

Smart Investments, Real Results:

A Net Impact Evaluation Framework for Minnesota's Workforce Development System and Initial Findings

Report to the Legislature
as required by 2014 Minnesota Statutes 116L.98, subd. 7

01/15/2015

Authors:

Nicholas Maryns, Associate Director, Governor's Workforce Development Council
Raymond Robertson, Professor of Economics, Macalester College

Minnesota Department of Employment and Economic Development

Total cost of salaries, printing, and supplies in
developing/preparing this report is \$9,074.97
(reported as required by Minn. Stat. 3.197).

Department of Employment and Economic Development

1st National Bank Building ■ 332 Minnesota Street ■ Suite E200 ■ Saint Paul ■ Minnesota 55101-1351
651-259-7114 PHONE ■ 800-657-3858 TOLL FREE ■ 651-296-4772 FAX ■ 651-296-3900 TTY ■ <http://mn.gov/deed>

An Equal Opportunity Employer and Service Provider

About this Document

This report details the development, methodology and initial results of a net impact evaluation framework for Minnesota's workforce development programs. The work described here builds off of the work of the Governor's Workforce Development Net Impact Initiative, and fulfills a requirement of [2014 Minnesota Statutes 116L.98, subd. 7](#).

About the Governor's Workforce Development Council

The Governor's Workforce Development Council works to strengthen Minnesota's workforce. Made up of leaders from business, education, labor, community organizations, and government, the GWDC forges practical solutions to Minnesota's workforce challenges and provides leadership on projects of strategic importance to the state. As Minnesota's state Workforce Investment Board, the GWDC is mandated and funded by the federal Workforce Investment Act and further defined by Minnesota Statutes, section 116L.665. More information can be found at www.gwdc.org.

About DEED

The Minnesota Department of Employment and Economic Development (DEED) is the state's principal economic development agency. DEED programs promote business recruitment, expansion, and retention; international trade; workforce development; and community development. Learn what DEED is all about and the ways we help job seekers, businesses and communities at mn.gov/deed.

About the Authors

Nicholas Maryns is Associate Director of the Governor's Workforce Development Council. Nick facilitates the GWDC's policy development efforts and leads up research, analysis, and communications for the Council. Prior to working for the GWDC, Nick was a fellow at the Congressional Hunger Center in Washington, DC and an AmeriCorps volunteer in the Minneapolis Public Schools. He has a Masters in Public Policy from the Harvard Kennedy School of Government and a BA in Mathematics with Distinction from St. Olaf College.

Raymond Robertson is Professor of Economics at Macalester College. His research focuses on the union of international, labor, and development economics. He has published in the American Economic Review, the Review of Economics and Statistics, the Journal of International Economics, Review of International Economics, the Journal of Development Economics, and others. He received his Ph.D. from the University of Texas after spending a year in Mexico as a Fulbright Scholar.

Acknowledgements

This report represents the work of many individuals who provided their expertise and input, helped share and analyze data, and provided support in countless other ways. The authors would especially like to thank the GWDC Net Impact Advisory Group (see Appendix page i) and the following individuals: Anne Barry, Stacy Becker, Deb Briggs, Richard Chase, Kasandra Church, Leslie Crichton, John Dahle, Dana DeMaster, Jose Diaz, Grant Duwe, Sarina Foreman, Amy Gehring, Karen Gibson, Jake Granholm, Gregory Gray, Jody Hauer, Jim Hegman, George Hoffman, Britta Holland, Kevin Hollenbeck, Anna Jacob, Chuck Johnson, Mike Jokinen, Pete Jonas, Chris King, David Kirchner, Larry Kontio, Leah Krotzer, Shelley Landgraf, Anne Lauer, Troy Mangan, Ruth Martin, Denese McAfee, Tom Pettet, Mary Phillippi, Ed Potter, Debbie Rielley, Shannon Scheurich, Jeannie Schwartz, Carolyn Schworer, Rachel Speck, Mark Toogood, Julie Toskey, Sarah West, John Wiersma, Bryan Wilson, and Derek Wolfe.

This document can be made available in alternative formats upon request.

Table of Contents

Executive Summary	1
Purpose and Goals	3
Project History	5
Pilot Study: Purpose and Scope	6
Net Impact Evaluation Design	8
Quasi-Experimental Approach	8
Propensity Score Matching Treatment and Control Cohorts	9
Difference-in-Differences Approach	10
Outcomes and Timeframes Analyzed.....	11
How Results are Disaggregated	12
Net Impact Results	13
Treatment Groups	13
Control Groups.....	15
Interpreting Statistical Tables.....	17
Assessing the Treatment-Control Group Match.....	18
Annual Earnings	19
Employment Likelihood	25
Quarterly Cash Benefits (MFIP and SNAP)	30
Limitations and Caveats	34
Conclusion and Ideas for Further Development and Implementation	36
Answering Important Policy and Performance Questions	36
Contextualizing Results to Encourage Proper Interpretation	37
Making Results More Useful and Accessible	37
APPENDICES	
APPENDIX I: GWDC Net Impact Advisory Group Members.....	i
APPENDIX II: Literature Review, Empirical Methodology, and Data Process.....	ii
APPENDIX III: References.....	x
APPENDIX IV: A Framework for Estimating Costs, Benefits, and Social Return on Investment	xiii
APPENDIX V: Technical Limitations and Caveats	xviii
APPENDIX VI: GWDC and OLA Recommendations on Net Impact Evaluation.....	xxi

Executive Summary

Purpose

The Governor's Workforce Development Council, in collaboration with a number of state agencies, has developed a framework for the net impact evaluation of the state's workforce programs. Led by an advisory group made up of a diverse array of stakeholders, the GWDC's Net Impact Initiative has been implementing a relatively small-scale pilot of the framework since 2012. The goal of this pilot has been to inform the development of a high-quality net impact evaluation framework, and to establish the necessary data procedures to make way for the implementation of a larger, ongoing performance management effort. This report details the development, methodology, and initial results of the net impact evaluation framework.

Methodology

In short, a net impact evaluation measures the outcomes of program participants compared against a control group of similar non-participants. It accounts for factors like participant demographics, work history, and local economic conditions, seeking to isolate the impact of the program itself. We use two approaches—kernel density propensity score matching and regression-adjusted difference-in-difference estimation—to estimate the net impacts of workforce programs with a high level of rigor.

Major Findings

This report presents a number of findings regarding the role of net impact evaluation in the continuous improvement of workforce services and participant outcomes.

1. **The framework developed by the GWDC advisory group is feasible and appropriate for many workforce programs.**

The advisory group has developed an evaluation methodology that can help the state understand the real net impacts of workforce programs, aid in continuous improvement, and provide a foundation for estimating associated costs and benefits (and thus return on investment). Given the initial investment that has been made to help automate and replicate the analyses, the approach will be straightforward to apply moving forward.

2. **State agencies have made significant progress in building data-sharing relationships that will continue to pay off into the future.**

Connections and relationships between DHS, DOC, and DEED¹ have identified areas of cooperation and common concern that will result in significantly lower data coordination costs in the future. Differences in data definitions and formats across different departments have required a significant investment in data cleaning and harmonization that will ensure lower costs in the future.

3. **Net impact evaluation faces challenges when applied to small programs and populations.**

The academic literature identifies a minimum sample size of at least 200. Some of Minnesota's workforce programs fall below this participant threshold; we also run into barriers when estimating net impacts for specific populations that may be small in size. Although the empirical methodology can be applied in these cases, the resulting estimates can be unreliable. Part of the challenge is data integrity; incomplete data result in a significant number of individuals being dropped from the analysis. Improving data quality would increase sample sizes and allow the framework to be applied more broadly.

¹ The Departments of Human Services, Corrections, and Employment and Economic Development, respectively.

4. **Looking forward, continued investment in this type of workforce evaluation will be increasingly important in a national context.**

An increasing number of states, Minnesota² included, have established evaluation requirements, while others are using net impact models to create pay-for-success models increasingly of interest to governmental and non-governmental funders at the national level. By continuing to invest in and improve data collection and analysis efforts, Minnesota can help ensure that its workforce education and training programs continue to be nation-leading.

Evaluation Findings

We were able to successfully apply the net impact evaluation methodology to the Workforce Investment Act (WIA) Adult and Dislocated Worker (both WIA and state-funded) programs for two cohort periods. This report details the findings of these analyses. The net impacts of these training programs are quite positive and generally statistically significant. In particular:

1. **Both programs are responsible for large net impacts on annual earnings and employment likelihood.** Program participants earned substantially more and were more likely to be employed than they would have been had they not participated in the program. For the earlier cohorts (2007-2008), these impacts appeared to decrease as time went on for, while impacts on the later cohorts (2009-2010) either stayed fairly constant (WIA Adult) or grew (Dislocated Worker) as time went on. These differing trends may be related to the changing state of the economy over the given periods.
2. **The programs appear to impact earnings and employment differently with regard to gender.** The WIA Adult program shows larger impacts for men while the Dislocated Worker program shows larger impacts for women.
3. **Three of the four cohorts produce greater earnings and employment impacts for individuals living in the seven-county metro area than for individuals living in Greater Minnesota.**
4. **The WIA Adult program produced notably larger earnings and employment net impacts for African Americans than for white participants.**
5. **Our analysis of net impacts on the amount of quarterly cash benefits received (MFIP and SNAP) yielded far fewer reliable results, in large part because of the smaller number of individuals in both the treatment and control groups receiving these benefits.** That said, three of the four cohorts (all but WIA Adult 2009-2010) appear to have produced reductions in the amount of cash benefits received, though by small amounts in the case of the Dislocated Worker program.

² See 2014 Minnesota Statutes 116L.98, subd. 7, a recently-enacted law requiring net impact evaluations every four years, available at: <https://www.revisor.leg.state.mn.us/statutes/?id=116L.98>

Purpose and Goals

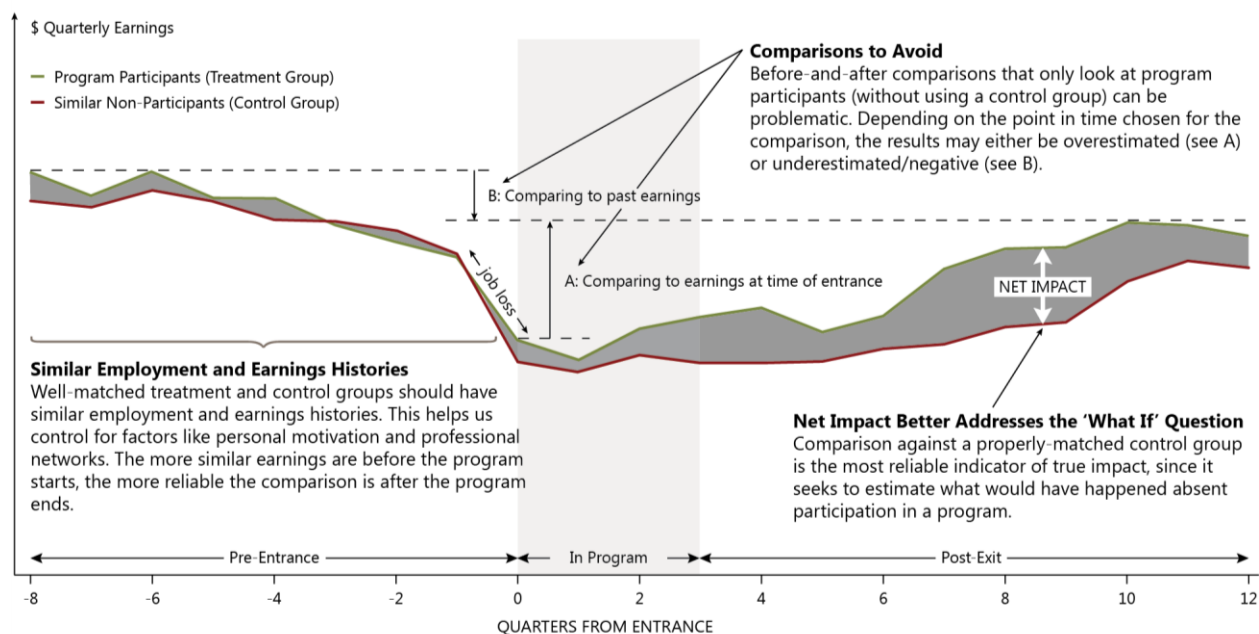
Workforce employment and training programs have long helped Minnesotans gain skills and find better jobs, and many have endeavored to measure their impact in one way or another. Yet very few rigorous, standardized evaluations of the true net impact have been implemented, let alone embedded into ongoing performance management practices. This has limited our ability to understand and replicate what works.

What is Net Impact?

In short, a net impact evaluation measures the outcomes of program participants compared against a control group of similar non-participants. It accounts for factors like participant demographics, work history, and local economic conditions, seeking to isolate the impact of the program itself.³

FIGURE 1: UNDERSTANDING NET IMPACT: COMPARISONS THAT ILLUMINATE AND COMPARISONS TO AVOID

To better understand the true impact of workforce education and training programs, program participants are compared to other individuals who are similar with regard to demographics and employment history, but who have not participated in a program.



Note: The above chart is for illustrative purposes only and does not reflect actual data or outcomes. However, the overall pattern is typical: average quarterly earnings among program participants typically decrease significantly the 1-2 quarters preceding entrance into a program, signifying job losses and/or reduced hours.

Net impact evaluation can help guide workforce strategy and investment.

Net impact evaluation is just one tool of many that can be used to understand and improve workforce programs—it is not the final word on the outcomes of programs. Given its rigor and depth, it is an important tool to continue developing and implementing.

Over the past five years, a net impact evaluation framework has been developed by the Governor’s Workforce Development Council (GWDC) in conjunction with a broad group of stakeholders and evaluation experts. The

³ See *Net Impact Evaluation Design* on page 8 for more information.

framework uses available data to understand program impacts, aid in continuous improvement, and support evidence-based policy making.

The framework is designed to produce insights on two fundamental types of questions:

- *What works?* What kinds of services and approaches have the greatest impact on participants?
- *Who is impacted?* How do services impact different populations, and how can program partners ensure high-quality outcomes across all populations?

Net impact evaluation increases the transparency and accountability of public investments in workforce development, and demonstrates the value of these investments to participants and the broader public. It improves our efforts to serve jobseekers, students, employers, and industry.

The work described below was initiated and sustained with a few broad goals in mind:

- The use of regular, standardized net impact evaluations produces insights that increase the ability of the workforce system to produce positive outcomes and eliminate disparities.
- Policymakers, funders, and service providers in workforce development and other areas see the value of data-informed decision-making and learn how to analyze and interpret results data.
- More broadly, Minnesota embraces a culture of evidence with regard to understanding and addressing public problems.

Project History

Across the country, workforce development programs have begun to harness the power of data and net impact evaluation methods to help drive strategy and investment. At least half a dozen states have used net impact evaluation to understand and improve the outcomes of workforce services.⁴ The federal Department of Labor has commissioned a number of net impact studies in recent years, and, as part of its Workforce Innovation Fund, has awarded grants to states to develop pay-for-success projects built on the net impact approach. Other national efforts, such as Results for America, the Benchmarking Project, and America Achieves, also support efforts to expand the use of data and net impact evaluation.

Minnesota Context

Over the last decade, Minnesota's leaders have increasingly noted the need for more rigorous and standardized approaches to evaluating workforce programs.

In 2009, Minnesota enacted a new set of "Uniform Program Accountability Measures"—known as UPAM—for DEED's economic development and workforce development programs.⁵ One of the measures stipulated by the UPAM law was return on investment. In response to this law, the GWDC launched its Return on Investment Initiative⁶ and convened an advisory group⁷ to study the issue and develop a standardized measure, in alignment with its statutory role to:

*Advise the governor on the development and implementation of statewide and local performance standards and measures relating to applicable federal human resource programs and the coordination of performance standards and measures among programs.*⁸

Made up of a broad range of partners, including key staff from relevant state agencies, workforce development service providers, business members, community-based organizations, and data evaluation experts, the GWDC advisory group set out to develop a standard return on investment methodology that could be applied to workforce programs administered or funded with public dollars.

To guide their work, the advisory group agreed to a number of shared values and goals for the methodology, namely that it should be transparent and credible, adaptable and sensitive to change, relatively simple to administer, and yield timely and relevant results.

Shift of Focus to Net Impact Evaluation

The advisory group chose to estimate return on investment through net impact evaluation, which takes a scientific approach to estimating and attributing program impacts, limiting the use of broad assumptions. The advisory group studied a number of net impact/return on investment evaluations in other states and at the federal level.

In early 2010, the Office of the Legislative Auditor released an evaluation report on workforce programs that provided some initial net impact findings, and recommended that, "DEED should adopt a set of standard approaches for assessing workforce program outcomes, including periodic comparisons of workforce program participants and non-

⁴ See Hollenbeck and Huang, "Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State," and Heinrich, Carolyn et al., "New Estimates of Public Employment and Training Program Net Impacts: A Nonexperimental Evaluation of the Workforce Investment Act Program."

⁵ Minnesota Statutes § 116J.997, which have since been repealed and replaced.

