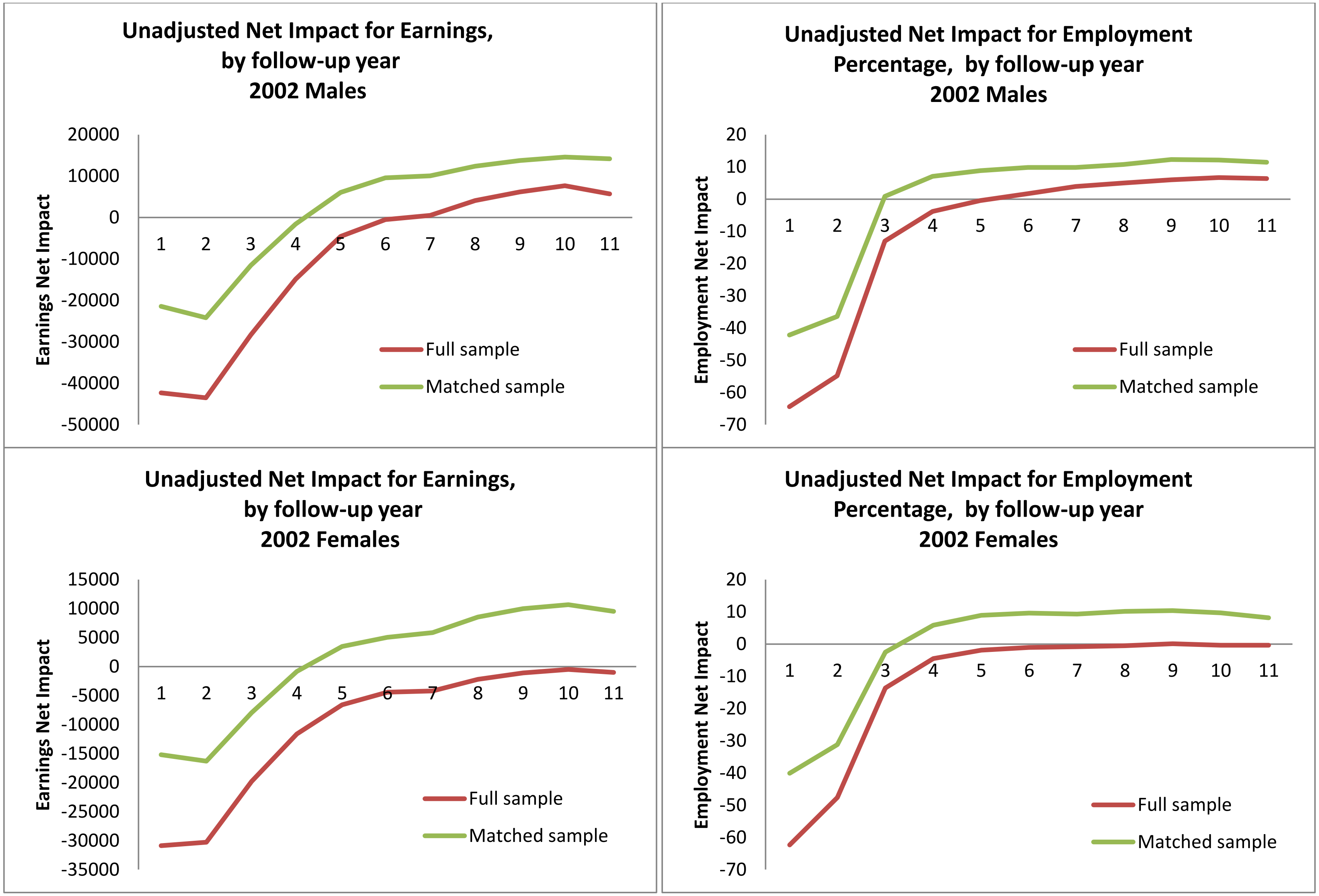


Figure 1 Unadjusted Net Impact of the TB Program on Earnings and Employment, by follow-up year, for 2002.

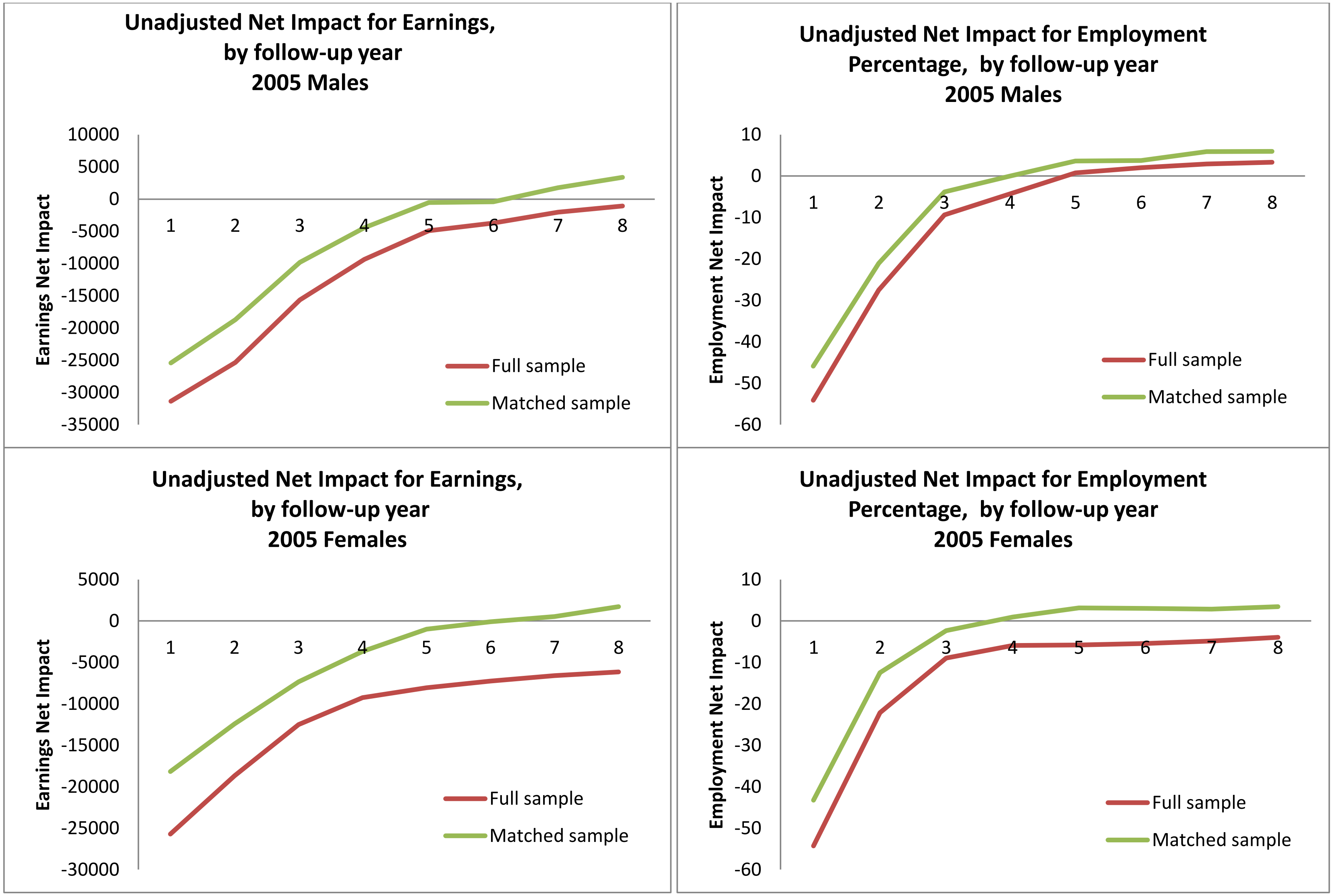


Figure 2 Unadjusted Net Impact of the TB Program on Earnings and Employment, by follow-up year, for 2005.

underestimate the impact of the TB program. This is consistent with the notion that the observable characteristics of the participants cause them to fare worse in the labor market than the average individual receiving UI benefits. Statistical matching results in having statistically identical characteristics for the treatment and comparison groups, which will help to identify the impact of the TB program.

Statistical matching can be done using several different techniques. In general, the process involves considering each of the treatment observations, and finding a non-treated observation that is very similar to it. Various options exist for “finding” the non-treated observation, and for defining “very similar.” For example, the matching can be done with or without replacement meaning that the non-treated observations may or may not be used multiple times. The matching can be done by defining the distance metric between a treatment observation and non-treated observations with several variables or with a single indicator, called a propensity score. The matching can be done by using a single observation from the non-treated group for each treatment observation (called one-to-one matching), or by using multiple observations from the non-treated group for each treatment observation (called many-to-one matching). ESD conducted their analyses with propensity score one-to-one matching without replacement.