⁶ This initiative has since been re-named the *Net Impact Initiative*, and is referred to as such throughout the report. Learn more at www.gwdc.org/net-impact/

⁷ See the appendix on page i for a list of advisory group members.

⁸ See Minnesota Statute 116L.665 Subd. 3c, available at: <https://www.revisor.mn.gov/statutes/?id=116L.665>.

participants.”⁹ This report further emphasized the importance of developing and implementing a standardized net impact framework for Minnesota’s workforce development programs—a foundation upon which return on investment estimates could be built.

The advisory group developed the broad parameters of the framework, such as which outcomes to measure, how to measure them, and how to estimate associated costs and benefits for the purpose of return on investment analyses.¹⁰ In doing so, the group balanced the varied interests and perspectives of its members to develop a consensus-based framework ready for implementation.

Pilot Study: Purpose and Scope

In 2012, the advisory group began to lay the groundwork for a relatively small-scale pilot evaluation. The goal of the pilot has been to inform the development of a high-quality net impact evaluation framework, and to establish the necessary data procedures to make way for the implementation of a larger, ongoing performance management effort.

The work of the pilot has involved building partnerships with various state agencies, studying state agency programs, and establishing data-sharing procedures across various data systems. Along the way, we have addressed numerous challenges, including a lack of precedent for the type and scope of cross-agency data sharing required, instituting an infrastructure for storing and sharing large datasets, and determining exactly which data are needed and how to properly link them. Inconsistencies in how (and whether) data are reported and defined across state data systems were also a barrier, requiring a great deal of work to standardize the data and prepare it for analysis.

Challenges notwithstanding, a major benefit of the approach has been that it uses data that are already available; it does not require service providers to collect additional data or program participants to self-report. The pilot has provided a clear pathway forward for future analyses, minimizing much of the upfront cost in large part due to the voluntary participation of our evaluator, Raymond Robertson.

The pilot evaluation, which is detailed below, analyzes the net impacts of two major workforce programs, the Workforce Investment Act (WIA) Adult Program and the Dislocated Worker Program (both WIA- and state-funded) operating between 2007 and 2010 in the midst of the Great Recession and its after-effects. The evaluation focuses on the impacts of these programs on employment, earnings, and participant use of cash benefits (namely MFIP and SNAP).¹¹ Workforce program participant outcomes are analyzed against the outcomes of comparison groups constructed from similar non-participants who either applied for unemployment insurance benefits or who registered at a WorkForce Center or online at MinnesotaWorks.net.¹² The work described here builds on previous analyses described earlier, using new statistical techniques to take a step forward in the field of workforce program evaluation.

Collaborative Approach

Along the way, the advisory group, evaluator, and staff have worked collaboratively with other evaluation experts in Minnesota to build consensus around a common framework for net impact and return on investment analysis. This includes Wilder Research, the Greater Twin Cities United Way, and Invest in Outcomes. The goal is that by achieving consensus, we may encourage further evidence-based policy-making efforts, foster the credibility and transparency of the framework, and save time and resources by avoiding unnecessary duplication. Thus another goal of the pilot project (and this report) is to provide clear guidance to state agencies and evaluators for reproducing and improving this work in the future.

⁹ Office of the Legislative Auditor, State of Minnesota, “Evaluation Report: Workforce Programs.” Available at:

<http://www.auditor.leg.state.mn.us/ped/pedrep/workforce.pdf>. See appendix page xxii for the full recommendation.

¹⁰ See the appendix on page xiv for the social return on investment parameters developed by the advisory group. These parameters can be used to estimate the costs and benefits associated with workforce programs and their net impacts.

¹¹ MFIP is the Minnesota Family Investment Program, the state’s TANF program; SNAP is the Supplemental Nutrition Assistance Program, formerly known as Food Stamps.

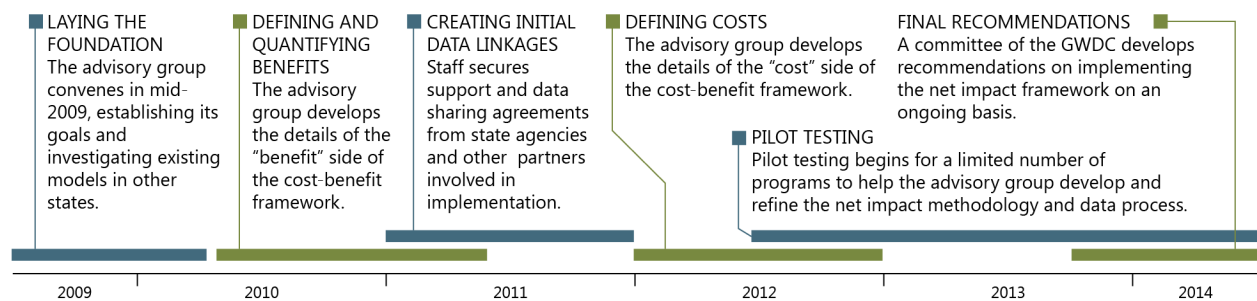
¹² See page 8 for details on the evaluation design.

GWDC Net Impact Policy Recommendations

In September of 2013, the GWDC established a committee of its members to 1) review and understand the work of the advisory group, staff, and evaluator, 2) consider the initial recommendations of the advisory group, and 3) make final recommendations to the full council for approval. The recommendations of the committee addressed the technical aspects of the evaluation model as well as policy recommendations on effectively implementing net impact evaluation to help workforce stakeholders continuously improve the performance of programs. The committee's recommendations were approved and included in the GWDC's 2014 policy advisory, and serve as a more policy-minded companion to this report.¹³

FIGURE 2: NET IMPACT INITIATIVE TIMELINE

The Governor's Workforce Development Council first convened an advisory group on net impact and return on investment in the summer of 2009. The group consists of a diverse array of stakeholders and experts with the goal of developing a standardized net impact/ROI methodology that can be applied to workforce programs administered or funded by public dollars, including nonprofit programs that receive pass-through funding.



¹³ Governor's Workforce Development Council, *Building Partnerships to Overcome Barriers*. Available at <http://www.gwdc.org/docs/publications/GWDC-Building-Partnerships-2014.pdf>. See appendix page xxii for the full recommendations.

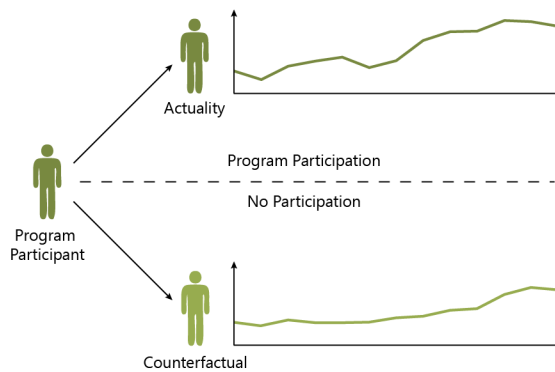
Net Impact Evaluation Design

Many factors impact an individual's employment and earnings outcomes: past education and work experience, race and gender, personal motivation and networks, and so on. For workforce development programs, the goal is to be a key factor in improving an individual's outcomes. But since so many factors are at play, it is difficult to understand the true impact of a program on an individual's employment and earnings. A program may appear to help a participant, but how much of the participant's success is due to the program itself, rather than, for example, the personal motivation that brought the participant to the program in the first place?

In seeking to understand the true impact of programs, we start with the question, "What would have happened to the program participant if they had not participated in the program?" This alternate reality, where the participant does not participate, is often called the counterfactual. The rest of the factors affecting the individual stay the same; the difference between the actual outcome and the counterfactual outcome has to do only with program participation—what we call the program's net impact.

FIGURE 3: THE ACTUAL AND THE COUNTERFACTUAL

In an idealized setting, we could observe not only what happens to program participants, but what would have happened to them if they had not participated in the program.



Since this is not possible, we instead seek to mimic the counterfactual by matching program participants with other individuals who are similar with regard to demographics and employment history, but who have not participated in a program.

Because we cannot observe both the actual outcome and its counterfactual, we seek to mimic it as best we can. This is often done through randomized controlled trials, where many individuals are randomly assigned to two groups, one that participates in a program and one that does not. Because the other factors at play (e.g. personal history and characteristics) are randomly spread across the two groups, their associated impacts are likely the same across both groups. So the only difference between the groups has to do with the program, and this difference is the program's net impact.

This kind of experimental approach, while considered the gold standard in program evaluation, is expensive and difficult to implement, in part because it involves selecting a large number of individuals to not receive services, and then tracking them closely.

Quasi-Experimental Approach

Fortunately, a quasi-experimental approach that relies on data already collected by the State of Minnesota can be used to achieve similar evaluation results. This is possible due to 1) recent advances in statistical research methods and 2) the fact that we can collect data on participants and similar non-participants both before and after program participation. This allows us to use two approaches—kernel density propensity score matching and regression-adjusted difference-in-difference estimation—to estimate the net impacts of workforce programs.¹⁴

The propensity score matching approach helps us match program participants with similar individuals who would be equally likely to participate in a workforce program (but did not). The difference-in-difference approach helps us account for unobservable factors that may affect outcomes, such as personal motivation or the strength of professional networks.

¹⁴ What follows is an attempt to describe the major features of our net impact approach, geared toward a lay audience. For a more in-depth, technical treatment, see the appendix on page ii.

Importantly, recent analyses comparing the validity of quasi-experimental approaches (such as ours) against experimental randomized trials have found that the quasi-experimental approach is just as valid and may even produce more appropriate estimates of changes in earnings.¹⁵

Propensity Score Matching Treatment and Control Cohorts

To estimate the true impact a program has on a participant, we need to understand what would have happened to the individual if they had not participated in the first place. While we cannot observe the same individual both participating in a program and *not* participating at the same time, we can use available data to construct a “match”—a non-participant that is similar to the participant.

In this evaluation, we study the outcomes of participants in two programs:

- The Workforce Investment Act (WIA) Adult Program
- The Dislocated Worker Program (both WIA- and state-funded)

For each program, we look at two cohorts (often referred to as treatment groups), which include individuals who exited the above programs in two timeframes:

- Between July 1, 2007 and June 30, 2008
- Between July 1, 2009 and June 30, 2010

Using propensity score matching, we construct groups of similar non-participants (often referred to as control groups) drawn from the following pools:

- Individuals who applied for Unemployment Insurance benefits in the same period that our treatment groups were entering programs
- Individuals who set up a user account at a WorkForce Center or online at www.MinnesotaWorks.net and last used these basic services during the same timeframe.

These pools were selected because they meet three conditions identified by Smith and Todd (2005) that are necessary for a proper control group:

- Comparison group workers should be from the same labor market and therefore experiencing the same labor market conditions.
- Data should be drawn from the same dataset.
- The data should contain a rich set of variables that affect both the outcome and the propensity to participate in the program.

In addition, earlier evaluations we reviewed used the same pools and made a strong case for their appropriateness. Using these pools, treatment and control groups were matched across a number of key characteristics:

- Past employment and earnings
- Gender
- Age
- Race
- Veteran status
- Highest level of education attained
- Geography
- Past enrollment in public benefits programs

¹⁵ See the meta-analyses of Greenberg et al. of 31 studies (2006) and Card et al. of 199 studies (2010) that conclude that experimental and quasi-experimental studies produce no statistically different estimates of program effectiveness.

Past employment and earnings are particularly important because they shed light on factors that are not directly recorded in the data, such as individual-level productivity and motivation.

Modeling the Selection Process

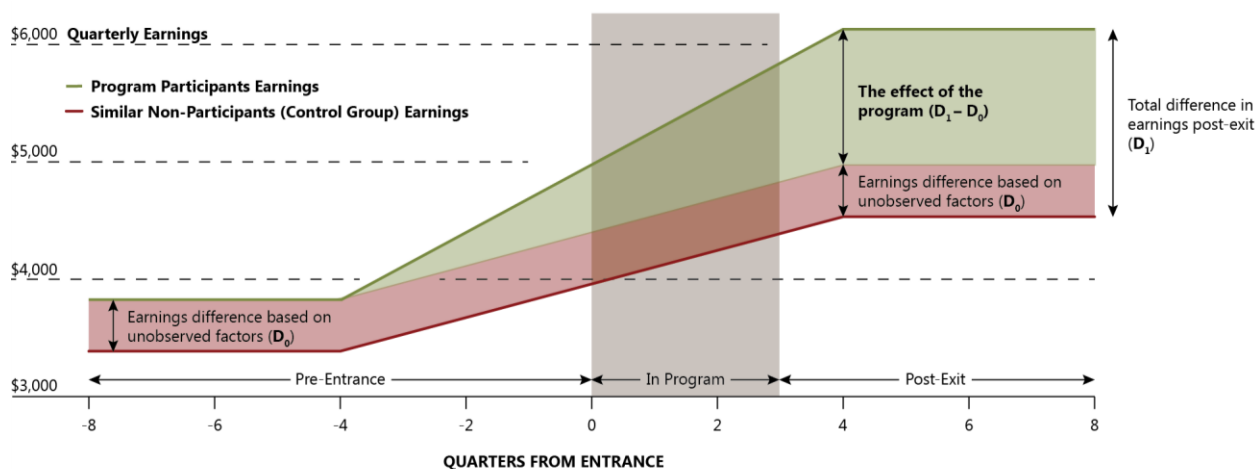
In the propensity score matching approach, we look at the types of individuals who participate in programs and ask: “What kinds of characteristics make a person likely to participate in a workforce program?” We are able to create a mathematical formula that estimates how likely any individual is to participate in the program (their propensity score) based on the above characteristics. We then match treatment and control individuals based on this propensity score. In this way, we seek to reduce selection bias—the possibility that our treatment and control groups differ in ways we cannot observe.

Difference-in-Differences Approach

How do we know net impacts are due to the program, and not to other factors like personal motivation? What if the treatment group is different from the comparison group in ways the data don’t reveal? The difference-in-differences technique addresses these questions. It estimates how unobservable factors may be affecting our results and removes those effects from the net-impact equation. Here’s how it works (see Figure 4 below):

- **Step 1:** Program participants are matched with similar non-participants across an array of factors, including demographics and past employment and earnings.
- **Step 2:** After matching, any differences that remain in pre-entrance earnings (D_0) are attributed to unobserved factors, like motivation, that don’t have to do with the program.
- **Step 3:** We look at the total difference in post-exit earnings between program participants and control group members (D_1). This total difference is made up of two factors: the effect of the program itself, and the unobserved factors (D_0) we saw in Step 2.
- **Step 4:** To isolate the effect of the program itself, we subtract unobserved factors (D_0) from the total difference (D_1). This amount ($D_1 - D_0$) is the program effect we’re interested in.

FIGURE 4: THE DIFFERENCE-IN-DIFFERENCES APPROACH



Note: The above chart is for illustrative purposes only and does not reflect actual data or outcomes.

Notably, the difference-in-differences approach is possible because we are able to use data on individual-level earnings and employment both before and after program participation. This richness of our data set helps to make our analysis more robust.

Outcomes and Timeframes Analyzed

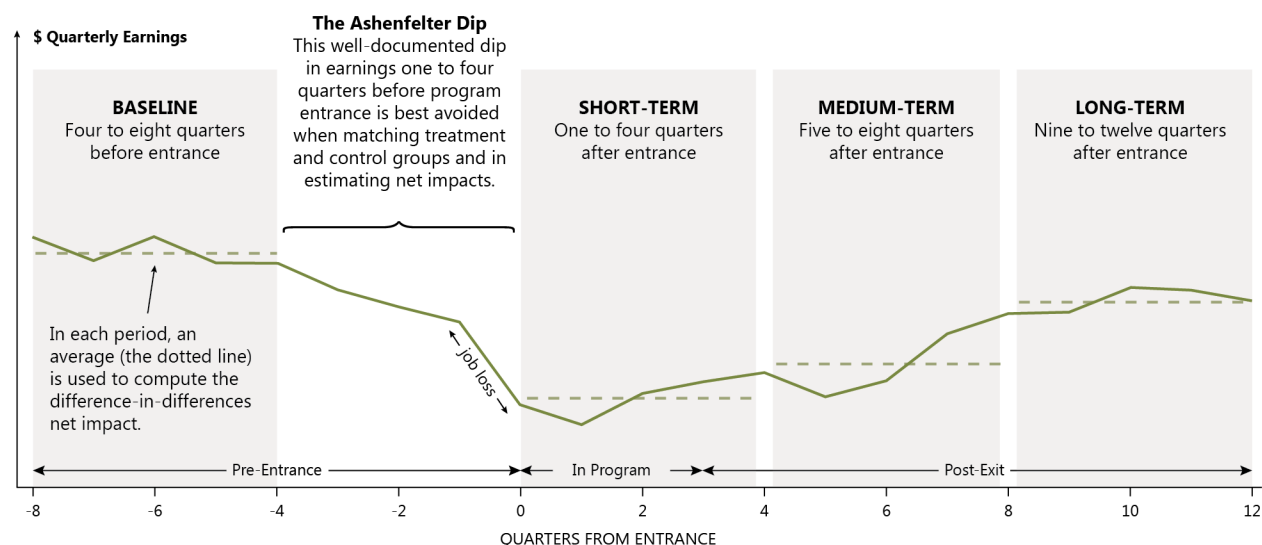
The framework analyzes the net impacts of workforce programs across an array of outcomes.¹⁶

- Earnings¹⁷
- Employment
- Usage of public cash benefits (e.g. MFIP, SNAP)¹⁸

Timeframes for Matching and Analysis

For the purposes of our difference-in-differences approach, we have defined a “baseline” period that takes place before program participation, and a few “follow-up” timeframes. Within each of these four-quarter timeframes, average earnings are used to compute difference-in-differences net impacts.