Finally, a net impact estimator needs to be chosen once a matching technique has been adopted. A natural choice is the difference in means between the treatment and comparison groups. However, because there is a random possibility that the difference in these means is caused by the variation in the other characteristics of the observations (called covariates), we usually statistically control for the covariate distribution by using regression analysis. ESD did a regression adjustment to derive their net impact estimates.

Figures 3 and 4 show the effect of regression adjusting the net impacts. These figures show the unadjusted (labeled Matched sample in the figures) and the regression-adjusted net impacts of the TB program on earnings and employment, by follow-up year, for the 2002 (figure 3) and 2005 (figure 4) male and female cohorts. The regression-adjusted net impacts are those produced by ESD. In this case, the differences between the lines in the graphs are not as sizeable as they are in the previous figures. In figure 3, it can be seen that the regression adjustment dampens the net impacts for both the 2002 male and female cohorts. In figure 4, there are no systematic differences between the unadjusted and estimated net impacts.

One purpose of the quantitative audit was to replicate the ESD results – trying to reproduce the ESD results using the same data and same algorithms. A second purpose was to test the stability of the ESD results by using different techniques and model specifications to see whether the results change. Arguably, we would have less confidence in the ESD results if we found that making minor changes in a parameter in the matching methodology or in the model specification resulted in large fluctuations in the results.

We ended up specifying nine alternative methods for estimating the net impacts of the TB program on employment, earnings, and UI benefits for each of the cohorts that we have tested. In all nine alternatives, the raw data that were used differed slightly from the raw data used by ESD.

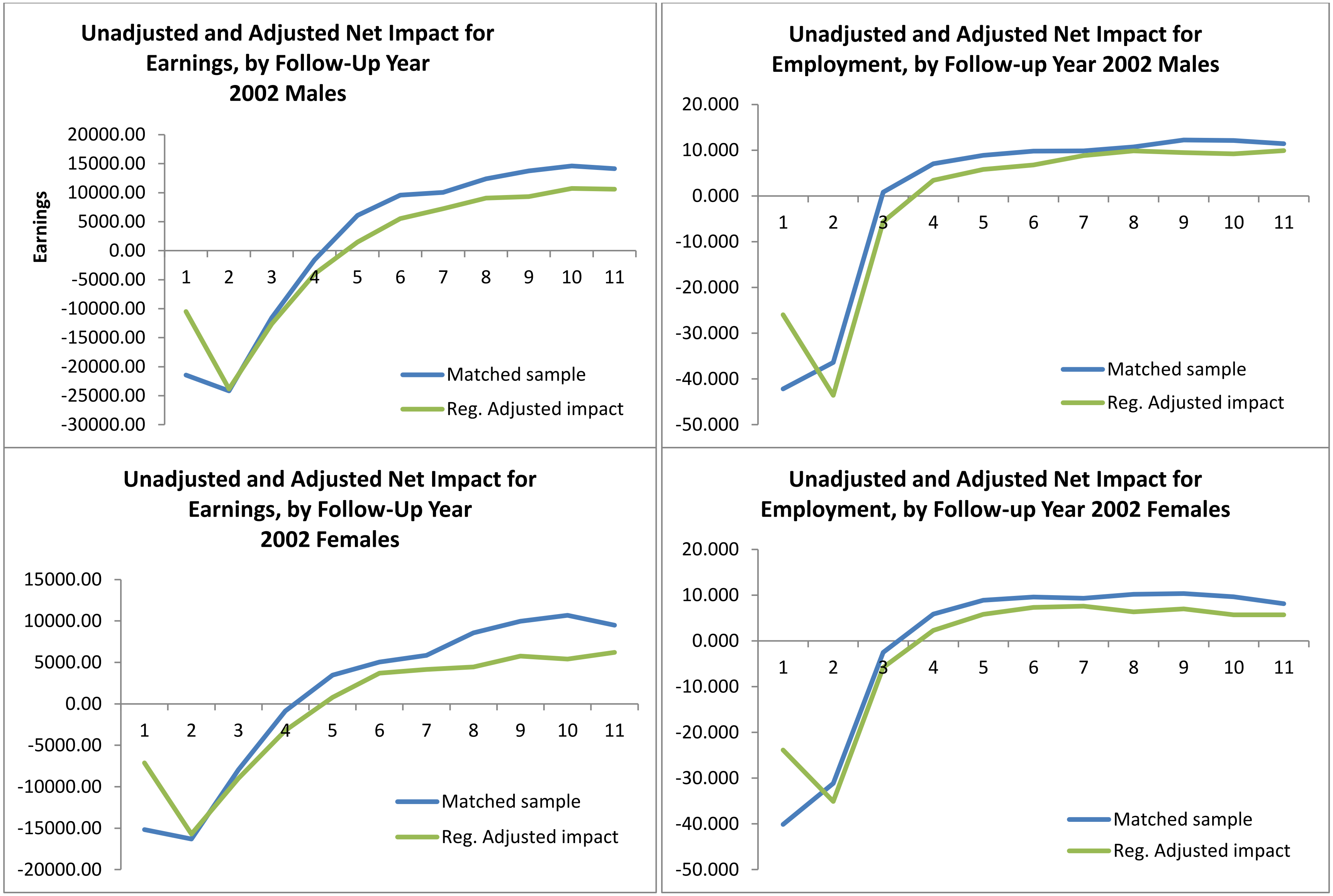


Figure 3 Unadjusted and the Regression-adjusted Net Impact of the TB Program on Earnings and Employment, by Follow-up Year, for 2002.

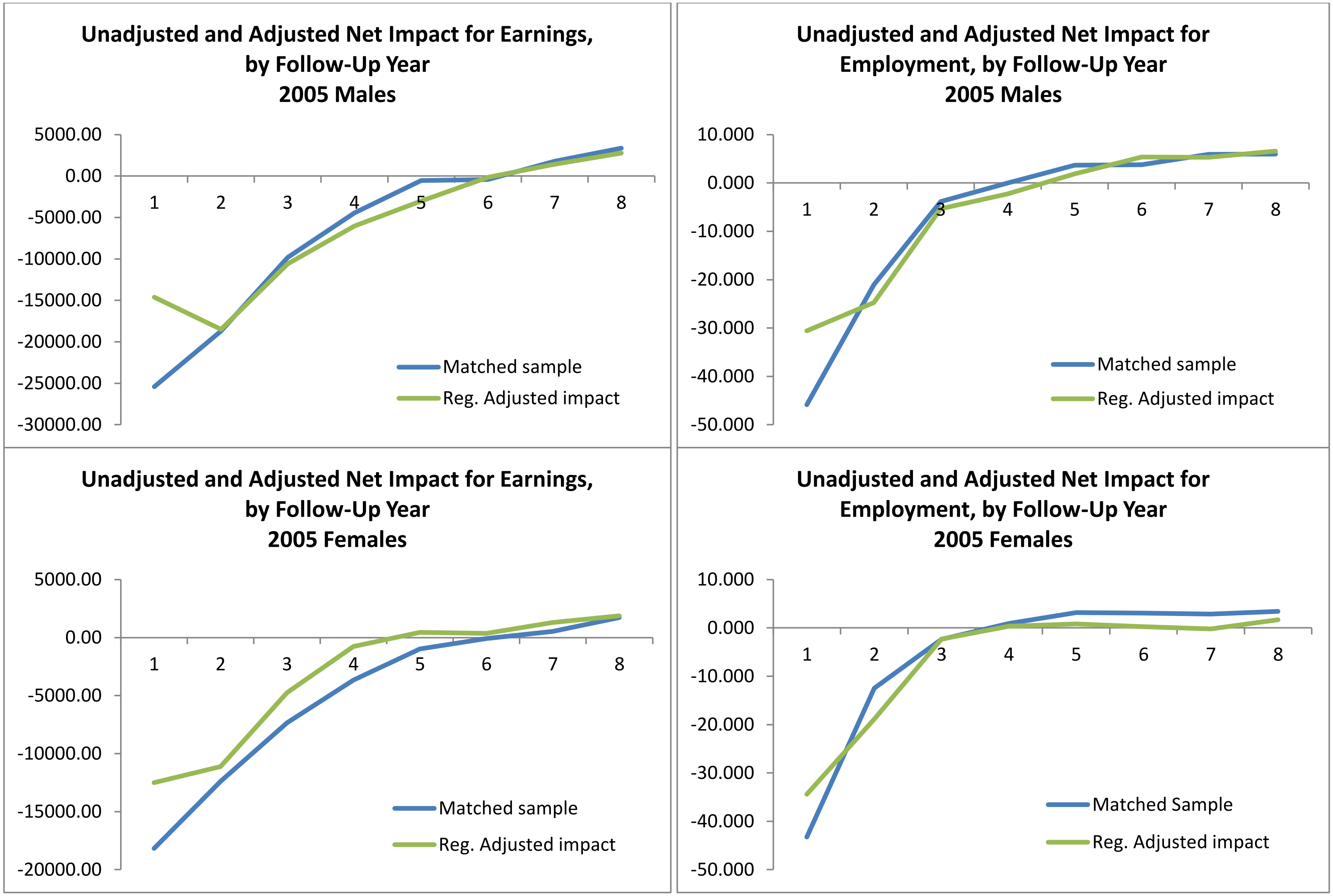


Figure 4 Unadjusted and the Regression-adjusted Net Impact of the TB Program on Earnings and Employment, by Follow-up Year, for 2005.

Table 1 Alternative Specifications Used in the Audit of the ESD Net Impact Evaluation of the Training Benefits Program

Acronym	Propensity score specification[a]	Matching method[b]	Outcome regression specification[c]
ESD-ESD-ESD replicate	ESD	ESD	ESD
ESD-ESD-UPJ	ESD	ESD	Upjohn
UPJ-ESD-UPJ	Upjohn	ESD	Upjohn
ESD-NNWR-ESD	ESD	Nearest neighbor, 1-to-1, with replacement	ESD
ESD-NNNR-ESD	ESD	Nearest neighbor, 1-to-1, no replacement	ESD
ESDUI-ESD-ESD	ESD + prior UI benefits	ESD	ESD
ESD-Block	ESD	Block Matching	--
ESD-IPWESD	ESD	Inverse propensity score weighted regression	ESD
UPJ-IPWESD	Upjohn	Inverse propensity score weighted regression	ESD

[a]ESD and Upjohn both use age, age squared, pre-program education, pre-program WDA, ethnicity/race, low-income earner status, veteran status, disability status, and pre-program occupation. The specifications differ in how pre-program employment and earnings and Ashenfelter dip variables are specified. In ESDUI-ESD-ESD, pre-program unemployment insurance benefits were added to the propensity score logit specification.

[b]ESD matching method is 1:1 match without replacement using propensity score values; first match is at 6 digits (5 decimals); second at 5 digits; and so forth to 2 digits (1 decimal). Nearest neighbor matching minimizes distance defined as the absolute value of the difference in propensity scores. Block matching combines the treatment and entire set of non-treated individuals into optimal number of blocks. In each block, the mean propensity score is not different among treated and non-treated cases. Inverse propensity score weighted regression is not a matching method. It is a weighted OLS regression that includes a treatment dummy among the regressors.

[c]For all specifications, the outcomes for employment and earnings are estimated with a difference-in-differences specification and the unemployment insurance outcome is based on levels. The Upjohn specification is identical to the Upjohn propensity score specification. The ESD specification differs slightly for each outcome: employment, earnings, and unemployment insurance. In all three outcomes, the ESD specification includes age, age squared, pre-program education, pre-program WDA, pre-training quarterly earnings, ethnicity/race, veteran status, disability status, and low income earner status. For ever employed, the ESD specification includes number of working to not working transitions in the pre-program period. For annual earnings, the ESD specification includes pre-training industry. For unemployment insurance benefits, the ESD specification includes Ashenfelter dip variables; pre-program earnings, occupation, union benefit claimant status, and industry; and post-program annual earnings.

In particular, we top-coded quarterly earnings at $99,999, and we corrected an error in the disability code for one of the cohorts.[3]

The nine alternative methods that we used to test the ESD analyses are enumerated in table 1. All nine of the methods required the estimation of a propensity score. Essentially this means that we have estimated a statistical model of applying for and being accepted into the TB program using the observed characteristics of each observation. The data base that we used is the universe of individuals who receive UI benefits in the cohort year. The dependent variable in this model is a 1 if the individual participated in the TB program, and 0 otherwise. The propensity score is the predicted value for an observation; in other words, it is the probability that the individual participated in the TB program. For each observation, it takes on a value between 0 and 1.

[3] In the 2005 cohorts, the disability variable in data supplied by ESD had the value 9 for several observations, which is an invalid code. Only 0 and 1 are valid codes. We changed the 9's to 0's, whereas ESD used the 9's in their analyses.

ESD analyses and the first seven alternatives in Table 1 used statistical matching. In these analyses, three different matching algorithms were used. These algorithms are documented in note b of the table. Suffice it to say that the ESD and nearest neighbor algorithms are very similar. Of the first 6 algorithms listed in the table, there is only one that would seem to be different from the ESD analyses in a consequential way. That potential outlier is the fourth one listed, in which the statistical matching was done with replacement. The results of allowing observations from the comparison group pool to be used more than once are that a better match is obtained at a cost of biased standard errors due to the clustering of data. Block matching is an entirely different algorithm. In this case, the entire set of non-participants and the treatment group is divided into mutually exclusive groups using the propensity score. So for example, observations with a propensity score between 0.08 and 0.095 might be one group, between 0.095 and 0.115 another group, and so forth. The groups are established so that there is statistical balance between the treatment cases and the non-participant cases in each of the groups. The net impact estimate is then a weighted average of the unadjusted mean differences for each of the groups.

The last two entries in table 1 do not rely on statistical matching at all. These methods use weighted least squares regression estimates. The equations that get estimated have the outcome variables as dependent variables and include a treatment dummy (takes on a value of 1 for treatment observations and 0 for non-treatment observations) as an independent variable. All other independent variables are the same as those used in the ESD regression-adjustment of outcomes model. The coefficient on the treatment dummy is the net impact estimator. The estimation techniques use differential weights so that observations with relatively high propensity scores are weighted more heavily than observations with relatively low scores[4]. (If an observation is more likely to be a participant or have similar characteristics to participants, it "counts" more than if the observation is unlikely to be a participant or have characteristics that are different from participants'.)

The last column of Table 1 differentiates between two different specifications of the outcome regression equation. As seen in the table, we most often relied on the ESD specification.

[4] The weight that gets used is p / (1 – p), where p is the propensity score. Note that if p is relatively high, say 0.70, then the weight is relatively high: 2.333. If p is relatively low, say 0.05, then the weight is low: 0.053.

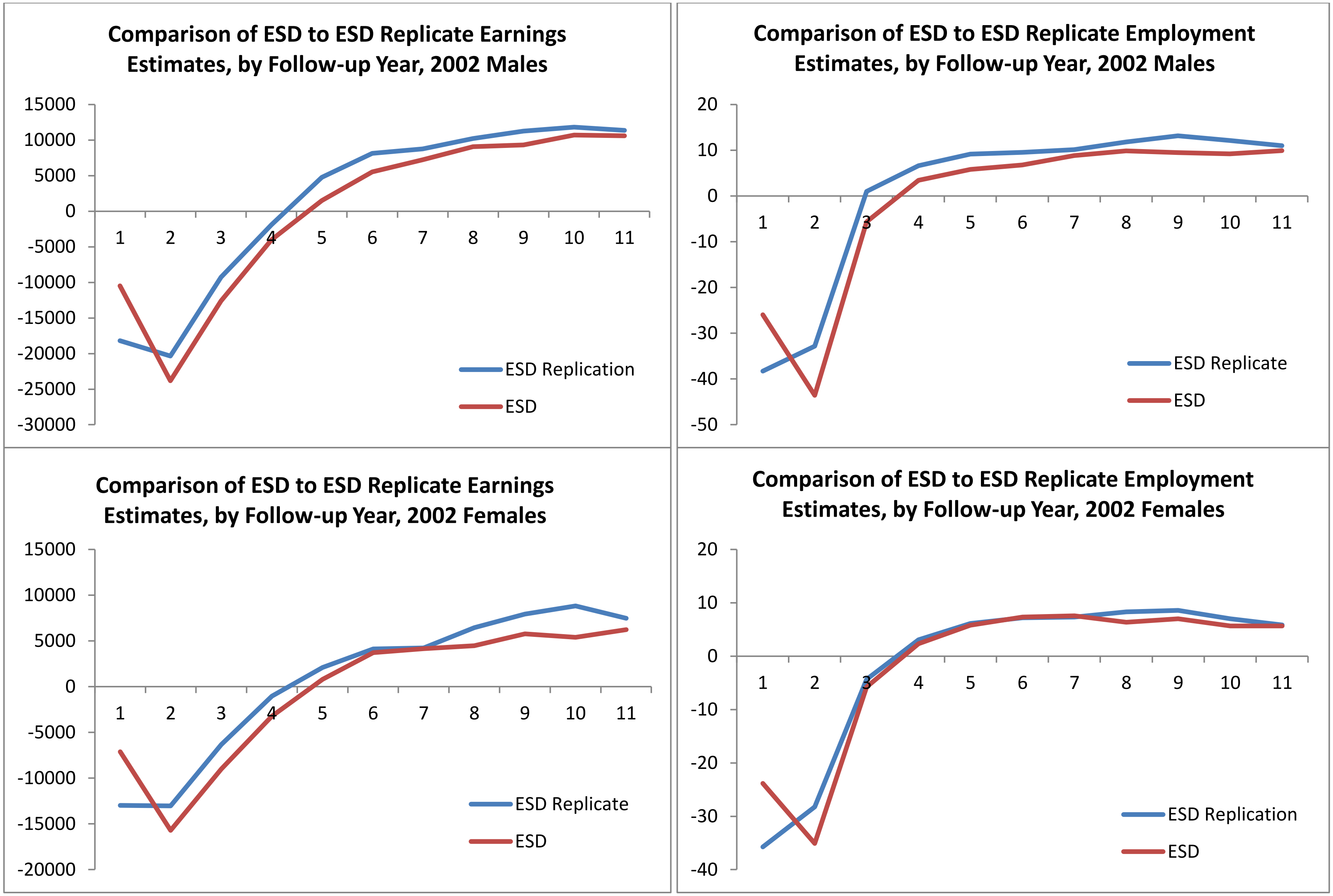


Figure 5 Comparison of ESD to ESD Replication, by Follow-up Year, for 2002.