FIGURE 5: TIMEFRAMES FOR ANALYSIS AND THE ASHENFELTER DIP



Note: The above chart is for illustrative purposes only and does not reflect actual data or outcomes. However, the overall pattern is typical: average quarterly earnings among program participants typically decrease significantly the 1-2 quarters preceding entrance into a program, signifying job losses and/or reduced hours.

Baseline: 4 to 8 quarters before entrance

It is particularly important to set a baseline that is at least four quarters before program entrance. The reason is the Ashenfelter Dip, a well-documented dip in earnings that occurs prior to program entrance. If participants are matched with non-participants based on their earnings during the dip, there is a risk that the participant will be matched to a non-participant whose earnings have not dipped, but are lower in general. This can create an inappropriate match and distort net impact estimates.

Follow-Up: Short-term: 1-4 quarters after entrance

Medium-term: 5-8 quarters after entrance

Long-term: 9-12 quarters after entrance

¹⁶ Impacts across three other outcomes—enrollment in public healthcare programs (e.g. Medical Assistance, MinnesotaCare), usage of Unemployment Insurance Benefits, and incarceration avoidance—have also been pursued in the pilot evaluation. The healthcare and Unemployment Insurance analyses are pending, but the analysis of incarceration avoidance presents additional challenges; the number of individuals in our evaluation who have any history of incarceration is so small that the statistical analysis is unreliable.

¹⁷ We deliberately chose to analyze overall earnings, as opposed to hourly wages. The Unemployment Insurance “Wage Detail” data we use captures both quarterly earnings and hours worked for roughly 98 percent of all Minnesotans. The “earnings” portion of this data set is known to be consistently and accurately reported by employers. The “hours worked” portion is not as consistently reported by employers. Therefore, an hourly wage rate calculation would be less certain than overall earnings.

¹⁸ Changes in benefit levels are observed directly from the Department of Human Services using data matching techniques.

An array of follow-up timeframes help us understand whether net impacts grow and/or persist over time. It is particularly important in any analysis of education and training impacts to allow for a lengthy follow-up period, since impacts often take time to appear and may last into the longer term.

How Results are Disaggregated

In addition to overall program net impacts, results can be broken down to provide greater insights and dimension. We are able to disaggregate results across characteristics that are captured consistently in the data and where there are enough individuals to provide a reliable result. Results are disaggregated based on:

- Age, across four groupings:
 - 18-24
 - 25-44
 - 45-54
 - 55-64
- Gender
- Race, across two categories:¹⁹
 - White
 - African American
- Geography, across two regions:
 - The seven-county metro region
 - Greater Minnesota
- Highest level of education attained, across four categories:²⁰
 - Less than a high school diploma
 - High school diploma or equivalent
 - Some postsecondary education
 - Associate degree or higher

¹⁹ Very small sample sizes made other racial categories and ethnic categories difficult to analyze.

²⁰ We chose to use groupings commonly used in other datasets, such as the U.S. Census, the American Community Survey, and in other labor market data tools produced by DEED. We are particularly interested in the outcomes of individuals with less than an associate degree. That said, the “some postsecondary education” category is particularly hard to interpret because it includes many different types of individuals such as those who may have completed postsecondary award at the sub-associate level and those who may have attended, but not completed, a four-year program. More fine-grained analyses may be possible in the future, barring too-small sample sizes.

Net Impact Results

In this section, we detail the results of the net impact pilot evaluation. We start by describing the treatment and control groups evaluated and ensuring that the groups are well-matched. We then evaluate the net impacts of the programs on three participant outcomes: earnings, employment, and the usage of public cash benefits (e.g. MFIP, SNAP). In other words, we evaluate the extent to which the programs are responsible for:

- Changes in average annual earnings
- Changes in the likelihood of employment
- Changes in the amount of quarterly cash benefits (MFIP and SNAP) received

See Figure 6 on page 17 for guidance on how to interpret the statistical tables that follow.

Treatment Groups

For the purpose of our pilot evaluation, the treatment population includes any individual who exited the following programs 1) between July 1, 2007 and June 30, 2008 or 2) between July 1, 2009 and June 30, 2010:

- The Workforce Investment Act (WIA) Adult Program
- The Dislocated Worker Program (both WIA- and state-funded)

The 2007-2008 cohorts exited their programs as the Great Recession was starting (the recession officially began December 2007), while the 2009-2010 cohorts exited as the recession was ending (the recession officially ended June 2009).

Workforce Investment Act (WIA) Adult Program

The WIA Title 1-B Adult program provides employment and training assistance to adults who face significant barriers to employment. Minnesota's Adult program prioritizes individuals who receive public assistance, individuals living with low incomes, and veterans within these groups. For each customer, the overarching goal is employment or enhancement within his or her occupation. Generally, Adult program customers work to increase their earnings, retain employment, and diversify their occupational skills.²¹

The Adult program provides services through a network of 48 WorkForce Centers Adult program counselors meet with customers, provide services, and coordinate training.

Dislocated Worker Program

The Minnesota Dislocated Worker program helps workers who lost their jobs through no fault of their own—that is, they neither quit nor were fired—find a new career.²² An individual typically must qualify for Unemployment Insurance benefits to be eligible. Dislocated Worker staff work with Unemployment Insurance to ensure Minnesota's workers are able to get stable jobs in high-demand occupations. The following groups of people are also eligible for services:

- Self-employed individuals who lose their jobs due to economic conditions
- Veterans leaving active duty with the armed forces
- Certain individuals leaving active duty of the National Guard or armed forces reserves

²¹ See WIA Annual Report, Program Year 2013, available at: www.doleta.gov/Performance/Results/AnnualReports/PY2013/MN-PY13_WIA_AnnualReport.pdf

²² For more information, see <http://mn.gov/deed/programs-services/dislocated-worker>

The Dislocated Worker program provides services through a network of 48 WorkForce Centers and several independent non-profit organizations. Dislocated Worker program counselors meet with customers, provide services, and coordinate training.

Definitions of Dislocated Workers

For the purposes of the program, a dislocated worker is defined as an individual who:

- Has been terminated or laid off, or has received a notice of termination or layoff from employment;
- Is eligible for or has exhausted unemployment insurance;
- Has demonstrated an appropriate attachment to the workforce, but not eligible for unemployment insurance and unlikely to return to a previous industry or occupation;
- Has been terminated or laid off or received notification of termination or layoff from employment as a result of a permanent closure or substantial layoff;
- Is employed at a facility, where the employer has made the general announcement that the facility will close within 180 days;
- Was self-employed (including employment as a farmer, a rancher, or a fisherman) but is unemployed as a result of general economic conditions in the community or because of a natural disaster; or
- The Rehabilitation Services Administration oversees grant programs that help individuals with physical or mental disabilities to obtain employment and live more independently through the provision of such supports as counseling, medical and psychological services, job training and other individualized services.

Common Aspects of Both Programs

The Adult and Dislocated Worker Programs share the same goals and general types of services.

Goals

- To increase employment, as measured by entry into unsubsidized employment
- To increase retention in unsubsidized employment six months after entry into employment
- To increase earnings received in unsubsidized employment for dislocated workers
- To enhance customer satisfaction for participants and for employers

Services

There are three levels of service:

- *Core services:* Includes outreach, job search and placement assistance, and labor market information available to all job seekers;
- *Intensive services:* Includes more comprehensive assessments, development of individual employment plans and counseling and career planning; and
- *Training services:* Customers are linked to job opportunities in their communities, including both occupational training and training in basic skills. Participants use an "individual training account" to select an appropriate training program from a qualified training provider.

Additional Services

- "Supportive" services such as transportation, childcare, dependent care, housing and needs-related payments are provided under certain circumstances to allow an individual to participate in the program.
- "Rapid Response" services at the employment site for employers and workers who are expected to lose their jobs as a result of company closings and mass layoffs are also available.
- Individuals whose layoff was created or affected by international trade, may access information and services under the Trade Act programs.

- States are responsible for program management and operations including enrollment, service delivery, and certification of training providers.

TABLE 1: DEMOGRAPHIC BREAKDOWNS OF TREATMENT COHORTS

	WIA Adult 2007-2008		WIA Adult 2009-2010		Dislocated Worker 2007-2008		Dislocated Worker 2009-2010	
	N	%	N	%	N	%	N	%
OVERALL TOTAL	745	100%	1,093	100%	6,977	100%	10,199	100%
AGE								
18-24	97	13.0%	231	21.1%	137	2.0%	298	3.1%
25-44	308	41.3%	548	50.1%	2,145	30.7%	4,031	41.7%
45-54	128	17.2%	201	18.4%	2,009	28.8%	3,545	36.7%
55-64	38	5.1%	66	6.0%	1,035	14.8%	1,795	18.6%
Other or No Data	174	23.4%	47	4.3%	1,651	23.7%	530	5.2%
GENDER								
Female	396	53.2%	596	54.5%	3,391	48.6%	4,508	44.2%
Male	349	46.8%	497	45.5%	3,586	51.4%	5,691	55.8%
RACE								
African American	115	15.4%	220	20.1%	495	7.1%	811	8.0%
White	414	55.6%	735	67.2%	4,321	61.9%	8,763	85.9%
Other or No Data	216	29.0%	138	12.6%	2,161	31.0%	625	6.1%
GEOGRAPHY								
Greater Minnesota	536	71.9%	687	62.9%	3,370	48.3%	3,478	34.1%
Seven-County Metro	209	28.1%	406	37.1%	3,607	51.7%	6,721	65.9%
EDUCATIONLEVEL								
Less than HS	113	15.2%	127	11.6%	290	4.2%	370	3.6%
HS Diploma or Equivalent	313	42.0%	466	42.6%	2,161	31.0%	2,770	27.2%
Some Postsecondary	259	34.8%	386	35.3%	2,771	39.7%	3,743	36.7%
AA, BA, and Above	60	8.1%	114	10.4%	1,755	25.2%	3,316	32.5%
		\$10,551		\$9,730		\$32,112		\$38,716

Control Groups

Control groups are constructed via propensity score matching techniques from pools of individuals from the following groups:

- Individuals who applied for Unemployment Insurance benefits in the same period that our treatment groups were entering programs
- Individuals who set up a user account at a WorkForce Center or online at www.MinnesotaWorks.net and last used these basic services during the same timeframe

Control groups were constructed from these pools because 1) individuals in these groups likely experienced a recent job loss and/or showed interest in employment and training services, 2) administrative data on these individuals was available and covered a large number of individuals from which a control group could be constructed, and 3) earlier evaluations we reviewed used the same pools and made a strong case for their appropriateness.

Unemployment Insurance Applicants

The first control group used in this analysis is made up of individuals who applied for Unemployment Insurance benefits. The Unemployment Insurance program provides a temporary partial wage replacement to workers who

become unemployed through no fault of their own. It is a stabilizer during economic downturns and helps maintain an available, skilled workforce. Workers may be paid up to 50 percent of their average weekly wage, subject to a state maximum (currently \$597) for up to 26 weeks.

Primary customers are the applicants who apply for benefits and employers who are subject to the unemployment insurance law. The program determines applicant eligibility for benefits, makes weekly benefit payments to eligible applicants, and—for applicants not attached to previous employment—referrals to WorkForce Centers for job-seeking assistance, job training, or other help. The program determines if employers are subject to the law, collects revenues, audits employer and applicant accounts to ensure proper payments are made, and provides impartial due process hearings for applicants and employers who appeal initial decisions. The unemployment insurance system is based on an insurance model, with employers' premiums based on their "experience" with the system; those with more layoffs have a higher tax rate.

Registrants at WorkForce Centers and at MinnesotaWorks.net

The second control group used in this analysis is made up of individuals who have registered an account at a WorkForce Center or with MinnesotaWorks.net, but who have not enrolled in an eligibility-based program.

Any individual can register an account and use the resource area at one of Minnesota's physical WorkForce Center locations. Customers are provided with access to useful websites, software, and other job, career, or educational resources. They are also informed of upcoming seminars, job fairs, and other events.

MinnesotaWork.net is an internet-based self-service system where employers and job seekers can find each other. Registration is encouraged because it allows full access to all the features of the system. Job seekers can post up to five resumes to be searched by employers. They can also search for job openings and be contacted by e-mail when new job postings meeting their search criteria are found by the system. Employers can post job openings. They can also search for job candidates, recruit job seekers online, and elect to receive emails when new resumes are found that match their requirements. MinnesotaWorks.net is a service provided by the Minnesota Department of Employment and Economic Development (DEED).

Interpreting Statistical Tables

Here's a quick guide on what to look for, and how to think about some key concepts.

FIGURE 6: INTERPRET STATISTICAL TABLES LIKE A PRO

DIFFERENCE-IN-DIFFERENCE

The difference-in-difference (DiD) result is a primary result to pay attention to. It tells us our estimated net impact roughly in terms of percent change. The corresponding percent change is given by the equation:

$$\% \Delta = e^{DiD} - 1$$

The difference-in-difference is calculated by **1** measuring the difference between the treatment and control group in the baseline period (before program entrance), **2** measuring the same difference in the follow-up period, and then taking the difference between the two: **2 - 1**.

See Figure 4 on page 10 to learn more about the difference-in-difference approach.

*Note: On its own, this number doesn't necessarily mean anything. Standard errors and *'s denoting statistical significance tell us just how confident we can be in our estimate.*

Cohort	Baseline Log Annual Earnings	Follow-Up Periods		
		Short Term Log Annual Earnings	Medium Term Log Annual Earnings	Long Term Log Annual Earnings
WIA Adult 2007-2008	9.375 (0.056)	9.651 (0.042)	9.553 (0.052)	9.503 (0.059)
Control Group	9.237 (0.016)	9.084 (0.032)	9.102 (0.037)	9.314 (0.035)
Difference	0.138 (0.058) 1	0.567 (0.053)	0.451 (0.064)	0.189 (0.069) 2
Difference-in-Difference		0.429 *** (0.078)	0.313 *** (0.086)	0.052 (0.090)
Net Impact (Percent Change in Annual Earnings)		+53.6%	+36.8%	+5.3%
Net Impact (Change in Annual Earnings, in Dollars)		+\$6.316	+\$4.333	+\$629
WIA Adult 2009-2010	9.189 (0.042)	9.253 (0.037)	9.455 (0.039)	9.674 (0.039)
Control Group	9.135 (0.013)	8.927 (0.029)	9.087 (0.032)	9.341 (0.028)
Difference	0.054 (0.044)	0.326 (0.047)	0.368 (0.050)	0.333 (0.048)
Difference-in-Difference		0.272 *** (0.065)	0.314 *** (0.067)	0.279 *** (0.065)
Net Impact (Percent Change in Annual Earnings)		+31.3%	+36.9%	+32.2%
Net Impact (Change in Annual Earnings, in Dollars)		+\$3.060	+\$3.611	+\$3.150

NET IMPACTS

The difference-in-difference (DiD) result is translated into net impacts in terms of percent change and dollars, using the equation above.

Example. Suppose we see a log earnings difference-in-difference of 0.052. This means we estimate the program was responsible for an average earnings increase of $e^{0.052} - 1 = 0.053$ or 5.3% for its participants, compared to the control group. In other words, participants' earnings are 5.3% higher on average than they would have been had they not participated in the program.

Using this percent change we can compute the dollar value of the impact, based on participants' average earnings in the baseline.

STANDARD ERROR

Denoted in parentheses, standard errors are an indication of the reliability of the estimated results directly above them.

In fact, a standard error tells us that we can be 95% confident that our estimated result is within two (actually 1.96) standard errors from the actual outcome.

Example. Suppose we have:

- a net earnings impact of 0.272 (a 31.3% increase), and
- a standard error of 0.065 (or 6.5%)

Then we can be 95% confident that the actual outcome is:

- at least 18.6% ($31.3\% - 1.96 \times 6.5\%$) and
- at most 44.0% ($31.3\% + 1.96 \times 6.5\%$)

In other words, a small standard error indicates that the result is a more accurate reflection of the actual outcome.

Note: A larger sample size will normally result in a smaller standard error. Conversely, small sample sizes mean higher standard errors, and less confident that our estimated result is accurate.

LOG EARNINGS or LOG BENEFITS

Instead of working with earnings and benefits in terms of dollars directly, we analyze the natural log of those amounts. This allows us to express the results in terms of percentage change (rather than changes in dollar amounts).

You can convert a log amount back into a dollar amount with the equation:

$$\text{\$ amount} = e^{\log \text{ amount}}$$

Example. Suppose we see a log annual earnings of 9.253. Then we know that the corresponding dollar amount is $e^{9.253} = \$10,436$. In Excel, you could use the formula $=\text{exp}(9.253)$.

Note: Setting up our evaluation design this way carries with it an implicit assumption: that we expect changes in earnings and benefits to be proportional to their initial levels. In other words, we expect that higher earners might expect to see bigger earnings impacts in terms of actual dollars. A 10% increase is larger if we start with \$20,000 than if we start with \$5,000.