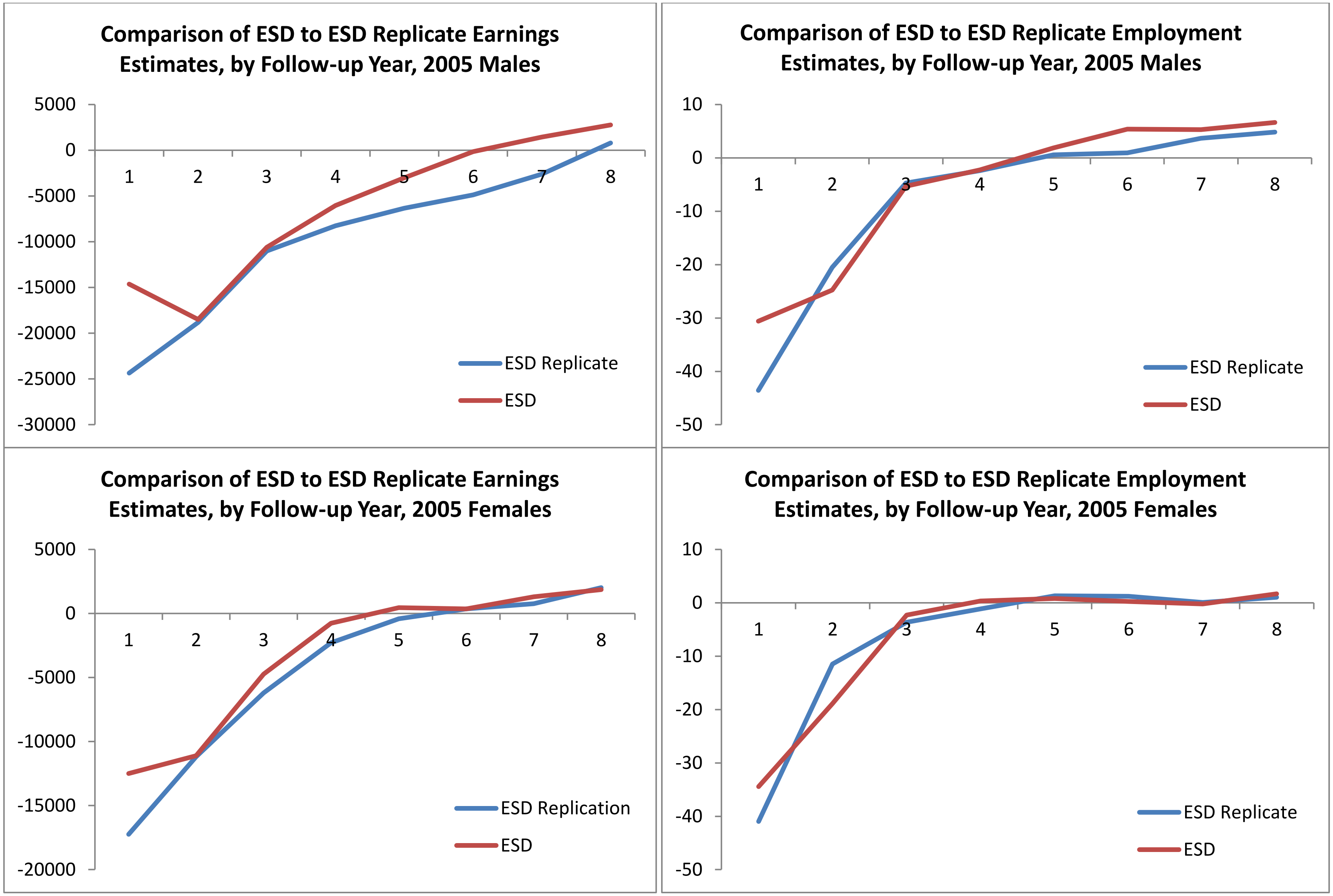


Figure 6 Comparison of ESD to ESD Replication, by Follow-up Year, for 2005.

Results

In this section of the report, we will summarize a few of the estimates that we produced. All of the results are available upon request. In figures 5 and 6, we show the results of attempting to replicate the ESD methodology. As far as we know, the only differences between the estimates should be caused by our top coding of earnings and our correction of the disability variable. With the exception of the first follow-up year, the two sets of net impact estimates do line up reasonably closely. However, the earnings impacts for 2002 females in the replication analyses are somewhat larger that the ESD estimates, and vice versa for the earnings impacts in follow-up years 4 to 8 for the 2005 male cohort. The ESD net impact estimates for the first follow-up year are quite different from our estimates, but otherwise, we are generally satisfied that we have closely replicated the ESD results.[5]

In figures 7–9, we show the ESD net impact estimates for employment and earnings with two alternative net impact estimates for the 2002, 2005, and 2007 male and female cohorts. The alternatives that are included are the block-matching estimates (labeled block in the figures) and the nearest neighbor propensity score matching with replacement estimates (labeled nnwr). The techniques used to generate these estimates are ones that we have used in prior work and would be ones that we would likely use to evaluate the TB program if we were to do so independently.

There are several interesting systematic results that are being displayed in the figures. First is the fact that the line graphs are very similar to each other. It turns out that all nine alternative estimation techniques that were used produced very similar results. Line graphs that include estimates from all nine techniques are virtually collinear. This is somewhat surprising since the underlying techniques are quite varied. Some of those techniques involve propensity score matching, and some of the techniques are regression estimates using the full sample. Some of the matching techniques use a comparison group and some use the whole comparison group pool.

A second general result shown in the figures is that the net impact estimates tend to be curvilinear. The horizontal axis at the value of 0 represents no difference between the treatment group and the comparison group, or in other words, a 0 net impact. At points above the axis, the net impact estimates are positive (TB participants have better outcomes than non-participants); at points below the axis, the net impact estimates are negative. In all three years of cohorts, the net impact estimates start below the 0-axis and then accelerate toward it. Results for the 2002 and 2005 cohorts actually cross the axis and become positive (not significant for the 2005 cohorts); however the results for the 2007 cohorts do not.

[5] The ESD net impact estimates for the first follow-up year seem quite anomalous. For the employment and earnings outcomes, the line graphs show that the ESD estimates are much less negative than our replication, and the ESD time series shows sharp declines to the 2nd follow-up year. Since the general explanation for negative results in the first follow-up years is the "lock-in" of training (participants are in training and therefore are delaying employment), one would expect the first follow-up year to have more negative outcomes since some of the TB participants will be in short-term training that will have ended in the first year – and "lock-in" would be absent for these individuals. This upward bias would tend to bias the benefit-cost findings as well. Note that the first follow-up year outcomes "look" anomalous in the log extrapolations graphically displayed in figures 5-2 – 5-4 as well.

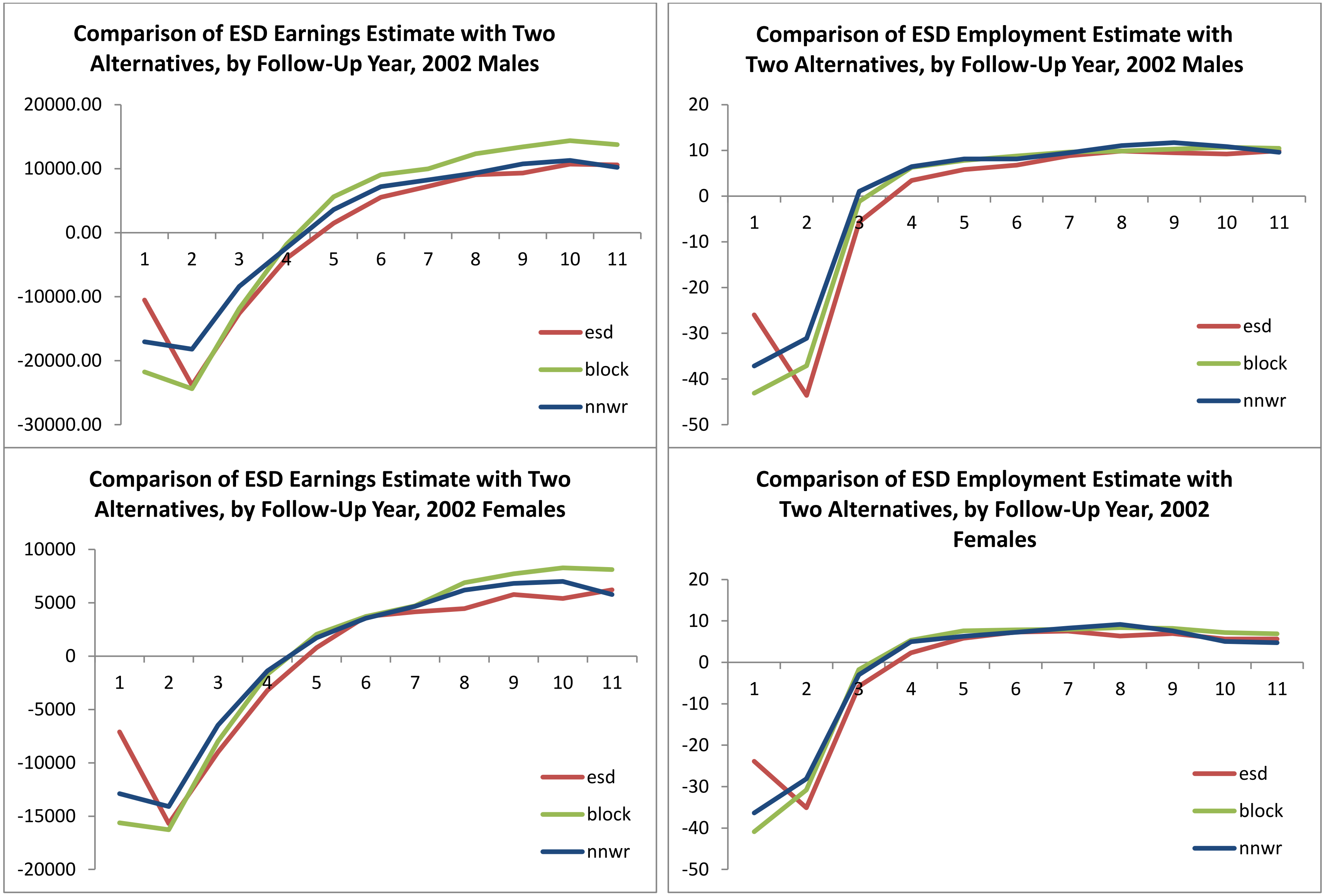


Figure 7 ESD Net Impact Estimates for Employment and Earnings with Two Alternatives, for 2002.

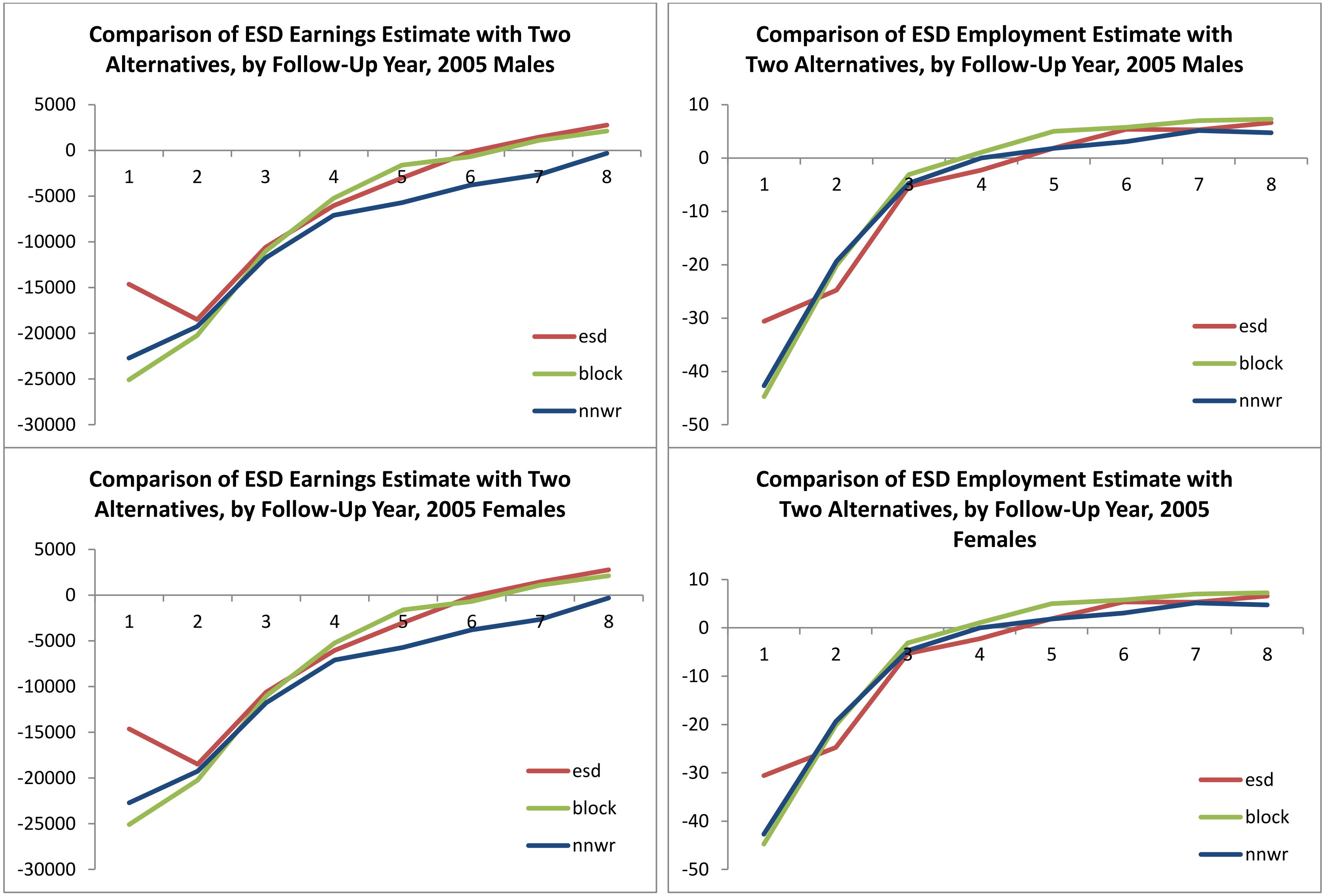


Figure 8 ESD Net Impact Estimates for Employment and Earnings with Two Alternatives, for 2005.