STATISTICAL SIGNIFICANCE

Denoted by the number of *'s you see next to a final result, the level of statistical significance tells us how confident we can be that our estimated result is reliable (i.e. actually attributable to the program, as opposed to chance or statistical noise).

- * means that there is a 10% likelihood that the result is due to chance.
- ** means that there is a 5% likelihood that the result is due to chance.
- *** means that there is only a 1% likelihood that the result is due to chance.

In other words, the more *'s the better. Results that lack statistical significance are more unreliable; you can look at the standard error to get a better sense.

Note: Statistical significance does not tell us anything about the magnitude (size) of an effect. In everyday use, "significant" can mean "big", but in statistical language "significant" is used to describe how real or actual a given result is likely to be.

Assessing the Treatment-Control Group Match

Before we can estimate the net impacts of programs, we must first ensure that we have properly-matched treatment and control groups. We want to make sure that our program participants (the treatment group) are very similar to the non-participants (the control group) we are comparing them against.

Treatment and control groups were matched across a number of key characteristics:

- Past employment and earnings
- Gender
- Age
- Race
- Veteran status
- Highest level of education attained
- Geography
- Past enrollment in public benefits programs

Past employment and earnings are particularly important because they shed light on factors that are not directly recorded in the data, such as individual-level productivity and motivation.

Table 2 shows how well our treatment and control groups are matched using our propensity score matching technique. For each cohort, you can see:

- A summary of the cohort's make-up with regard to key demographic variables
- A summary of the cohort's control group across the same variables
- The difference between the two (we want this difference to be as small as possible)
- The "p-value" for the difference, which tells us if the difference is statistically significant (denoted by *'s). In this case, we do not want the differences to be statistically significant.²³

As you can see, our treatment and control groups are generally well-matched. The most notable exception is with regard to the pre-program earnings of the Dislocated Worker 2007-2008 cohort and its control group, which have a statistically significant difference of roughly 10 percent (-0.102). Differences in geography, average level of education, and pre-program cash benefits are also present. But overall, our matches are robust enough to produce statistically significant estimates of program net impacts, as seen in the following tables.

²³ In other words, we want any differences we observe to be most likely due to chance, and not to real differences between the groups. See Figure 6: Interpret Statistical Tables like a Pro on page 17 for a discussion of statistical significance.

TABLE 2: PRE-TREATMENT MATCH DIAGNOSTICS

Cohort	Mean Weighted Variables								
	Cohort Size (N)	Log Annual Earnings	% Female	% Age	% White	% Veteran	Education Level	% Seven-County Metro	Quarterly Cash Benefits
WIA Adult 2007-2008		9.234							
p-value									
p-value									
p-value									
Control Group	19,060	10.568	43.5%	44.7	85.7%	8.3%	4.10	65.5%	\$26
Difference						0.4%	-0.03	0.4%	\$6
p-value				0.162	0.6065	0.2164	0.0115**	0.4402	

Notes:

- In matching treatment and control groups, the goal is that any observed differences are not statistically significant.
- Means and t-tests are estimated by linear regression
- Inference: *** p<0.01; ** p<0.05; * p<0.1
- Education level is a weighted average using the following coding: 0 - other and/or missing; 1 - grade school or less; 2 - incomplete high school; 3 - high school diploma or equivalent; 4 - some postsecondary education; 5 - associates or bachelor degree; 6 - masters or doctorate
- "% white" is calculated based on those providing information

Annual Earnings

Our analysis shows that both programs have large, statistically significant net impacts on annual earnings overall (see Table 3). Over the medium term (five to eight quarters after program entrance) the WIA Adult program was responsible for an increase in annual earnings of roughly 31 percent. In other words, program participants had earnings 31 percent higher than they would have if they not participated in the program. This percentage increase translates into positive impacts of \$4,333 and \$3,611 in annual earnings for the 2007-2008 and 2009-2010 cohorts, respectively.

The Dislocated Worker program was also responsible for earnings increases in the medium term of 31.7 percent and 13.2 percent for the 2007-2008 and 2009-2010 cohorts, respectively. Translated into dollars, this means participants earned on average \$10,349 and \$5,121 more, respectively, than they would have if they had not participated in the program.²⁴

Trends Over the Short-, Medium-, and Long-Term

There is no clear trend across all cohorts with regard to whether net impacts on earnings grow or dissipate over time. For both 2007-2008 cohorts, we observe that impacts on earnings decrease as time goes on from the short-term to the long-term. For the 2009-2010 cohorts, impacts either stayed fairly constant (WIA Adult) or grew (Dislocated Worker) as time went on. These differing trends may be related to the changing state of the economy over the given periods. The 2007-2008 cohorts exited their programs as the Great Recession was starting (the recession officially

²⁴ It is important to note that the net impacts in this report measure the difference between the observed participant outcome and what we estimate *would have happened if the participant had not participated*. In other words, they do not measure before-and-after changes in outcomes. So, for example, it is not precise to say only that participant earnings *went up* or *went down*. In fact, a program can have a positive net impact even if its participants have lower earnings than they did before participation. In this case, the control group would have had even lower earnings. The comparison between the treatment and control is what matters.

began December 2007), while the 2009-2010 cohorts exited as the recession was ending (the recession officially ended June 2009).

Net Impacts Across Various Populations

Table 4 details how the programs impacted the annual earnings of various populations over the medium term. A few trends stand out.

- First, the programs appear to impact participants differently with regard to gender. The WIA Adult program shows much larger, statistically significant impacts for men, while the Dislocated Worker program shows larger, statistically significant impacts for women.
- Second, three of the four cohorts (all but WIA 2007-2008) show greater, statistically-significant impacts for individuals living in the seven-county metro area than for individuals living in Greater Minnesota.
- Third, the WIA Adult program produced notably larger, statistically significant net impacts for African Americans (increases in annual earnings of 93.7 percent for the 2007-2008 cohort and 98.8 percent for the 2009-2010 cohort) than for white participants.

TABLE 3: NET IMPACTS ON AVERAGE ANNUAL EARNINGS

See Figure 6 on page 17 for a guide to interpreting these results.

Cohort	Baseline	Follow-Up Periods		
		Short-Term	Medium Term	Long Term
		Log Annual Earnings	Log Annual Earnings	Log Annual Earnings
WIA Adult 2007-2008	9.375 (0.056)	9.651 (0.042)	9.553 (0.052)	9.503 (0.059)
Control Group	9.237 (0.016)	9.084 (0.032)	9.102 (0.037)	9.314 (0.035)
<i>Difference</i>	0.138 (0.058)	0.567 (0.053)	0.451 (0.064)	0.189 (0.069)
Difference-in-Difference		0.429 *** (0.078)	0.313 *** (0.086)	0.052 (0.090)
Net Impact (Percent Change in Annual Earnings)		+53.6%	+36.8%	+5.3%
Net Impact (Change in Annual Earnings, in Dollars)		+\$6,316	+\$4,333	+\$629
WIA Adult 2009-2010	9.189 (0.042)	9.253 (0.037)	9.455 (0.039)	9.674 (0.039)
Control Group	9.135 (0.013)	8.927 (0.029)	9.087 (0.032)	9.341 (0.028)
<i>Difference</i>	0.054 (0.044)	0.326 (0.047)	0.368 (0.050)	0.333 (0.048)
Difference-in-Difference		0.272 *** (0.065)	0.314 *** (0.067)	0.279 *** (0.065)
Net Impact (Percent Change in Annual Earnings)		+31.3%	+36.9%	+32.2%
Net Impact (Change in Annual Earnings, in Dollars)		+\$3,060	+\$3,611	+\$3,150
Dislocated Worker 2007-2008	10.395 (0.011)	9.952 (0.015)	10.094 (0.016)	10.113 (0.017)
Control Group	10.481 (0.008)	9.702 (0.020)	9.905 (0.018)	9.981 (0.017)
<i>Difference</i>	-0.086 (0.014)	0.25 (0.025)	0.189 (0.024)	0.132 (0.024)
Difference-in-Difference		0.337 *** (0.028)	0.275 *** (0.028)	0.218 *** (0.028)
Net Impact (Percent Change in Annual Earnings)		+40.1%	+31.7%	+24.4%
Net Impact (Change in Annual Earnings, in Dollars)		+\$13,102	+\$10,349	+\$7,964
Dislocated Worker 2009-2010	10.566 (0.007)	9.686 (0.014)	10.178 (0.012)	10.326 (0.011)
Control Group	10.568 (0.004)	9.718 (0.012)	10.056 (0.010)	10.182 (0.009)
<i>Difference</i>	-0.002 (0.008)	-0.032 (0.019)	0.122 (0.015)	0.143 (0.014)
Difference-in-Difference		-0.03 (0.020)	0.124 *** (0.018)	0.145 *** (0.016)
Net Impact (Percent Change in Annual Earnings)		-3.0%	+13.2%	+15.6%
Net Impact (Change in Annual Earnings, in Dollars)		-\$1,147	+\$5,121	+\$6,053

- Timeframes: Baseline 4-8 quarters before entrance; short-term 1-4 quarters post-entrance; medium-term 5-8 quarters; long-term 9-12 quarters
- Robust Standard Errors are in parentheses
- Means and Standard Errors are estimated by linear regression
- Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

The following figures show the average annualized earnings of the treatment and control groups prior to propensity score matching (PSM). Thus the overall picture these charts provide may differ from that of the results tables, which were computed after PSM.

FIGURE 7: AVERAGE ANNUALIZED EARNINGS: WIA ADULT 2007-2008 COHORT vs. UNMATCHED CONTROL GROUP

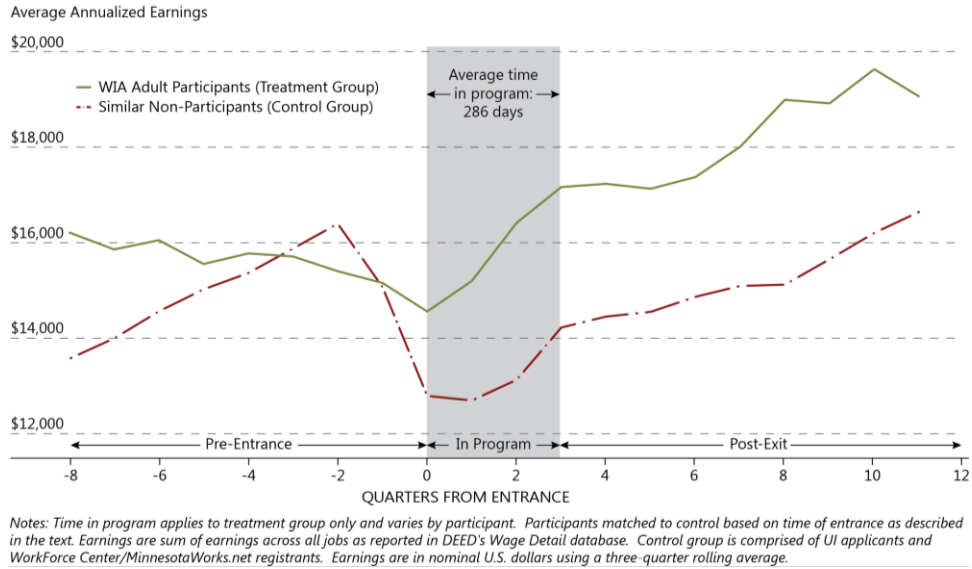
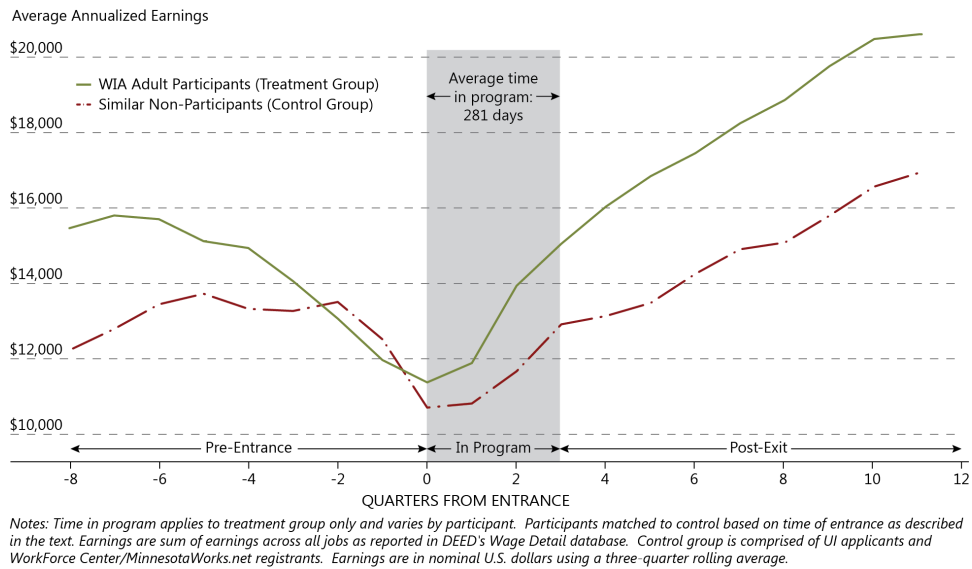
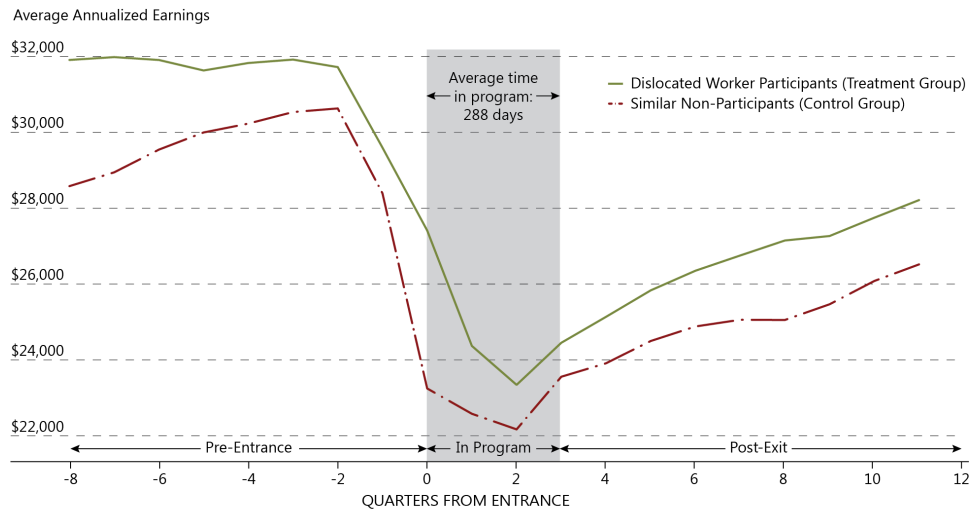


FIGURE 8: AVERAGE ANNUALIZED EARNINGS: WIA ADULT 2009-2010 COHORT vs. UNMATCHED CONTROL GROUP



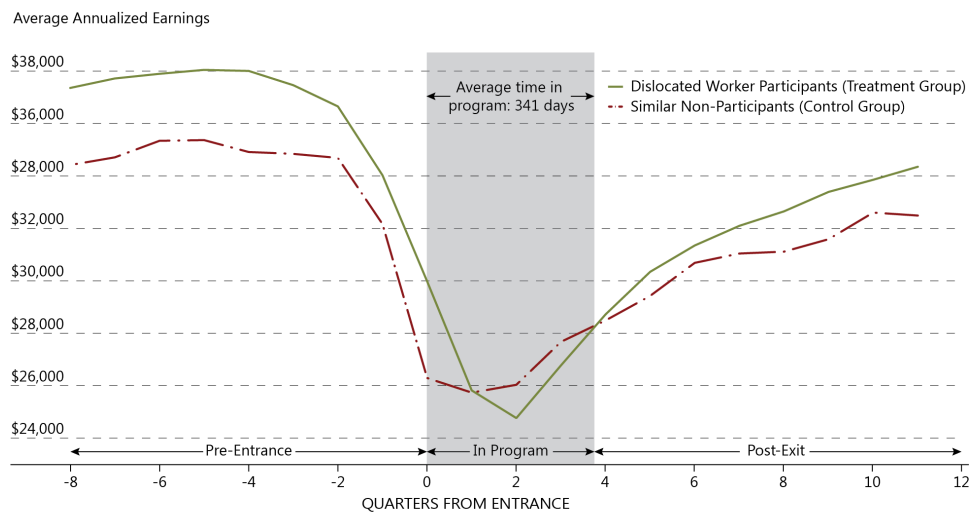
The following figures show the average annualized earnings of the treatment and control groups prior to propensity score matching (PSM). Thus the overall picture these charts provide may differ from that of the results tables, which were computed after PSM.

FIGURE 9: AVERAGE ANNUALIZED EARNINGS: DISLOCATED WORKER 2007-2008 COHORT VS. UNMATCHED CONTROL GROUP



Notes: Time in program applies to treatment group only and varies by participant. Participants matched to control based on time of entrance as described in the text. Earnings are sum of earnings across all jobs as reported in DEED's Wage Detail database. Control group is comprised of UI applicants and WorkForce Center/MinnesotaWorks.net registrants. Earnings are in nominal U.S. dollars using a three-quarter rolling average.