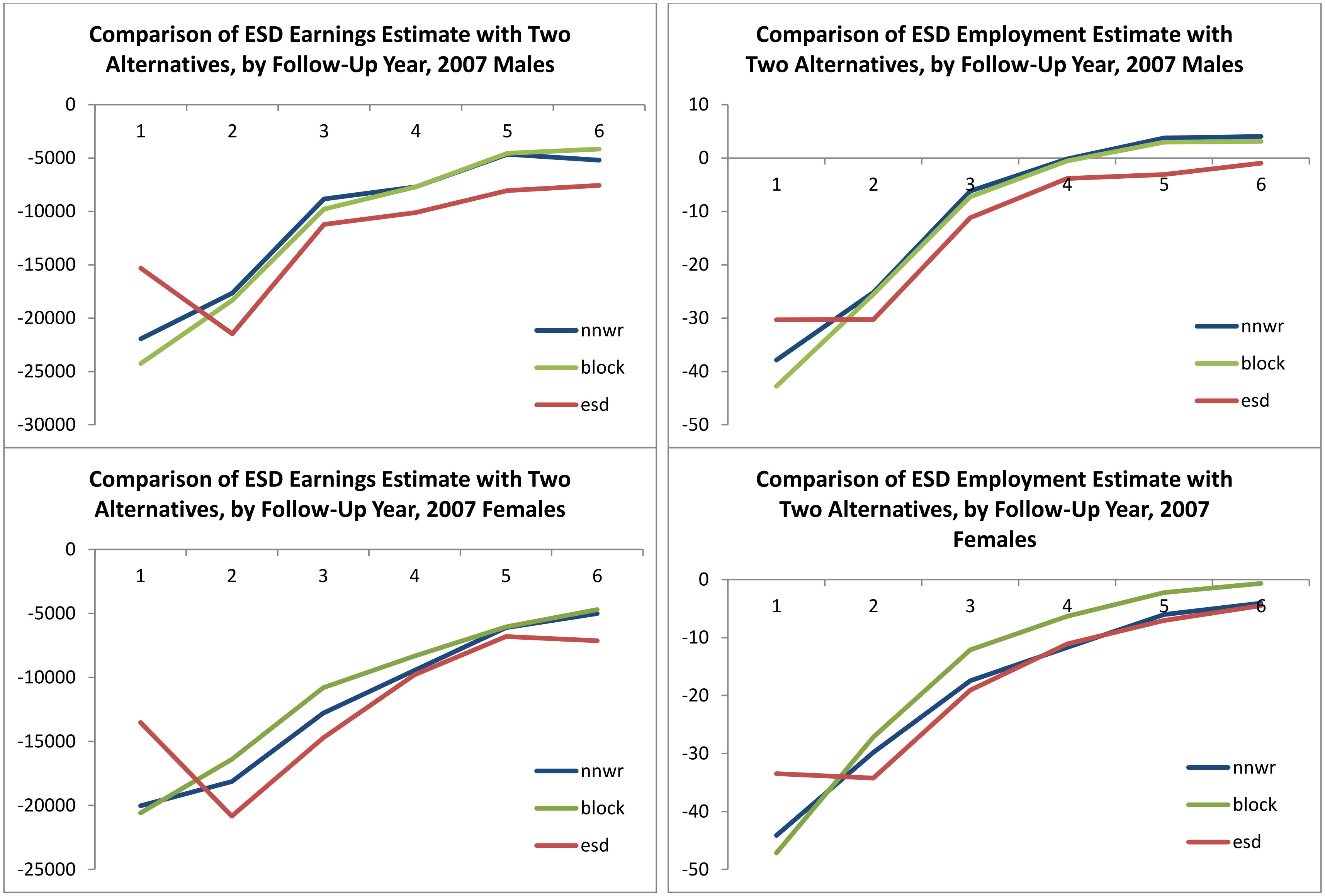


Figure 9 ESD Net Impact Estimates for Employment and Earnings with Two Alternatives, for 2007.

A third general result is that the employment results seem to "flatten out" after the third follow-up year. Both of the 2002 cohorts seem to have a positive employment rate impact of about 10 percent starting in the 4th follow-up year; both of the 2005 cohorts seem to have a positive employment rate impact of about five percent starting in the 5th follow-up year; and the 2007 cohorts' leveling off is at or near 0.

Fourth, the recovery of the treatment group seems to be faster for employment than earnings, especially for the 2002 and 2005 cohorts. For both males and females, the employment rate impacts in these two cohorts hits approximately 0 in the third follow-up year and then become positive afterward. However, the earnings net impacts hit 0 approximately a year later.

Finally, for some inexplicable reason, the ESD net impacts for the 1st follow-up year are all much less negative than the alternative estimated net impacts that we calculated.

Interpretation

Having a very stable pattern of results for all of the alternative estimates has two implications. First, the results are very robust, which we believe gives them more credibility; we can have more confidence in them. Quite frankly, we are surprised by how stable the estimates are. Our prior assumption was that there was going to be more variability, and that we would end up arguing about whether to use propensity score or block matching, to match with or without replacement, to use a one-to-one or many-to-one matching, or even to do matching versus regression. In our test cases, these questions are moot because the pattern and level of results are quite stable. The second implication was noted above – there is absolutely no evidence to suggest that ESD biased the results in any direction.

As we have discussed before, the pattern of net impact results, which is consistent across cohorts (and estimation methods), is one of negative impacts for the first few follow-up years often followed by impacts that are approximately zero and then turn positive. This is quite consistent with the phenomenon called "lock-in." Individuals pursuing training or education are "locked" into that status until the training is over, so they are unlikely to become employed and have earnings. The facts that the 2005 cohorts just barely get into positive territory and the 2007 cohorts never become positive even after 6 years is concerning, however.

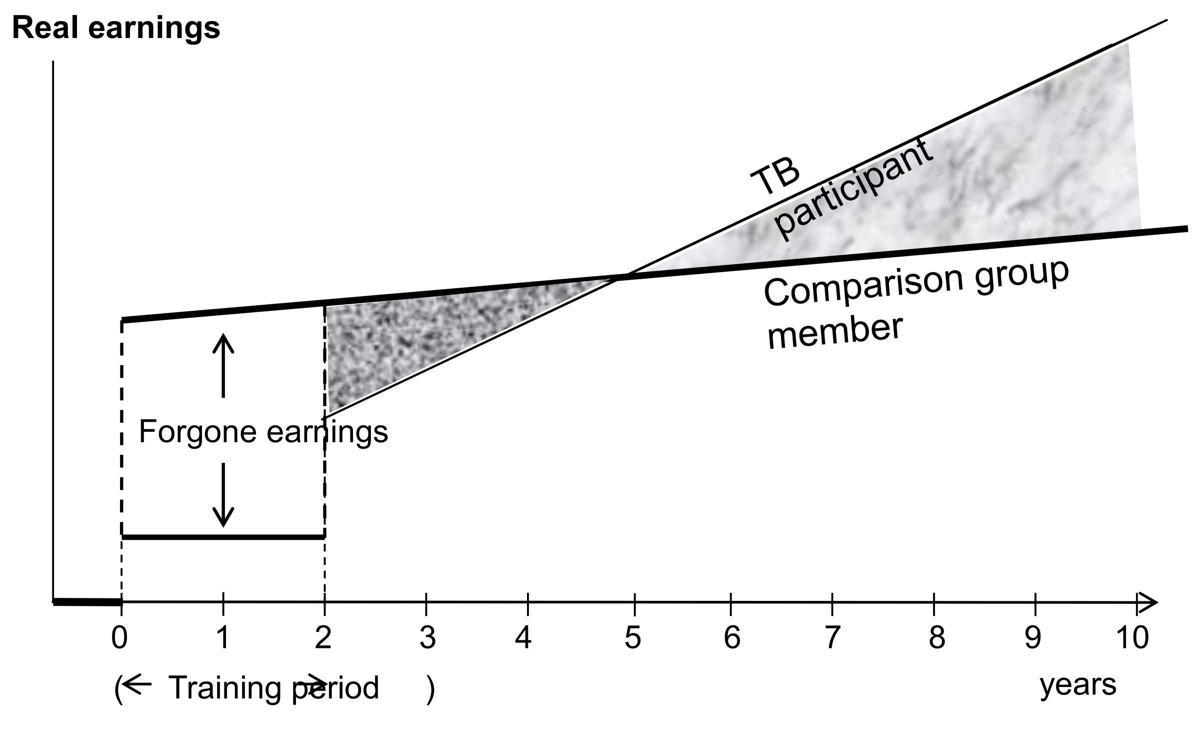
The fact that employment impacts are "stronger" than earnings impacts, i.e., employment rates hit the 0-axis sooner than earnings, suggests that the TB program is resulting in employment opportunities, but in the early stages of employment, either wages or hours are not as high, on average, as the comparison cases. This might be expected if the treatment observations are changing occupations relatively more than the comparison cases.

Can occupational changes explain the lackluster outcomes of the TB program?

Even though chapters 2 and 3 of the ESD draft report discuss how the employment impact and the earnings impact, respectively, become positive by the 4th or 5th follow-up year, these impacts are not statistically significant for all but the 2002 and 2003 cohorts. If one sticks to a conventional test of hypothesis that an impact is not nonzero unless the p-value is less than 0.05 (or even if you use a looser test using a p-value of 0.10), then figure 2-3 in the ESD report indicates that the 2002 and 2003 male cohorts attain a positive employment impact in the 4th or 5th follow-up year, and none of the other cohorts ever achieve a positive net impact. The same is true for females as shown in figure 2-4 in the ESD report. The story is the same for the pre-tax earnings net impacts. In figures 3-3 and 3-4 in the ESD report, the impacts do not turn positive (using p-values of 0.05 or 0.10) for the 2002 and 2003 cohorts until the 6th follow-up year, and never turn non-negative for any of the other cohorts.

In several chapters of the ESD draft report, the suggestion is put forward that these outcomes result from the fact the TB participants are more likely to be changing occupations than members of the comparison group, and the cohorts after 2003 had to re-enter the labor market just as the Great Recession and its post-recession "softness" occurred. Having less job experience and its occupational human capital in a new occupation caused the TB participants to be at a disadvantage.[6] This is definitely a sensible and plausible hypothesis and is empirically consistent with the findings that TB participants who returned to their prior employer have better outcomes than those who do not, since it is less likely that these participants changed occupations.

In my opinion, there may be a different, simpler explanation for the non-positive results. That explanation is basically what has been referred to in this paper as "lock-in." Figure 10 (often referred to as a Mincer diagram) exhibits a classic theoretical framework for the impact of training. The vertical axis in the figure is annual earnings (assumed to be adjusted for inflation



[6] The ESD draft report also suggests that re-entering the ranks of the employed in a new occupation may result in more part-time employment. If true, this would help to explain the lags in the earnings net impact, but would not help to explain the employment net impacts.

Figure 10 Theoretical Payoff to TB Training.

and also assumed to be equal to worker productivity). The horizontal axis is time, measured in years after the start of training. We have arbitrarily set time = 0 when the TB participant starts training, and the individual that is matched to the TB participant starts employment.[7] The bolded line represents the (real) earnings of the comparison group member. It is upward-sloping because workers generally become more productive over time.

The TB participant's earnings are set at a very low level during training to reflect part-time or part-year employment. When the training ends, the participant, who now has new or upgraded skills, begins to look for full-time employment but two to three years after the comparison group member has begun full-time employment. That is the "lock in" phenomenon.