FIGURE 10: AVERAGE ANNUALIZED EARNINGS: DISLOCATED WORKER 2009-2010 COHORT VS. UNMATCHED CONTROL GROUP



Notes: Time in program applies to treatment group only and varies by participant. Participants matched to control based on time of entrance as described in the text. Earnings are sum of earnings across all jobs as reported in DEED's Wage Detail database. Control group is comprised of UI applicants and WorkForce Center/MinnesotaWorks.net registrants. Earnings are in nominal U.S. dollars using a three-quarter rolling average.

Employment Likelihood

Our analysis shows that both programs have sizable net impacts on employment likelihood; participants were more likely to be employed than they would have been if they had not participated in programs (see Table 5). Over the medium term (five to eight quarters after program entrance), we observe that the WIA Adult program is responsible for positive employment impacts of 14.5 percent and 15 percent for the 2007-2008 and 2009-2010 cohorts, respectively. The Dislocated Worker program was responsible for employment impacts of 13.3 percent and 8.2 percent for the 2007-2008 and 2009-2010 cohorts, respectively.²⁵

Trends Over the Short-, Medium-, and Long-Term

There is no clear trend across all cohorts with regard to whether net impacts on employment grow or dissipate over time. As with earnings, we observe differences that may be related to the changing state of the economy over the given cohort periods. For the 2007-2008 cohorts, we observe that impacts on employment either decrease over time (from 18.5 percent to 11.5 percent for WIA Adult) or stay flat (at 12.1 percent for Dislocated Worker). For the 2009-2010 cohorts, impacts grew with time, from 14.7 percent to 15.6 percent for WIA Adult and from 3.1 percent to 8.9 percent for Dislocated Worker.)

TABLE 5: CHANGE IN EMPLOYMENT LIKELIHOOD

See Figure 6 on page 17 for a guide to interpreting these results.

Cohort	Baseline	Follow-Up Periods		
		Short-Term	Medium Term	Long Term
	% Likelihood of Employment	% Likelihood of Employment	% Likelihood of Employment	% Likelihood of Employment
WIA Adult 2007-2008	85.6%	87.0%	86.7%	76.3%
Control Group	100.0%	82.9%	86.6%	79.2%
<i>Difference</i>	-14.4%	4.1%	0.1%	-2.9%
Difference-in-Difference (Percent Change in Employment)		+18.5%	+14.5%	+11.5%
WIA Adult 2009-2010	83.2%	77.0%	85.5%	80.7%
Control Group	100.0%	79.1%	87.2%	81.8%
<i>Difference</i>	-16.8%	-2.1%	-1.8%	-1.1%
Difference-in-Difference (Percent Change in Employment)		+14.7%	+15.0%	+15.6%
Dislocated Worker 2007-2008	91.3%	72.5%	74.2%	71.4%
Control Group	99.3%	68.4%	68.9%	67.3%
<i>Difference</i>	-8.0%	4.0%	5.3%	4.1%
Difference-in-Difference (Percent Change in Employment)		+12.1%	+13.3%	+12.1%
Dislocated Worker 2009-2010	93.2%	63.0%	75.7%	77.6%
Control Group	100.0%	66.7%	74.4%	75.6%
<i>Difference</i>	-6.8%	-3.7%	1.4%	2.1%
Difference-in-Difference (Percent Change in Employment)		+3.1%	+8.2%	+8.9%

- Timeframes: Baseline 4-8 quarters before entrance; short-term 1-4 quarters post-entrance; medium-term 5-8 quarters; long-term 9-12 quarters
- Difference-in-difference estimates are based on matched samples, but have not been formally tested for statistical significance.

²⁵ It should be noted that the employment results described here are based on a simplified difference-in-differences analysis using matched treatment and control groups; the results have not been regression-adjusted to further control for demographic characteristics or to provide formal tests of statistical significance.

The following figures show the average annualized earnings of the treatment and control groups prior to propensity score matching.

FIGURE 11: EMPLOYMENT LIKELIHOOD: WIA ADULT 2007-2008 COHORT VS. UNMATCHED CONTROL GROUP

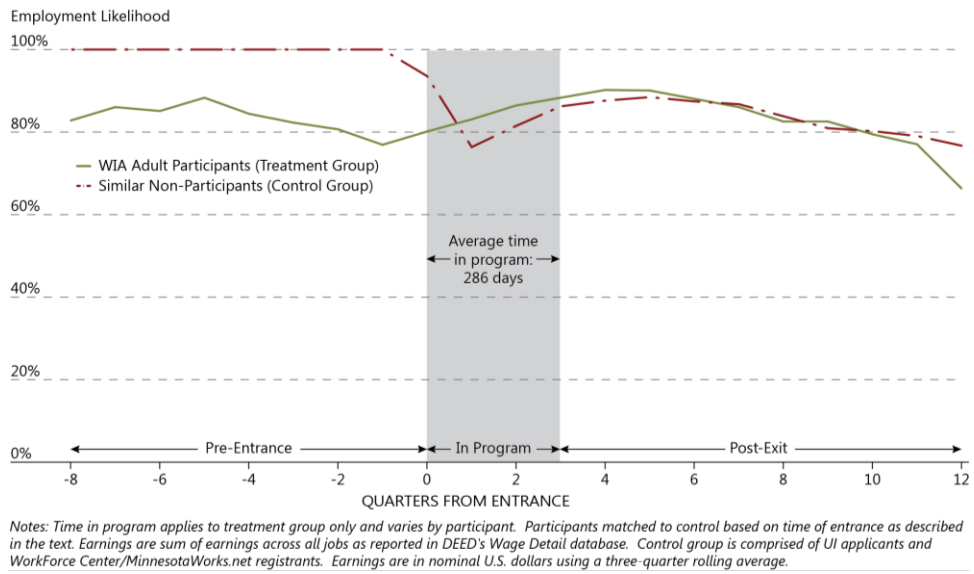
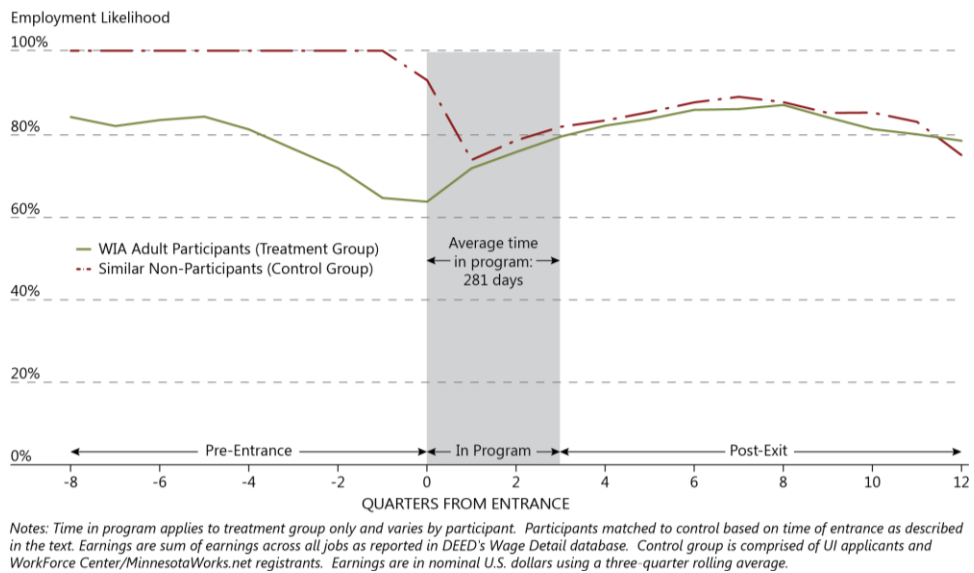


FIGURE 12: EMPLOYMENT LIKELIHOOD: WIA ADULT 2009-2010 COHORT VS. UNMATCHED CONTROL GROUP



The following figures show the average annualized earnings of the treatment and control groups prior to propensity score matching.

FIGURE 13: EMPLOYMENT LIKELIHOOD: DISLOCATED WORKER 2007-2008 COHORT VS. UNMATCHED CONTROL GROUP

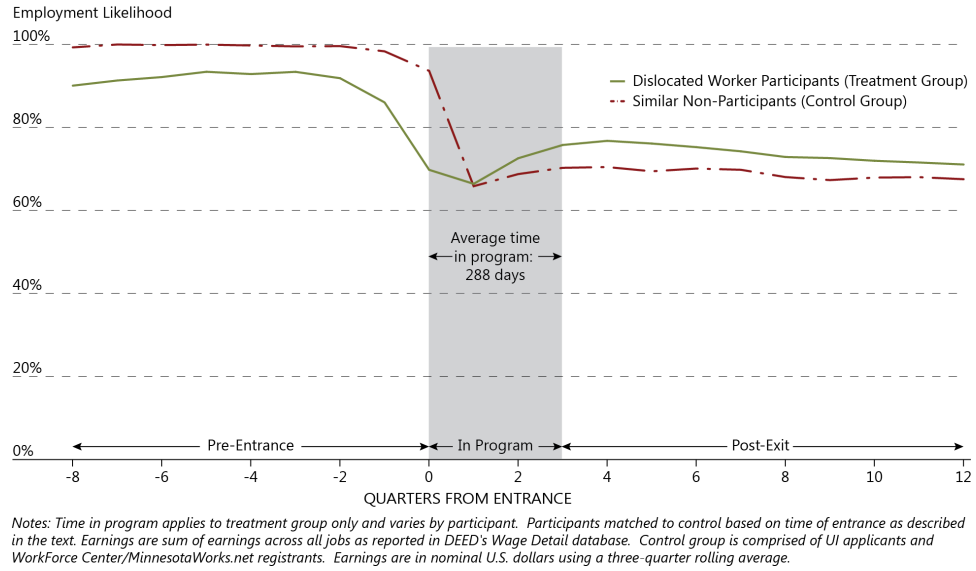
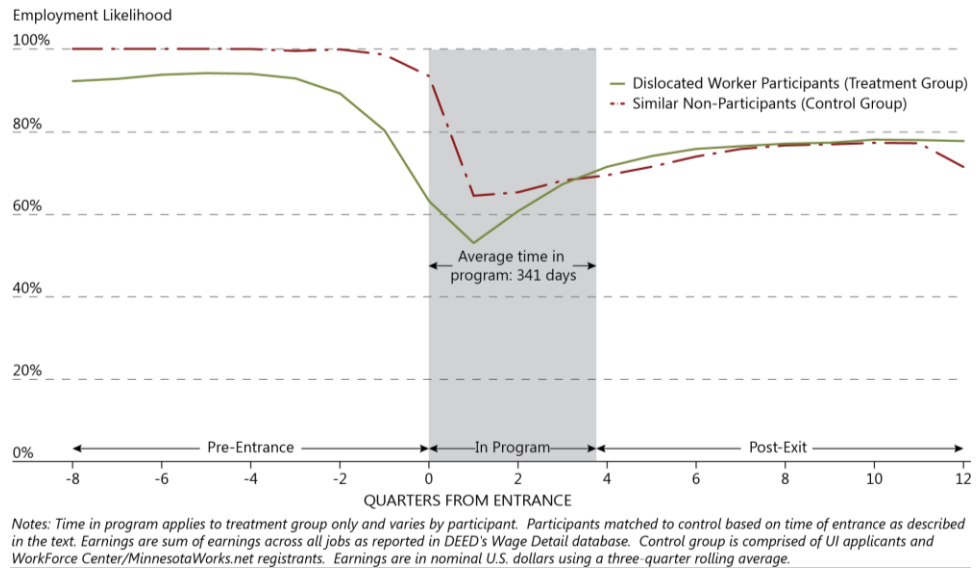


FIGURE 14: EMPLOYMENT LIKELIHOOD: DISLOCATED WORKER 2009-2010 COHORT VS. UNMATCHED CONTROL GROUP



Net Impacts Across Various Populations

Table 6 details how the programs impacted the employment likelihood of various populations. A few trends are apparent:

- First, the programs appear to impact participants differently with regard to gender. The WIA Adult program shows larger impacts for men, while the Dislocated Worker program shows larger impacts for women.
- Second, three of the four cohorts (all but WIA 2007-2008) show greater impacts for individuals living in the seven-county metro area than for individuals living in Greater Minnesota.
- Third, the WIA Adult program produced notably larger impacts for African Americans (increases in employment likelihood of 19.7 percent for the 2007-2008 cohort and 21.9 percent for the 2009-2010 cohort) than for white participants.
- Fourth, with regard to age, employment impacts were consistently lowest (though still positive) for the group aged 25-44.

TABLE 6: CHANGE IN EMPLOYMENT LIKELIHOOD OVER THE MEDIUM TERM, DISAGGREGATED

See Figure 6 on page 17 for a guide to interpreting these results.

	WIA Adult 2007-2008 Difference-in-Difference (% Change in Employment Likelihood)	WIA Adult 2009-2010 Difference-in-Difference (% Change in Employment Likelihood)	Dislocated Worker 2007-2008 Difference-in-Difference (% Change in Employment Likelihood)	Dislocated Worker 2009-2010 Difference-in-Difference (% Change in Employment Likelihood)
18-24				
25-44				
45-54				
55-64				
BY GENDER				
Female	+12.7%	+15.2%	+16.2%	+8.5%
Male	+16.3%	+17.1%	+15.2%	+6.6%
African American				
Greater Minnesota		+13.1%	+10.2%	+7.9%
Seven-County Metro	+16.0%	+20.2%	+16.8%	+8.4%
Less than HS	+21.1%	+22.9%	+20.8%	+12.6%
HS Diploma or Equiv.	+15.7%	+14.6%	+10.5%	+11.9%
Some Postsecondary	+12.4%	+15.0%	+10.3%	+8.9%
AA, BA, and Above	+16.4%	+11.0%	+13.2%	+7.7%

- Timeframes: Baseline 4-8 quarters before entrance; medium-term 5-8 quarters
- Difference-in-difference estimates are based on matched samples, but have not been formally tested for statistical significance.

Quarterly Cash Benefits (MFIP and SNAP)

Our analysis of net impacts on the amount of quarterly cash benefits (MFIP and SNAP) received yielded far fewer statistically significant results, in large part because of the smaller number of individuals in both the treatment and control groups receiving these benefits in the first place (see

Table 7).

The most apparent, statistically significant result overall is that the Dislocated Worker program reduces the amount of cash benefits received, albeit by very tiny amounts (less than 21 cents per quarter). For WIA Adult, the 2007-2008 cohort showed a negative impact in terms of cash benefits received (though the results were not statistically significant) and the 2009-2010 cohort showed a positive impact. While the WIA Adult impacts were larger in magnitude (as high as \$180 per quarter), only the short-term impact on the 2009-2010 cohort was statistically significant.

Net Impacts Across Various Populations

Table 8 details how the programs impacted the quarterly benefits received by various populations over the medium term. Again, there were very few statistically significant results, but we do find the following:

- For the WIA Adult 2007-2008 cohort, we find a statistically significant increase in quarterly cash benefits for African Americans and a statistically significant decrease for those with less than a high school diploma.
- For the WIA Adult 2009-2010 cohort, we find statistically significant increases in quarterly cash benefits for two groups: men and individuals ages 55-64.
- For the Dislocated Worker 2009-2010 cohort, we find a statistically significant increase in quarterly cash benefits for individuals aged 45-54. This increase is quite large in percentage terms (311 percent).

TABLE 7: NET IMPACTS ON AVERAGE QUARTERLY CASH BENEFITS (MFIP AND SNAP)

See Figure 6 on page 17 for a guide to interpreting these results.