The TB participant's earnings trend upward, but they start at a point that is lower than the comparison group member because that comparison group member has been working. However, because the training has improved the participant's skills and knowledge, his/her productivity should eventually surpass the comparison group member. In the figure, this "crossover" occurs in follow-up year 5.

Unfortunately for the 2004 to 2007 cohorts, the recession hit the Washington labor market at the time when the TB participants would have been entering employment. So these participants were "locked into" training and "locked out" of employment, and then when they were ready to enter employment, they met the headwinds of the recession. Thus the non-positive impacts for these cohorts are not surprising because the participants are being compared, for the most part, to individuals who were employed prior to the recession.

Benefit-Cost Analysis

The last chapter of the ESD draft report presents a benefit-cost analysis (BCA) and rate of return analysis of the TB program. In my own work, I use a slightly different framework from the one used in that chapter because I think that the perspective of employers should be explicitly presented. The *raison d'etre* for training is to improve the trainee's skills so that he/she may become productive, and employers are the beneficiaries of the enhanced productivity. It is unfortunate that employers get lumped in with taxpayers and others in the framework presented in chapter 5, although because productivity is difficult to value, that may be the best that can be done.

Figure 11 presents, in tabular form, the components of a full BCA for the TB program in a manner that differs from figure 5-1 in the ESD report. The final row displays the net benefits to each of the parties and is derived by summing the columns. The final column of the table represents the total net benefits to society and is derived by summing across the rows. This column represents the social return noted in the ESD analysis. The entries in the table represent the expected costs (−) or benefits (+) to the group.

[7] Obviously, the events will not happen at exactly the same time, but on average, the comparison group member will become employed a few weeks before or a few weeks after the participant starts training.

Program costs are in the first and second row. In the TB program, fees and tuition are paid by participants, who are also investing their time and effort. Furthermore, as shown in figure 10, participants have substantial forgone earnings. All in all, TB program participation is a cost for participants. Employers do not bear any of the direct cost of the TB program, whereas taxpayers and others bear the administrative costs of the TB program and may bear tuition subsidy costs, if the training undertaken increases the number of postsecondary students in the state. The major public cost of the program is shown in the second row – the extended weeks of UI benefits. This is clearly a benefit for the participants and a cost for taxpayers who bear the costs of the extra weeks of unemployment compensation. We assume that the extended weeks of UI do not alter the experience rating of the employers, and thus there is no cost for them.

Rows 3–8 of the table represent potential benefits from program participation. The new or upgraded skills that TB program participants gain are intended to lead to job placement. When individuals become employed, they become productive members of the workforce. In row 3, we show that employers benefit because they are able to sell more and higher-quality goods and services, and society benefits from the availability of the additional goods and services.

Benefit or cost	Participants	Employers	Taxpayers and rest of society	All
1. Training costs	−	0	−	−
2. Unemployment insurance benefit extension	+	0	−	0
3. Productivity of individuals who become employed	0	+	+	+
4. Higher earnings	+	−	0/+	0/+
5. Fringe benefits	+	−	0	+/0/–
6. Less unemployment/ lower turnover	−	+	0	0
7. Lower social safety net program costs	−	0	+	0
8. Higher taxes	−	–	+	0
9. Net benefits	+	+	+/0/–	+

Figure 11 Components of a Benefit Cost Analysis.

Row 4 shows that trained workers receive higher earnings (through increased employment, wages, and hours). Those earnings are a cost to employers. We add a potential benefit for the rest of society in this row because of the multiplier effect that program participants' higher earnings may engender.

The fifth row shows that program participants who become employed will typically receive fringe benefits over and above their earnings. We indicate that the additional fringe benefits may be a net benefit or cost to society depending on whether or not workers value the fringe benefits more or less than what employers pay for them. In the sixth row, we show reduced levels of unemployment and turnover due to skills learned. This is actually presumed to be a cost to program participants because they are losing nonwork or leisure time plus they may be losing future unemployment compensation benefits. The reduction in unemployment and turnover is a benefit to employers because they will have lower hiring costs and future unemployment compensation payments.

The seventh row indicates that participants are less likely to receive social safety net program benefits because of higher earnings and employment. This is a cost to them, but a gain to the rest of society. On net, the benefit is zero because these payments are transfers from the rest of society to recipients. With higher levels of earnings and employment come higher tax liabilities. These are denoted in row 8. Workers and employers will pay higher payroll taxes. The rest of society benefits because presumably the government will spend the money on social benefits or cut taxes.

The bottom row shows that the net benefits to TB participants are positive (or they would not participate). Their increased earnings (net of taxes) will exceed their time and financial costs, if any, and reduced transfer income. We would expect the net benefit to employers to be positive also. Employers' compensation costs will be less than the productivity of the worker (that is the profit margin of the employer) plus they are likely to have lower turnover costs.

The net benefit to taxpayers and the rest of society is indeterminate. This sector of the economy bears the costs of the extension of UI weeks and the administration of the TB program, and its major return will take the form of additional goods and services available in the economy, higher levels of government spending/lower taxes, and potentially lower future social safety net expenses. Adding across the bottom row, it is likely to be the case that the social net benefits are positive. In the empirical implementation of the TB program's BCA, the main "drivers" of the results are likely to be the plus sign in the fourth row for participants and the minus sign in the second row for taxpayers and the rest of society.

The framework that ESD employs in its benefit-cost analysis in chapter five is comparable to figure 11 above. The ESD framework is displayed in figure 5-1 in ESD's report. The first row splits the participant output into three components: In-program output, Before-tax earnings, and Fringe benefits. The latter two are rows four and five in figure 11. The in-program output is not represented in figure 11. If trainees were productive during training, then in figure 11, the 0 in the Employer column would become a +. The point is moot, however, since ESD does not include this component in their analyses.

The second row of figure 5-1 is identical to the 8th row of figure 11. The third row splits the direct participant costs into three components: Tuition and fees; Books, transportation, clothing, tools, supplies; and Forgone earnings. These components comprise the cost that is in the participant column in the first row of figure 11. The ESD figure includes three rows (5, 6, and 7) to represent scholarships, grants, or loans that TB participants might receive to help finance the TB training. ESD does not attempt to estimate these financial aids, but I think that they should be included in the direct cost of the TB training, so that 3a and 3b in figure 5-1 become net costs. Furthermore, in line 6, somebody is providing the scholarships or grants, so there has to be a cost in the Society column, and in line 7, since student loans bear interest, the government or taxpayer is getting a benefit if the loans are public, or somebody in society (e.g., banks) is getting a benefit if they are private. The fourth line in figure 5-1 exhibits Government/taxpayer subsidies to the direct cost of training. This would be the cost that is displayed in the first row of figure 11 in the Taxpayer and Rest of Society column.

The administrative cost of managing the TB program, in column 8 of figure 5-1, is included in the first row of figure 11. In figure 5-1, line 9 splits transfer payments into UI benefit payments and all other transfer programs such as TANF or SNAP. We have a slight quibble here in that we think that In-program UI benefits should be displayed separately as we have done in figure 11. They are a benefit for TB participants and a cost for the Government/Taxpayers. However, in follow-up years, the TB training is intended to increase individuals' employability and earnings, so the expectation should be a reduction in transfer payments. So the TB participants would bear a cost and the Government/Taxpayer would get a benefit.

Finally, if we include the productivity of TB trainees in follow-up years as a non-monetary benefit, then the last row in figure 5-1 would include the third row in figure 11 above.

All in all, our critique of the benefit-cost framework in the ESD draft report would be the following:

- Line 6 should have a Cost in the Government or Taxpayer column since someone in the private sector is transferring the scholarship or grant to the TB participant
- Line 7 should have a benefit in the Government or Taxpayer column since the loans will be repaid with interest, and thus Society should have a benefit since the interest will be over and above the principal, which is a transfer.
- Line 9b should disaggregate UI benefits between In-program and Follow-up years

TECHNICAL ASSISTANCE AND REPORT REVIEW

Technical Assistance

In addition to the quantitative audit of findings, we made two trips to ESD to gain an understanding of the evaluation methodology and data, and to talk through various decisions about these items. These visits could be essentially considered technical assistance. The first visit occurred on March 27 – 28, 2014, and the second occurred on December 15 – 16, 2014. Highlights of these visits are summarized in the next few paragraphs.

The topics that were covered in the first trip were as follows:

- Understanding Program and Changes Over Time
- Implications of Adding UI Benefits as Additional Outcome
- Design Issues
- Definition of Treatment
- Definition of Comparison Group Pool (ENPs)
- Selection of Comparison Group (P-score matching)
- Net Impact Analyses
- Analysis of CB/ROI
- Report Presentation

It was during this meeting and its discussion of the definition of the treatment that it was decided to define a cohort by the effective date of the individual's first UI benefit payment. If an individual's application for UI was approved at the end of December 2001, but they did not receive a benefit until January 2002, then they would be considered part of the 2002 cohort. This decision was made with the belief that this represented the first time that an individual might consider participating in the TB program.

Another definitional decision that was made concerning the treatment was to only include individuals who had been approved for the TB program and completed at least one course at a community or technical college. This excluded a small number of individuals who attended a proprietary school. It was also decided that the exclusion of individuals under the age of 20 or over the age of 60 that was done in the previous study would not be repeated in this study.

Several technical issues were discussed such as whether a regression model could be estimated, whether it was necessary to do an exact match on the sex of the individual, how to conduct balance and specification testing, whether a difference-in-difference specification was appropriate, and whether any demand side variables could be added to the specification.

Finally, we advised the ESD staff to present findings in graphical form, which they indicated they would do.

The second in-person (technical assistance) visit occurred in December 2014. Prior to the meeting, ESD had sent several documents:

- Project timeline
- Current codebook
- Email exchange with Ashenfelter re: appropriateness of DiD
- Detailed report outline and table shells

I reviewed these materials and the 2013 Annual Report prior to our meeting and formulated a number of questions.

The meeting started out with a presentation of the propensity score model estimates for each of the cohorts—males and females from 2002 to 2012. ESD had estimated the propensity score equations for all 22 groups. Staff members had experimented with different specifications and chose a model that with the following variables in it:

- Ashenfelter dip (decrease in earnings prior to UI spell)
- Pre-training occupation
- Pre-training quarterly earnings for quarters -12 to -2
- Age and age squared
- Education
- WDA
- Race/ethnicity
- Veteran status

The ESD analyst had undertaken a statistical match using one-to-one matching without replacement and exhibited diagnostics – tables that compared distributions of treatment group and matched comparison group characteristics **and** line graphs that showed the earnings profiles of the treatment and matched comparison groups. The diagnostics all suggested that excellent balance had been achieved between the treatment and comparison groups. Nevertheless, we spent much of the day dissecting the propensity score specification.

I asked whether it would be possible to use TB applicants who were deemed not eligible or who did not complete the application process as the comparison pool. This group may/may not make a better pool, but it was determined to be a moot question since these data were not available.

I asked how the team was going to use the employment transition variables that were in the codebook. This led to a lengthy discussion in which there was general agreement that "turnover propensity" is an important behavioral characteristic and may be an important source of variation. I suggested that in addition to the way the codebook defined transitions, the team might want to use a variable that I have used in past studies – pre-training quarters with an employer change. The ESD analyst noted that in her specification experimentation, she used some of the transition variables (as defined in the codebook), and that they did not improve the results.

We discussed the issue of which deflator to use to convert nominal dollars into inflation-adjusted real dollars. This is not an area in which I have expertise, but I had done some investigation prior to the meeting, and fully support the use of the CPI-W, which is unchained.

I questioned the definition of the Ashenfelter dip that was in the codebook. We agreed that it was misdefined, and that the team will re-define it and re-estimate the propensity score equations with a new, correct definition, which may involve three separate variables.

Much of the meeting was spent discussing how the SBCTC data on course-taking would be used in the study. The discussion was predicated on the assumption that quite detailed data would be forthcoming. If that assumption is correct, then the team plans to estimate "black box" outcome equations with a treatment dummy only, and "dosage" equations that fully interact the treatment dummy with course-taking variables. It was furthermore decided that, if available, it would make sense to include pre-training course- taking variables in the propensity score and outcome equations.

The ESD evaluation design calls for estimating the treatment effect through a regression model that explains the outcomes of interest. This regression model will have a treatment dummy (in the black box specification) or interactions between treatment and course-taking variables (in the dosage model specification.) The outcomes include post-treatment quarterly earnings, post-treatment employment rate (percent of quarters during the year that the individual has earnings > $100), and post-treatment UI benefits. If the evaluation were conducted with random assignment, then a reasonable estimate of the treatment effect would be the difference in means between the treatment and control samples.

The analog in a quasi-experimental framework is to use a regression model estimated with both the treatment and comparison sample. The dependent variables in the regression models can be outcomes in time t (post-treatment year) or can be the change in outcomes in time t from a base year. Call these two a *levels model* and a *change model*. The ESD team is calling the latter a difference-in-difference model. The design that was being followed was to estimate the change model using the third year prior to treatment (quarter -12 to -9) as the base year.

ESD had estimated the outcome equations for a couple of cohort/sex groups using the matched sample that had been produced prior to these meetings. The results for employment and earnings had statistically significant treatment effects that seemed signed in the way that was expected and that showed a time path indicating that the TB program participants started behind the comparison group and then bypassed them after a few quarters.

ESD also showed results when UI benefits were the outcome of interest. I suggested that the base period UI benefits might differ systematically between the treatment and comparison group if the former included a higher share of dislocated workers who would likely have had high employment rates, and thus low UI benefit recipiency in the base year. We seemed to reach a consensus that perhaps a levels model would be a better choice for this outcome. The ESD team would like to then put pre-training levels of UI benefits on the right-hand-side, which is fine, although if I am correct, the coefficients on those variables should be insignificant.

The last item of business was that I asked whether the variable disability status was in the propensity score model or the outcomes equations. I had realized that it seemed to be missing when I was preparing this memo. Indeed, it seemed to have "slipped between the cracks," and all agreed that it should be in the models.

As a final word, we would note our appreciation for the time, effort, and objectivity that the ESD staff members have displayed. Their work and the process they followed to produce their findings have been exemplary as far as this auditor is concerned.

Appendix 1

This figure appeared in the ESD's draft report but does not appear in the final published report.

Figure 5-5. Social cost-benefit projections with the 11-year log function estimates of net lifetime earnings*
Washington State, 2002 through 2013
Source: Employment Security Department/LMPA

Group	Discounted social cost-benefit estimates					
	One-year training period		Two-year training period		Three-year training period	
	0% annual decay rate	3.4% annual decay rate	0% annual decay rate	3.4% annual decay rate	0% annual decay rate	3.4% annual decay rate
Panel 1-a: 3 percent discount rate, all participants						
Total sample	$68,570	$61,535	$64,433	$57,398	$60,417	$53,382
Male	$98,503	$89,209	$94,365	$85,072	$90,349	$81,056
Female	$35,930	$30,487	$31,793	$26,350	$27,777	$22,333
Panel 1-b: 4 percent discount rate, all participants						
Total sample	$51,944	$46,044	$47,886	$41,986	$43,985	$38,084
Male	$76,280	$68,497	$72,222	$64,439	$68,320	$60,537
Female	$25,102	$20,498	$21,044	$16,440	$17,142	$12,538
Panel 1-c: 10 percent discount rate, all participants						
Total sample	-$3,220	-$5,370	-$6,847	-$8,998	-$10,144	-$12,295
Male	$3,240	$409	-$387	-$3,218	-$3,685	-$6,516
Female	-$10,889	-$12,653	-$14,516	-$16,281	-$17,814	-$19,578
Panel 2-a : 3 percent discount rate, participants who did not return to their employer of record						
Total sample	$26,019	$20,897	$21,882	$16,760	$17,865	$12,743
Male	$44,354	$37,796	$40,217	$33,659	$36,200	$29,643
Female	$3,029	-$1,339	-$1,108	-$5,476	-$5,124	-$9,492
Panel 2-b :4 percent discount rate, participants who did not return to their employer of record						
Total sample	$14,351	$10,099	$10,293	$6,041	$6,391	$2,140
Male	$28,468	$23,050	$24,410	$18,992	$20,508	$15,090
Female	-$4,076	-$7,761	-$8,134	-$11,818	-$12,036	-$15,720
Panel 2-c : 10 percent discount rate, participants who did not return to their employer of record						
Total sample	-$22,792	-$24,231	-$26,420	-$27,859	-$29,717	-$31,156
Male	-$21,778	-$23,568	-$25,406	-$27,195	-$28,703	-$30,493
Female	-$26,333	-$27,720	-$29,960	-$31,348	-$33,258	-$34,645
Panel 3-a: 3 percent discount rate, participants who returned to their employer of record						
Total sample	$137,816	$128,636	$133,679	$124,499	$129,662	$120,482
Male	$172,646	$161,540	$168,509	$157,403	$164,492	$153,387
Female	$105,505	$99,048	$101,368	$94,911	$97,352	$90,894
Panel 3-b : 4 percent discount rate, participants who returned to their employer of record						
Total sample	$115,327	$107,534	$111,270	$103,476	$107,368	$99,574
Male	$144,295	$134,905	$140,237	$130,848	$136,335	$126,946
Female	$88,525	$83,046	$84,467	$78,988	$80,566	$75,086
Panel 3-c : 10 percent discount rate, participants who returned to their employer of record						
Total sample	$37,404	$34,325	$33,776	$30,698	$30,479	$27,400
Male	$47,475	$43,843	$43,847	$40,216	$40,550	$36,918
Female	$28,754	$26,608	$25,126	$22,980	$21,829	$19,683