Cohort	Baseline	Follow-Up Periods		
		Short-Term	Medium Term	Long Term
	Log Quarterly Benefits	Log Quarterly Benefits	Log Quarterly Benefits	Log Quarterly Benefits
WIA Adult 2007-2008	6.273 (0.116)	6.701 (0.113)	6.719 (0.138)	6.996 (0.117)
Control Group	5.773 (0.086)	6.204 (0.081)	6.387 (0.094)	6.541 (0.074)
<i>Difference</i>	0.5 (0.144)	0.497 (0.139)	0.332 (0.167)	0.455 (0.138)
Difference-in-Difference		-0.003 (0.200)	-0.168 (0.221)	-0.045 (0.200)
Net Impact (Percent Change in Quarterly Benefits)		-0.3%	-15.5%	-4.4%
Net Impact (Change in Quarterly Benefits, in Dollars)		-\$2	-\$82	-\$23
WIA Adult 2009-2010	6.28 (0.080)	7.125 (0.065)	6.956 (0.076)	6.88 (0.079)
Control Group	5.988 (0.079)	6.544 (0.111)	6.45 (0.137)	6.467 (0.291)
<i>Difference</i>	0.291 (0.112)	0.581 (0.129)	0.506 (0.157)	0.413 (0.113)
Difference-in-Difference		0.29 * (0.171)	0.215 (0.193)	0.122 (0.159)
Net Impact (Percent Change in Quarterly Benefits)		+33.6%	+24.0%	+13.0%
Net Impact (Change in Quarterly Benefits, in Dollars)		+\$180	+\$128	+\$69
Dislocated Worker 2007-2008	0.25 (0.017)	0.479 (0.024)	0.402 (0.022)	0.44 (0.023)
Control Group	0.186 (0.013)	0.489 (0.021)	0.521 (0.022)	0.571 (0.023)
<i>Difference</i>	0.064 (0.022)	-0.01 (0.032)	-0.119 (0.031)	-0.131 (0.033)
Difference-in-Difference		-0.074 * (0.038)	-0.183 *** (0.038)	-0.196 *** (0.039)
Net Impact (Percent Change in Quarterly Benefits)		-7.1%	-16.7%	-17.8%
Net Impact (Change in Quarterly Benefits, in Dollars)		-\$0.09	-\$0.21	-\$0.23
Dislocated Worker 2009-2010	0.171 (0.010)	0.437 (0.016)	0.439 (0.017)	0.456 (0.017)
Control Group	0.116 (0.006)	0.439 (0.012)	0.495 (0.013)	0.52 (0.013)
<i>Difference</i>	0.055 (0.012)	-0.002 (0.020)	-0.056 (0.021)	-0.064 (0.021)
Difference-in-Difference		-0.057 ** (0.024)	-0.111 *** (0.024)	-0.119 *** (0.025)
Net Impact (Percent Change in Quarterly Benefits)		-5.5%	-10.5%	-11.2%
Net Impact (Change in Quarterly Benefits, in Dollars)		-\$0.07	-\$0.12	-\$0.13

Notes:

- Timeframes: Baseline 4-8 quarters before entrance; short-term 1-4 quarters post-entrance; medium-term 5-8 quarters; long-term 9-12 quarters
- Robust Standard Errors are in parentheses
- Means and Standard Errors are estimated by linear regression
- Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

TABLE 8: NET IMPACTS ON AVERAGE QUARTERLY CASH BENEFITS (MFIP AND SNAP) OVER THE MEDIUM TERM, DISAGGREGATED

See Figure 6 on page 17 for a guide to interpreting these results.

	WIA Adult 2007-2008		WIA Adult 2009-2010		Dislocated Worker 2007-2008		Dislocated Worker 2009-2010	
	Log Quarterly Benefits Diff-in-Diff	Net Impact (% Change in Benefits)	Log Quarterly Benefits Diff-in-Diff	Net Impact (% Change in Benefits)	Log Quarterly Benefits Diff-in-Diff	Net Impact (% Change in Benefits)	Log Quarterly Benefits Diff-in-Diff	Net Impact (% Change in Benefits)
OVERALL	-0.168 (0.221)	-15.5%	0.215 (0.193)	+24.0%	-0.183 *** (0.038)	-16.7%	-0.111 *** (0.024)	-10.5%
BY AGE								
18-24	-0.339 (0.565)	-28.8%	0.358 (0.329)	+43.0%	0.432 (0.672)	+54.0%	0.000 (0.000)	+0.0%
25-44	-0.148 (0.267)	-13.8%	0.229 (0.224)	+25.7%	-0.038 (0.262)	-3.7%	-0.068 (0.188)	-6.6%
45-54	-0.383 (0.550)	-31.8%	0.203 (0.315)	+22.5%	0.941 (0.668)	+156.3%	1.414 ** (0.594)	+311.2%
55-64	-0.137 (0.564)	-12.8%	0.58 * (0.318)	+78.6%	0.952 (0.687)	+159.1%	0.000 (0.000)	+0.0%
BY GENDER								
Female	-0.067 (0.245)	-6.5%	0.109 (0.149)	+11.5%	-0.281 (0.211)	-24.5%	0.026 (0.230)	+2.6%
Male	-0.368 (0.542)	-30.8%	0.697 ** (0.338)	+100.8%	0.195 (0.388)	+21.5%	0.107 (0.248)	+11.3%
BY RACE								
African American	0.225 *** (0.434)	+25.2%	0 (0.326)	0.0%	0.166 (0.330)	+18.1%	-0.124 (0.357)	-11.7%
White	0.257 (0.268)	22.7%	0.01 (0.194)	1.0%	0.34 (0.262)	28.8%	0.023 (0.234)	+2.3%
	-0.077 (0.317)	-7.4%	0.174 (0.151)	+19.0%	-0.255 (0.316)	-22.5%	-0.114 (0.263)	-10.8%
	0.038 (0.326)	3.7%	0.076 (0.283)	+7.9%	0.143 (0.251)	+15.4%		
BY EDUCATION LEVEL								
Less than HS	-0.608 *** (0.608)	-45.6%	-0.251 (0.393)	-22.2%	0.426 (0.460)	+53.1%	0.014 (0.424)	+1.4%
HS Diploma or Equiv.	0.063 (0.321)	+6.5%	0.33 (0.218)	+39.1%	-0.32 (0.374)	-27.4%	-0.058 (0.321)	-5.6%
Some Postsecondary	0.335 (0.340)	+39.8%	0.123 (0.270)	+13.1%	-0.239 (0.441)	-21.3%	0.105 (0.278)	+11.1%
AA, BA, and Above	0.101 (1.344)	10.6%	-0.726 (0.904)	-51.6%	-0.998 (0.784)	-63.1%	0.018 (0.790)	+1.8%

- Timeframes: Baseline 4-8 quarters before entrance; medium-term 5-8 quarters
- Robust Standard Errors are in parentheses
- Means and Standard Errors are estimated by linear regression
- Inference: *** p<0.01; ** p<0.05; * p<0.1

Limitations and Caveats

Thoughtful implementation and interpretation of net impact analyses will ensure their value. For the net impact framework to have a true impact, it has to be credible, consulted thoughtfully, and useful to a wide array of audiences. Here are five lessons we have learned in developing the net impact framework:

1. Net impact is just one tool of many necessary to understand the value of workforce programs.

Every tool has its limitations. While the framework provides important data-driven insights, it does not tell us everything we need to know to understand complex programs and diverse participants. It is one of several tools decision makers should use to assess workforce strategies and investments. Those using this framework should understand its purpose and limitations, and use it in the context of other tools, considerations, and available information.²⁶

2. Good performance looks different for different programs.

A standardized net impact methodology invites comparisons between programs that vary in terms of who they serve, their programmatic goals, and their local conditions. Impacts across programs may vary significantly, but differences don't necessarily imply that one program is better than another. The goal is to encourage the right kinds of comparisons. Ideally, a program is compared to itself over time, to other very similar programs, or to impact targets that take into account the many particulars of any given program.

3. Results should be timely and responsive.

Rigorous net-impact analyses require years to collect the necessary data. This creates major lags in the feedback loop, making important insights harder to discern and react to in a timely fashion. This is a challenge particularly to service providers, who are interested in using real-time data to understand and improve their performance. Leading indicators that help predict longer-term outcomes can be used to address this issue.

4. Performance metrics are most useful when they are actively integrated into ongoing continuous improvement efforts.

Many one-time net impact evaluations are valuable for a time, only to end up as footnotes. The information and insights generated by the framework do not themselves lead to continuous improvement, but are instead the start of a more informed conversation about what works and for whom. Organizational leads, program staff, and broader cross-functional teams should meet at regular intervals to review net impact findings and to develop strategies and plans for continuous improvement.

5. Investments in data systems can pay large dividends in the long term.

Net impact evaluation is very data-intensive. Improvements to data management infrastructure and policy can help the evaluations run more smoothly. First, state statutes do not allow for the ongoing sharing of individual-level data across relevant state agencies for the purposes of performance measurement and continuous improvement. Currently, these types of the analyses can only be built around one-time data sharing arrangements for research purposes, requiring a partial reinvention of the wheel each time. Second, inconsistencies in how (and whether) data are reported and defined reduce the validity of standardized performance measures. Third, the pilot project has highlighted the need for more robust, integrated, and

²⁶ This could include studies of educational and employment disparities and their causes, analyses of specific skill gaps and occupations in demand around the state, and customer satisfaction surveys, among other things.

user-friendly data systems at the state level. Recent and ongoing efforts like the Workforce Data Quality Initiative (WDQI) and the Statewide Longitudinal Data System (SLEDS) are moving Minnesota in the right direction, but further support would be beneficial, particularly in terms of reducing long-term costs.

6. Net impact evaluation faces challenges when applied to small programs and populations.

The academic literature identifies a minimum sample size of at least 200.²⁷ Some of Minnesota's workforce programs fall below this participant threshold; we also confront barriers when estimating net impacts for specific populations that may be small in size. Although the empirical methodology can be applied in these cases, the resulting estimates can be unreliable.

For additional detail about technical limitations and caveats associated with the evaluation design, see page xviii.

²⁷ Liebman, Jeffrey et al. "Social Impact Bonds: Lessons Learned So Far"

Conclusion and Ideas for Further Development and Implementation

Minnesota's workforce development system continues to make strides in the use of data to aid continuous improvement. According to 2014 Minnesota Statutes 116L.98, subd. 7:²⁸

(b) The [DEED] commissioner shall contract with an independent entity to conduct an ongoing net impact analysis of the programs included in the net impact pilot project under paragraph (a) and any other programs deemed appropriate by the commissioner. The net impact methodology used by the independent entity under this paragraph must be based on the methodology and evaluation design used in the net impact pilot project under paragraph (a).

(c) By January 15, 2017, and every four years thereafter, the commissioner must report to the committees of the house of representatives and the senate having jurisdiction over economic development and workforce policy and finance the following information for each program subject to paragraph (b):

(1) the net impact of workforce services on individual employment, earnings, and public benefit usage outcomes; and

(2) a cost-benefit analysis for understanding the monetary impacts of workforce services from the participant and taxpayer points of view.

As Minnesota continues to use net impact evaluation and cost-benefit analysis to improve workforce services and participant outcomes, the framework described in this report can be further developed to (1) answer important policy and performance questions, (2) contextualize results to encourage their appropriate interpretation, and (3) make the results more useful and accessible. The GWDC advisory group identified a number of areas for further development and implementation that fell outside the scope of the initial pilot evaluation. These are detailed below.

Answering Important Policy and Performance Questions

Net Impact Results for Specific Populations

Net impact results can be disaggregated in additional ways to shed light on the outcomes of specific populations. This might include further study of:

- A wider array of racial and ethnic groups
- The long-term unemployed
- Individuals in specific industries or occupations
- Individuals requiring developmental education

Additional Outcomes

Additional outcomes, particularly those that could impact cost-benefit analyses can be evaluated. This might include further study of:

- Changes in enrollment in public healthcare programs
- Changes in recidivism / incarceration avoidance

²⁸ For more information, see: <https://www.revisor.leg.state.mn.us/statutes/?id=116L.98>

- Changes in the amount of Unemployment Insurance benefits received
- Longer-term earnings and employment outcomes (four years after program entrance and beyond).

Analysis of Specific Services and Co-Enrollments

To shed light on what works, further analysis of a wider array of programs models and specific types of services can be implemented. Considering the increasing use career pathways models and other collaborative service designs involving an array of programs and partners working together, analysis of the impacts of co-enrollment in programs may be particularly fruitful.

Cost/Benefit Analysis and Social Return on Investment

The GWDC Advisory Group has developed a framework for estimating the costs and benefits of workforce programs, using net impact evaluation results as the foundation. This framework is described in Appendix IV on page ii.

Contextualizing Results to Encourage Proper Interpretation

Good performance looks different for different programs. As net impact evaluation is applied to a wider array of programs, it is important to thoughtfully consider the broader context around programs and participants, and to encourage appropriate comparisons across programs. The Advisory Group has identified a few ways to do this:

Contextualized Goals

Statistical techniques can be used to develop net impact targets that are adjusted to account for program-specific and provider-specific factors. These targets should be used to identify useful benchmarks, encourage appropriate comparisons, and understand programs and providers in context. Such an approach could be modeled after similar efforts, including existing regression-adjusted performance target methodologies²⁹ and the Minnesota Department of Human Services Self-Sufficiency Index for Counties.

Comparing Similar Programs

To the extent that net impact results are compared across programs, these comparisons should take place primarily among programs that are similar whether in terms of their stated goals, the population served, program intensity, or other factors. A framework for defining and comparing “similar” programs would be useful in this regard; the Corporation for a Skilled Workforce’s Benchmarking Project provides some thoughtful guidance in this area.³⁰

Making Results More Useful and Accessible

Leading Indicators

Since net impact analyses can require years to collect the necessary data, insights can be difficult to discern in a timely fashion. This is a challenge particularly to service providers, who are interested in using real-time data to understand and improve their performance. Leading indicators that help predict longer-term outcomes can be used to address this issue. Statistical techniques can be used to develop leading indicators that are predictive of longer-term net impacts. These leading indicators could include near-term participant outcomes (such as placement in a job or the attainment of a certain wage) or programmatic progress points (such as completion of a training module or a score on a particular assessment).

²⁹ Eberts, Randall W., Wei-Jang Huang, and Jing Cai, “A Methodology for Setting State and Local Regression-Adjusted Performance Targets for Workforce Investment Act Programs.”

³⁰ See Miles, Marty, and Stacy Woodruff-Bolte. “Apples to Apples: Making Data Work for Community-Based Workforce Development Programs.”

Online, Interactive Dashboard

An online, interactive dashboard tool that meets high standards for usability can ensure the transparency and accessibility of net impact evaluation results and insights. Results should be presented with a high level of context and guidance for proper interpretation and use.

APPENDIX I: GWDC Net Impact Advisory Group Members

The Net Impact Advisory Group was initially convened by the GWDC in 2009; organizations listed reflect the members' affiliation at that time.

Anthony Alongi	Department of Employment and Economic Development
Paul Anton	Wilder Research
Art Berman	Twin Cities RISE!
Mark Brinda	City of Minneapolis
Susan Carter	Minnesota State Colleges and Universities
Suzanne Cerrato	Department of Employment and Economic Development
Steve Ditschler	ProAct, Inc.
Kathy Dolphin	Dolphin Group
Carol Dombek	Southwest Minnesota Private Industry Council
Terry Donovan	Department of Employment and Economic Development
Steve Erbes	Department of Employment and Economic Development
Randy Johnson	Workforce Development, Inc.
Cristine Leavitt	Department of Employment and Economic Development
Susan Lindoo	Department of Employment and Economic Development
Bryan Lindsley	Governor's Workforce Development Council
Nicholas Maryns	Governor's Workforce Development Council
Devon Meade	Greater Twin Cities United Way
Brian Paulson	Greater Twin Cities United Way
Sarah Radosevich	Minnesota Chamber of Commerce
Dr. Raymond Robertson	Macalester College
Mary Rothchild	Minnesota State Colleges and Universities
Deb Serum	Department of Employment and Economic Development
JoAnn Simser	Minnesota State Colleges and Universities
Richard Todd	Federal Reserve Bank of Minneapolis
Todd Wagner	Minnesota Department of Education / Adult Basic Education
Luke Weisberg	LukeWorks, LLC
Annie Welch	Department of Employment and Economic Development
Inez Wildwood	Governor's Workforce Development Council

APPENDIX II: Literature Review, Empirical Methodology, and Data Process

The effectiveness of training programs is an important question, so it is not surprising that there has been ample debate in the academic literature about both the appropriate approach and resulting estimates from attempts to evaluate training programs. Since our approach is firmly grounded in this literature, we briefly describe the debate surrounding the evaluation of other training programs, present some of the estimates that have emerged from this literature, and then describe our approach in detail.

The methodological debate surrounding program evaluation

The goal of the program evaluation literature is to properly determine the correct way to identify the effect of a particular program. We only briefly review this literature here. For those interested in a very thorough overview of the program evaluation literature, Imbens and Wooldridge (2009) provide an overview of different approaches to program evaluation found in this literature and discuss the estimation issues that arise with the different methods.

In theory, program evaluation is quite simple. A program's effect on an individual participant is the difference between the outcome that a participant would have experienced had (s)he not participated in the program and the outcome the participant actually experienced. Unfortunately, it is impossible to observe the outcome for same individual with and without the program at a given point in time. Therefore, the program evaluation literature centers on attempts to accurately identify the most relevant comparison for program participants.

While comparing wages (or other outcome measures) before and after program participation is tempting, such comparisons are insufficient because if a worker was motivated enough to complete the program, they may have been motivated enough to succeed without the program. An intuitive and popular alternative is to compare participants with non-participants after the program. This approach, however, is also problematic because program participants may be different (e.g. more motivated) than those who chose not to participate, and it is possible, and perhaps even likely, that such differences (and not the program *per se*) explain differences in post-participation outcomes.

To deal with these problems, the literature identifies three main approaches: randomized experiments, propensity score matching, and difference-in-difference (DiD) models.

Randomized experiments involve randomly assigning potential participants into two groups – program participants and nonparticipants – and then comparing the outcomes of the two groups. Randomized controlled trials have long been considered the 'gold standard' in program evaluation because, in theory, they control for all potential confounding variables. When participants are randomly assigned to either the experimental or control conditions, the assumption is that all potential confounding variables are also randomly distributed between the two groups. Thus the only difference between the two groups is that one received training and the other did not.

The U.S. national Job Training Partnership Act (JTPA) included a randomized evaluation component, but such experiments are rare. While theoretically rigorous, they are not without their own concerns. For example, often experimental design assumes no treatment for the control group and can include participants that enroll but do not receive treatment in the treatment group, biasing estimates of program effect downward (King and Heinrich, 2011). Others have raised the possible ethical concern that if there is a reasonable belief that the program might be helpful (even to the point that the evaluation's goal is to estimate the magnitude of the positive effect) then putting people in the control group denies them of the opportunity to improve their lives. A third concern about such experiments is the cost: they can be expensive to set up and conduct. Since a randomized evaluation requires an experimental set-up and implementation, the experimental approach is not feasible for existing programs.

Perhaps the most popular alternative falls under the umbrella of “propensity score matching” (PSM) methods. Instead of randomly assigning participants to either a treatment or control group, PSM creates control groups to match the participant group on a number of observable variables (e.g. race, sex, education status, and previous wages). If these observable characteristics are correlated with unobserved worker traits (like productivity and effort), then PSM may generate an appropriate control group. In this approach, the probability of program participation is estimated as a function of observed individual characteristics. This probability is known as the “propensity score.” Individuals who participated in the program are then matched with individuals who did not participate based on the similarity of the propensity scores. Dehejia and Wahba (1999, 2002) suggest that PSM approaches can be used to evaluate programs that were not implemented with a random assignment evaluation component. Recent attempts to evaluate programs (e.g. Hollenbeck et al. 2005) rely almost exclusively on PSM methods.

Of course, PSM approaches have attracted their share of criticism. Heckman et al. (1998) argue that if the required PSM assumptions are inaccurate, matching methods may make the selection issue worse. Smith and Todd (2005), in particular, raise several concerns about the PSM approach. One of the concerns is that Dehejia and Wahba’s (1999, 2002) results are sensitive to the specification of the equation used to estimate the propensity score and to achieve the balance between the treatment and comparison groups required for successful matching.

The third approach is known as difference-in-differences (DiD). Imbens and Wooldridge (2009) highlight Smith and Todd’s (2005) argument that the Difference-in-Difference (DiD) approach provides an alternative to the PSM approach that addresses the same concern driving the PSM approaches. The Difference-in-Difference (DiD) approach involves comparing workers in and out of the program before and after the study. The main idea is that any unobserved differences between participants and nonparticipants would affect wages prior to participating in the program. As long as these differences are constant (that is, they are unique to the individual program participants), changes in the difference between participants and non-participants before and after the study (that is, the difference in the difference between the workers) can be reasonably attributed to the program. Smith and Todd (2005) find that the DiD approach “did exhibit better performance than cross sectional estimators” in which the cross-sectional estimators include the PSM approaches described above.

Estimates of Program Effects from the Literature

Two national workforce development programs in the United States come from the Workforce Investment Act (WIA) of 1998 (King, 2004). Although mandated and funded federally, WIA operates on a state-by-state basis. State implementations of these programs contribute to the variation of program impacts. Therefore, most evaluations focus on statewide impacts of WIA programs.

Evaluations of WIA generally find positive wage trends for participants (see Imbens and Wooldridge (2009) highlight results of WIA evaluations). Estimates vary as to how much wages increase depending on participant characteristics like program and gender. An evaluation of the Job Training and Partnership Act (JTPA), a comparable predecessor to WIA, found estimates of increase in quarterly wage around \$700 for men and \$750 for women (Mueser, Troske, and Gorislavsky, 2006). Hollenbeck (2009) found estimates that ranged from \$349 to \$549 dollars. In contrast, Andersson et al. (2013) found positive impacts on earnings for participants of WIA Adult, but not for participants in the Dislocated Worker program (Andersson, Holzer, Lane, Rosenblum, and Smith, 2013). Most evaluations find smaller impacts on the wage for WIA Adult, but Andersson et al. stand out in that they finds no positive impacts from Dislocated Worker. Methodological differences in defining impacts result in these differing estimates. While Mueser et al. (2006) and Andersson et al. (2013) examine the net impact of WIA programs when compared to the broader population, Heinrich et al. (2009) evaluate the added benefit of training compared to individuals already participating in WIA.

Contributing to the abundance of conversation on this topic is the fact that, although WIA is federally funded, it operates locally and is subject to local politics and local implementation. States have power over aspects of WIA like the percentage of applicants that will be accepted into training programs. Selectivity can range from less than 40% to over 60% of applicants (Heinrich, Mueser, & Troske, 2008). Studies also vary in the data that they have access to for evaluation purposes.

The main difference in the existing program literature discussed above lies in the unit of analysis or different outcome variables. When using other non-training WIA participants as a control group, Heinrich et al. (2008) find increased earnings of around \$600 a quarter. Hollenbeck (2009) defines program impact as the net benefits to society and estimates a return on investment for workforce development of over 20% (Hollenbeck, 2009). Andersson et al. (2013), however, evaluate impact only as the added benefit from receiving training in a WIA program. Core WIA service receivers served as the control group in this analysis and the researchers found negative impact on wages for WIA Dislocated Worker.

Which Approach is Best?

Much of the debate in the program evaluation literature compares experimental design with PSM. Surprisingly, it is not clear that experimental design, while theoretically superior, generates results that are significantly different than other methods. Card et al. (2010) perform a meta-analysis of labor market evaluations and find no statistically significant difference in the results produced by experimental and observational studies. King and Heinrich (2011) build on this result and suggest that having an experimental design is not necessarily superior to observational approaches. Other studies, such as Heinrich et al. (2008) and Andersson et al. (2013) and argue that quasi-experimental methods, specifically using propensity score matching to create control groups, is not only a valid approach but also that it produces more appropriate estimates of change in wages. This assumption is supported by meta-analyses from Greenberg et al.'s (2006) analysis of 31 studies and Card et al.'s (2010) analysis of 199 studies (2010). Both conclude that experimental and quasi-experimental studies produce no statistically different estimates of program effectiveness³¹. Greenberg et al. (2006) regress programs outcomes on the type of evaluation and find that the evaluation type coefficient is small in magnitude with small standard errors. Card's studies also find a statistically insignificant effect of experimental design but with large standard errors. Therefore studies have consistently shown constant findings with slight variation between experimental and quasi-experimental. Given the increased cost in dollars and time of conducting an experimental evaluation, quasi-experimental design is the more efficient, equally accurate choice for program evaluation³².

While the DiD approach seems to perform better than PSM, PSM studies are far more common because the DiD approach is much more data-intensive than PSM approaches. The DiD approach requires comprehensive information about workers before and after the program, as well as information from a comparison group of workers before and after the program. These data requirements are rarely met, making the DiD approach relatively less common in the academic literature.

³¹ Greenberg et al. only use non-experimental designs that include a correction for selection bias while Card et al. do not specify which programs they allow into the quasi-experimental category.

³² King reflects an uncommon shift in opinion on this topic in two studies. While in 2004 he states "[d]espite enhancements in quasi-experimental methods for evaluating training programs in recent years..., the most reliable and credible evidence of the impacts of training comes from well-designed and structured experiments relying on randomly assigned treatment and control groups", his opinion shifts by 2011 (p. 66). Heinrich and King (2011) reads, "research designs used in quasi-experimental studies are generally unbiased" and furthermore "quasi-experimental designs are more likely to estimate the impact of the 'treatment on the treated' (p. 14).

Data Methods

The sample programs presented in this report are WIA Adult and Dislocated Worker (both WIA and MN). For each program, we defined two cohorts based on participant time of exit (July 2007- June 2008 and July 2009- June 2010). The data used in the analysis come from several sources. The Minnesota Department of Employment and Economic Development (DEED) is our main data provider. DEED has worker-level "Administrative Data." All employers in Minnesota are required to report individual level wage data for every employee per quarter. We can tie individual wages from multiple employers to that person. For each quarter we have accurate total wage information for every individual³³.

Wage data completeness is vital to our examination of the effect of job training because it ensures accurate matching between treatment and control groups, and it eliminates the need for programs to report (follow) the wages of participants after they exit the program³⁴.

Demographic data for training participants come from the job training programs. Once again because these questions are program-specific, we have harmonized responses to basic questions in our analysis. When collapsing these categories, we attempt to retain the highest common specificity in the definitions to give the most descriptive benefit. Finally, treatment programs report social security numbers (SSNs) as well as basic program information³⁵ for each participant.

Comparison group individuals are drawn from three separate databases: Unemployment Insurance (UI), Minnesota Works (MNW), and Customer Registration System (CRS). All three are state agencies to aid unemployed workers and therefore provide a logical pool from which to draw our comparison groups (see Heinrich et al., 2008 and Andersson et al., 2009).

Unemployment Insurance is a temporary partial wage replacement for workers who are unemployed due to layoffs. Workers may be paid up to 50 percent of their average weekly wage, subject to a state maximum (currently \$597) for up to 26 weeks. The first control group used in this analysis is made up of individuals who applied for Unemployment Insurance benefits in the same time period that our treatment groups were entering programs. Certain Unemployment Insurance applicants were excluded from control groups. This includes individuals receiving pensions, those who had re-filed for administrative or technical reasons (only first-time filers were included), those working out-of-state or for the military or federal government (since there may be gaps in wage detail data for these individuals), and those filers deemed ineligible for non-monetary reasons (e.g. individuals who may have quit or been discharged from employment).

The second control group used in this analysis is made up of individuals who registered an account at a WorkForce Center or with MinnesotaWorks.net in same timeframes that our treatment groups were entering programs cohort timeframes, but who did not enroll in an eligibility-based program.

Any individual can register an account and use the resource area at one of Minnesota's WorkForce Center locations. Customers are provided with access to useful websites, software, and other job, career, or educational resources. They will also be informed of upcoming seminars, job fairs, and other events.

MinnesotaWork.net is an internet-based self-service system where employers and job seekers can find each other. Registration is encouraged because it allows full access to all the features of the system. Job seekers can post up to

³³ Several notable exceptions include proprietors or unincorporated firms. However, almost all employees working in the state are included in this resource.

³⁴ Follow up is a challenge because programs often do not maintain contact with participants after exit.

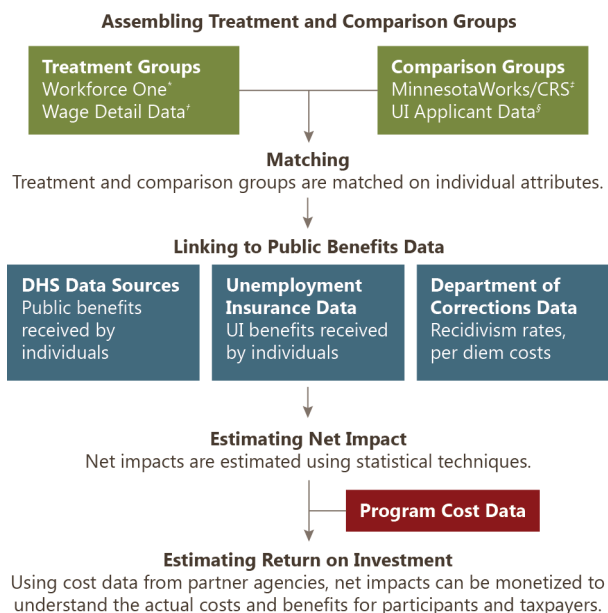
³⁵ Including start dates, exit dates, and types of training provided.

five resumes to be searched by employers. They can also search for job openings and be contacted by e-mail when new job postings meeting their search criteria are found by the system. Employers can post job openings. They can also search for job candidates, recruit job seekers online, and elect to receive emails when new resumes are found that match their requirements. MinnesotaWorks.net is a service provided by the Minnesota Department of Employment and Economic Development (DEED).

The Minnesota Department of Human Services (DHS) provided four kinds of welfare payment data: MFIP or MFMF is the Minnesota Family Investment Program (MFIP) cash grant. MFFOOD or MFFS is the MFIP food portion (Minnesota has a waiver to combine food assistance with cash assistance in MFIP. Without such a waiver, SNAP and TANF are separate programs). The third is Food Support (FS), which was recently renamed SNAP in MN to match the federal name. The final is the Diversionary Work Program (DWP), a 4-month employment-focused program that some MFIP applicants get diverted to. They can transfer to MFIP after the 4 months if eligible at that time.

The DHS data also include information on case (family) size (variable hh). For MFIP and DWP, the case is basically a nuclear family, caregiver(s) plus child(ren) minus those who are ineligible for some reason. For stand-alone FS (i.e., cases getting food assistance excluding MFIP cases getting the food portion – nearly all MFIP cases get the food portion), a household includes everyone in a household eligible for SNAP who prepare food and eat together. All these programs are distributed via the EBT card, but the food amounts can only be spend on approved food items. For more program info, go to <http://mn.gov/dhs/people-we-serve/children-and-families/economic-supports>.

FIGURE 15: THE NET IMPACT EVALUATION DATA PROCESS



¹ Workforce One is DEED's case management system for workforce programs.
² Refers to data on taxable wages and employment collected from employers on a quarterly basis and required by the federal Unemployment Insurance program.
³ These datasets provide information on those registering online or at a WorkForce Center but not receiving services from eligibility-based programs.
⁴ The Minnesota Unemployment Insurance program provides data on individuals who have lost their jobs and have applied for benefits (whether or not they receive them).

Exhibit: Data Provided by DHS

Cash/Food Assistance

- Monthly amount of MFIP cash assistance received by the case containing the individual
- Monthly amount of MFIP food assistance received by the case containing the individual
- Monthly amount of DWP (Diversionary Work Program) assistance received by the case containing the individual
- Monthly amount of Food Support (SNAP) cash benefits received by the case containing the individual

Healthcare Enrollment

- Monthly enrollment status in MinnesotaCare
- Monthly enrollment status in General Assistance Medical Care
- Monthly enrollment status in Medical Assistance
- For all enrollments, major program and eligibility type

Data Harmonization

Primary analysis began with a thorough assessment of the best methods of data harmonization. Original data sources varied in the ways they presented every necessary variable. For example, gender was coded as a binary 1 or 0 in some datasets and Male or Female in others. Although variables like gender are relatively straightforward, more nuance is required to harmonize different definitions of education levels. Generalizing education data from five different sources requires reducing all data to the least common denominator. Unfortunately this method loses some of the specificity provided by some programs for the purpose of harmony. Finally, while some programs offer specific details on the types of services provided, not all do; definitions of services vary across programs and datasets.

Once demographic data was harmonized across programs, we obtained individual wage detail data for all individuals in our sample. Wages are reported quarterly to the state by employers. Wage data are the linchpin of our analysis because they provide both the dependent variable (variable of interest) and a key control variable. Although we can control for a number of demographic factors with data provided, wages pre enrollment can be used as an instrument for factors that are challenging to quantify, like motivation.

Our analysis uses program entrance date rather than exit for several reasons. The analysis began under the assumption that most individuals who exit the program at the same time would be clustered around the same enrollment time for that program. We discovered that variation in enrollment times were much larger than anticipated. Adjusting the unit of analysis to enrollment date rather than exit date allows for similar comparisons across all programs. When a participant enters the program, we 'start the clock' (so to speak) and measure effects from that point in time. Naturally average program length needs to be accounted for when using this analysis to make generalizations across programs, but WIA Adult and Dislocated Worker both have average program lengths under one year.

The methodological approach isolates the effect of participating in workforce development programs on wages. Unlike randomized trials, the treatment group for the pilot was not randomly selected. To create a control group as similar to the treatment participants as possible, we used kernel density Difference-in-Differences estimation techniques. Kernel density matching is a subset of propensity score matching which estimates a log likelihood function of each individual participating in training given the demographics of the individuals who did participate in each program. Kernel density matching analyzes the composition of the dependent variables of the treatment group and creates a weighted value for control individuals based on their similarity to the treatment composition. The final control cohort is a weighted average of all control individuals.

Because our control samples were so large (millions of potential control individuals), we eliminated individuals with wages that were extremely different from treatment participants. On the graphs presented in the Appendix, we attempt to have matching quarters (quarters negative eight to negative four) as close in average wages as possible (see Analysis for further discussion of pre-program entrance matching).

Matching and Net Impact Estimation

We match participants and non-participants on three main factors: their demographic profile, their time of program enrollment, and their pre-enrollment wages. The actual variables used to estimate the propensity scores are gender, age, race (whether or not the individual listed their race as Caucasian), veteran status, education, residence (whether or not the individual lived in the seven-county metro area), and benefits. As described earlier, to create a control group as similar to the treatment participants as possible, we use kernel density difference-in-differences estimation techniques.

To create the matched cohort groups, we used a two-tier approach. The control group sample sizes were several orders of magnitude larger than the treatment group sample sizes. To get smaller control group samples, we first applied a randomized restriction algorithm that randomly dropped control group individuals that were sufficiently dissimilar to the treatment group means. The resulting control group means were compared to the treatment group means and the process was repeated until the control group samples were (usually) several times the size of the treatment group and the resulting means were similar. The resulting samples were then analyzed using the PSM DiD approach.

Empirical Methodology

Since this study evaluates existing programs, experimental design is not a feasible evaluation approach. Instead, this study takes advantage of a very unique dataset of matched administrative data of both program participants and a comparison group of workers. Both the participants and the comparison groups are tracked over 40 consecutive quarters ranging from 2002-2012. Since we have an uncommonly-high-quality dataset, we combine PSM with DiD using three different groups of workers as a comparison group. This section describes our approach in detail.

PSM is a general term. In practice, there are many ways to use propensity scores to match treatment and control groups. PSM can match individuals one-to-one, one-to-many, or many-to-one. Nearest neighbor matching matches all individuals to their closest PSM fit and can have multiple nonparticipants for every treatment individual or find the best matches only using nonparticipants once. In any matching procedure, tradeoffs exist between sample size and match quality. Thus the researcher determines how close of a match is close enough to balance these conflicting interests.

To create the most appropriate control group, we use an approach called Kernel density matching. While one-to-one matching can limit match possibilities, Kernel density matching uses the propensity scores as weights and calculates a weighted average of matched nonparticipants for each participant.³⁶ Thus multiple individuals combined create a 'control match' that balances characteristics and predicts wages based on that balance. In this way, better matches are given higher 'weight' and therefore more influence in the control group.

Acknowledging that our matching may not completely control for unobserved characteristics, we then subject the matched individuals to the Difference-in-Difference approach adapted from Imbens and Wooldridge (2009). Difference-in-Difference is especially appropriate in program evaluation because it eliminates both time effects and group-specific effects from the 'noise' and leaves the main program effect. Assuming matches are similar between the two groups, the treatment group's post-program wages minus the control group's post-program wages result in net program impacts. Thus, one main advantage of using Difference-in-Differences is that even if the two groups are not

³⁶ One important matching characteristic is pre program wages. Characteristics like motivation cannot be directly measured, so wages are used as an instrument. Training participants' wages have been found to dip in the periods leading up to training due, a phenomena called the 'Ashenfelter dip' (Heinrich, Mueser, & Toske, 2008) and are often followed by quarters of stagnant wages due to the opportunity cost in decreased wages of dedicating time to training (often called 'lock in' effects). Thus for matching purposes we have used eight quarters of data, from twelve to four quarters prior to program start date to avoid the Ashenfelter dip.

similar in ways that are consistent across time, it eliminates group impacts and leaves only the effect of treatment on the treated.

We further control for the possibility that serial correlation biases our results (Bertrand et al. 2004) by reducing our comparison to “pre” and “post” treatment. For the “pre” period, we take the sum of all earnings for each individual i in quarters 5-8 prior to entering the program for workers in the treatment group. For the control group, we take the sum of wages in quarters 5-8 prior to entering control group status (e.g. becoming unemployed if the control individual is drawn from the sample of unemployed workers). Our “post” period is the sum of wages in the 5th through 8th quarter after entering the status that qualifies the individual for treatment or control group status.

Our guiding equation is adapted from Imbens and Wooldridge (2009) and is implemented with Villa’s (2011) *diff* command in Stata.³⁷ Assume that we have N individuals indexed $i=1\dots N$ for whom we observe (G_i, w_{i0}, w_{i1}) where:

G_i =Group membership

\bar{w}_{i0} =Average wages for the first four quarters pre participation

\bar{w}_{i1} =Average wages for the last x quarters after enrollment

Thus the estimation equation becomes:

$$\bar{w}_{i1} - \bar{w}_{i0} = \beta + \tau DID * G_i + \epsilon_i$$

The coefficient of interest is τDID or the predicted change in average wages pre-program participation to post program participation for the treatment group. We expect this coefficient to be positive and significant. Although the group effect (G_i), would be interesting to examine, Difference-in-Differences does not estimate this variable explicitly. It only assumes that the group effects are constant to isolate the effect of the treatment on the treated. Group effects are not reported in the final Difference-in-Differences output.

The number of quarters over which to estimate program effects depends on average program participation length. Some programs are more intensive than others and will have a larger primary negative ‘lock in’ effect on wages for the first quarters after participation (Card et al., 2010).

Increasing the follow up period can improve overall impact accuracy. Wage changes are important, but just as critical to the analysis is how long increases are sustained. If an individual sees a sustained increase of \$500 a quarter for the rest of their career then the impacts of that program would be much larger over time than for a program where impacts taper off after three or four quarters.

³⁷ Villa, “DIFF: Stata Module to Perform Differences in Differences Estimation.”

APPENDIX III: References

- Abadie, Alberto. "Semiparametric Difference-in-Difference Estimators" *The Review of Economic Studies* 72, no. 1 (2005): 1-19.
- Andersson, Fredrik, Harry Holzer, Julia Lane, David Rosenblum, and Jeffrey Smith. Does federally-funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms, no. w19446 (2013). National Bureau of Economic Research.
- Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings" *Review of Economics and Statistics* 6, no. 1 (1978): 47-57.
- Athey, Susan, and Guido W. Imbens. "Identification and Inference in Nonlinear Difference-in-Difference Models" *Econometrica* 74, no. 2 (2006): 431-497.
- Basu, Anirban. "Estimating Person-Centered Treatment (PET) Effects Using Instrumental Variables." National Bureau of Economic Research Working Paper 18056 2012.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119, no.2 (2004): 249-275.
- Card, David, Jochen Kluevel, and Andrea Weber. "Active Labor Market Policy Evaluations: A Meta-Analysis." National Bureau of Economic Research Working Paper 16173, 2010.
- Cattaneo, Matias "Efficient Semiparametric Estimation of Multi-Valued Treatment Effects Under Ignorability" *Journal of Econometrics* 155 (2012): 138-154.
- Cattaneo, Matias, David Drukker, and Ashley Holland. (forthcoming) "Estimation of Multivalued Treatment Effects Under Conditional Independence" *The Stata Journal*.
- Dehejia, Rajeev and Sadek Wahba "Propensity Score Matching Methods for Nonexperimental Causal Studies" *The Review of Economics and Statistics* 84, no. 1 (2002): 151-161.
- Donald, Stephen and Kevin Lang "Inference with Difference-in-Differences and Other Panel Data" *The Review of Economics and Statistics* 89, no. 2 (2007): 221-233.
- Eberts, Randall W., Wei-Jang Huang, and Jing Cai. "A Methodology for Setting State and Local Regression-Adjusted Performance Targets for Workforce Investment Act Programs." Upjohn Institute Working Paper 13-189. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2013.
- Greenberg, David, Charles Michalopoulos, and Philip Robin. Do experimental and nonexperimental evaluations give different answers about the effectiveness of government-funded training programs? *Journal of Policy Analysis and Management*, 25 no. 3 (2006): 523-552.
- Heckman, James and Jeffrey Smith. The determinants of participation in a social program: Evidence from a prototypical job training program . National Bureau of Economic Research, no. w9818 (2003).
- Heckman, James, Hidehiko Ichimura, and Petra Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies* 64, no. 4 (1997): 605-654.

- Heckman, James and Jeffrey Smith. "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies" *The Economic Journal* 109, no. 457 (1999): 313-348.
- Heckman, James; Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. "Characterizing the Selection Bias Using Experimental Data" *Econometrica* 66, no. 5 (1998): 1017-1098.
- Heinrich, Carolyn et al. "Workforce Investment Act Non-Experimental Net Impact Evaluation" Impaq International (2008).
- Heinrich, Carolyn et al. "New Estimates of Public Employment and Training Program Net Impacts: A Nonexperimental Evaluation of the Workforce Investment Act Program" IZA DP No. 4569 (2009).
- Hollenbeck, Kevin. Workforce Investment Act (WIA) net impact estimates and rates of return. Upjohn Institute Conference Papers, 1-23 (2009) .
- Hollenbeck, Kevin and Wei-Jang Huang. "Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State." Upjohn Institute (2006).
- Hollenbeck, Kevin; Daniel Schroeder; Christopher King; and Wei-Jang Huang "Net Impact Estimates for Services Provided through the Workforce Investment Act" Employment and Training Administration Occasional Paper 2006-06, U.S. Department of Labor (2005).
- Imbens, Guido and Jeffery Wooldridge. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47, no. 1 (2009): 5-86.
- King, Christopher. The effectiveness of publicly financed training in the United States: Implications for WIA and related programs, Upjohn Institute Press Book Chapters (2004): 57-99.
- King, Christopher and Carolyn Heinrich. "How Effective are Workforce Development Programs? Implications for U.S. Workforce Policies." Association for Public Policy Analysis and Management's Fall Research Conference (2011): 1-26.
- King, Christopher et al. "Returns from Investments in Workforce Services: Texas Statewide Estimates for Participants, Taxpayers, and Society." Ray Marshall Center for the Study of Human Resources.
- LaLonde, Robert. "Evaluating the Econometric Evaluations of Training Programs" *American Economic Review* 76, no. 4 (1986): 604-620.
- Liebman, Jeffrey et al. "Social Impact Bonds: Lessons Learned So Far." John F. Kennedy School of Government, Harvard University. Community Development Investment Review.
- Miles, Marty, and Stacy Woodruff-Bolte. "Apples to Apples: Making Data Work for Community-Based Workforce Development Programs." Corporation for a Skilled Workforce (2013).
- Mueser, Peter, Kenneth Troske, and Alexey Gorislavsky Using state administrative data to measure program performance. *The Review of Economics and Statistics*, 89, no. 4 (2007): 761-783.
- Nichols, Austin. "Causal Inference with Observational Data" *The Stata Journal* 7, no. 4 (2007): 507-541.
- Office of the Legislative Auditor, State of Minnesota. "Evaluation Report: Workforce Programs" (2010).

Plesca, Miana and Jeffrey Smith. "Evaluating multi-treatment Programs: Theory and Evidence from the U.S. Job Training Partnership Act Experiment" *Empirical Economics* 32 (2007): 491-528.

Smith, Jeffrey and Petra Todd "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125, no. 1-2 (2005): 305-53.

Specktor, Carole. "Using Return-on-Investment Measures for Workforce Development Programs in Minnesota." Hubert H. Humphrey Institute of Public Affairs (2004).

Villa, Juan M. "DIFF: Stata Module to Perform Differences in Differences Estimation." Statistical Software Components. Boston College Department of Economics (2011).

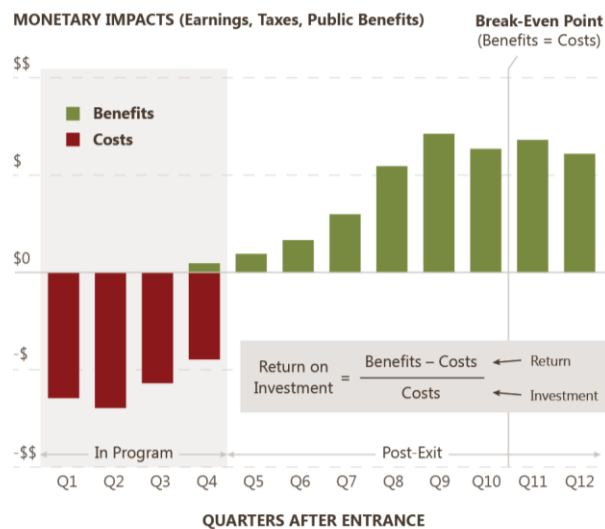
APPENDIX IV: A Framework for Estimating Costs, Benefits, and Social Return on Investment

Net Impact Evaluation: A Strong Foundation for estimating Social Return on Investment

Net impacts can be translated into dollars and cents in a standardized way, allowing us to understand the costs and benefits of programs and their outcomes. While this report does not undertake cost/benefit or social return on investment (SROI) analysis, this appendix details a standardized methodology that can be applied to future net impact evaluations, developed through consensus by the GWDC Net Impact Advisory Group. Using the SROI framework outlined below, the monetary effects stemming from workforce programs can be estimated over the same short-, medium-, and long-term follow-up periods used by the net impact evaluation framework. The final step—estimating the costs associated with programs, and thus an overall SROI—is perhaps the most challenging given the complexities and inconsistencies of current cost accounting practices, and should be treated with care.

FIGURE 16: RETURN ON INVESTMENT: WEIGHING COSTS AND BENEFITS

When a business or individual makes an investment decision, they consider the amount of money the investment will make relative to the initial cost of the investment. This ratio of the amount gained (or lost) to the initial amount invested is known as return on investment (ROI). For example, an ROI of seven percent (.07) means that every dollar invested returns seven cents of profit on top of returning the initial investment.



Note: The above chart is for illustrative purposes only and does not reflect actual data or outcomes. However, the overall pattern is typical: costs accrue early on while the program serves an individual, with benefits accruing later as individuals move back into the workforce.

Benefit and Cost “Perspectives”

The social return on investment (SROI) framework accounts for costs and benefits from both the participant and taxpayer points of view. This allows us to gauge the real value of programs to participants and the return on taxpayer investments.

In analyses of total social return on investment (SROI), costs and benefits are typically measured from the points of view of the participant and the “taxpayer” (or “government”). Together, the participant and taxpayer perspectives give an approximation of the total social return on investment. The approximation is rough because it leaves out other

difficult-to-measure impacts that accrue to broader society, beyond the more direct (and monetizable) impacts that accrue to the participant and the taxpayer.

Notably, some costs and benefits act as transfers between participants and taxpayers and therefore “net out” of the sum total social return on investment. For instance, a decrease in a participant’s MFIP payment is equally a cost to the participant and a benefit (or cost savings) to the taxpayer. In this case, the net effect is zero, since one’s benefit cancels out the other’s cost.

TABLE 9: BENEFIT- AND COST-ACCOUNTING FOR DIFFERENT GROUPS

The net impact framework accounts for costs and benefits to different groups. This table is for illustrative purposes only; actual effects may differ.

Benefit Categories	Training Participant	Taxpayers[†]	Society[†]
Change in Earnings and Fringe Benefits	Earnings up	No effect	Overall earnings up
Change in Taxes (Federal and State Income, Sales, and Payroll)	More taxes paid	Tax revenues up	No overall effect
Change in Public Benefits (e.g. MFIP, SNAP, MA, MinnesotaCare, UI)	Less benefits received	Greater savings	No overall effect
Change in Incarceration Costs	No effect	Greater savings	Greater savings
Change in Worker Productivity	Greater productivity	No effect	Greater productivity

Cost Categories	Training Participant	Taxpayers[†]	Society[†]
Program Expenditures	N/A	Cost to taxpayers	Cost to taxpayers
Foregone Participant Earnings While in Training	Cost to participant	N/A	Cost to participants
Foregone Tax Receipts While in Training	N/A	Cost to taxpayers	Cost to taxpayers
Tuition Costs Paid by Participant	Cost to participant	N/A	Cost to participants

No Employer “Perspective”. Benefits and costs to employers are not incorporated into the analysis for two reasons. First, data on worker productivity increases and their effects are not conclusive. Second, there is controversy about the incidence of taxes paid by the employer (e.g. FICA, UI taxes) – it is difficult to determine who actually bears those costs. Employer benefit-cost perspectives, where they are present in the literature, are often more “art” than “science.” That said, the employer “perspective” should continue to be examined and understood so as to one day incorporate it into the analysis.

Estimating the Monetary Effects of Workforce Program Participation

The SROI framework accounts for a number of different effects of workforce programs that are directly monetizable:

- Changes in earnings and employment
- Changes in fringe benefits earned
- Changes in income taxes (federal and state) paid
- Changes in payroll taxes paid
- Changes in sales taxes paid
- Changes in MFIP, DWP and SNAP benefits received
- Changes in MinnesotaCare & Medical Assistance enrollment
- Changes in Unemployment Insurance benefits received
- Changes in incarceration costs

TABLE 10: METHODS FOR ESTIMATING MONETARY EFFECTS OF WORKFORCE PROGRAMS

The table below provides detail on how to estimate the monetary effects associated with workforce program participation.

Effect	Methods / Assumptions	Sources for Data and Assumptions / Notes
Change in Earnings and Employment	Earnings are defined as the individual's total earnings over the baseline period and follow-up periods defined above.	Earnings and employment status are observed directly from wage data collected by the Minnesota Unemployment Insurance program.
Change in Fringe Benefits	Fringe benefits (healthcare, retirement, vacation) are estimated to equal 10 percent of gross earnings.	Assumption based on annual Bureau of Labor Statistics Employer Costs for Employee Compensation Report www.bls.gov/news.release/ecec.nr0.htm . Legally required benefits (e.g. Social Security, Medicare) are more likely to represent a transfer from employers to government in the short term.
Change in Income Taxes (Federal and State)	Income tax liability calculators like TaxSIM (for federal taxes) and those developed by the Minnesota Department of Revenue can be used to estimate changes in taxes paid at the individual level.	TaxSim: http://users.nber.org/~taxsim/
Change in Payroll Taxes	Change in taxes paid equals the individual's change in earnings multiplied by the statutory payroll tax rate (7.65%).	The employer portion of the payroll tax is not counted since it represents a transfer from one non-participant entity (the employer) to another (government).
Change in Sales Taxes	Taxes paid equals the individual's change in earnings multiplied by the average marginal sales tax rate for the given income.	Average marginal sales tax rates are provided by biennial Minnesota Department of Revenue Tax Incidence Studies. Annual Minnesota Tax Incidence Study
Change in MFIP, DWP and SNAP Benefits	Changes in benefit levels are observed directly from the Department of Human Services using data matching techniques.	DHS Administrative Data
Change in MinnesotaCare & Medical Assistance Benefits	Changes in public health care coverage eligibility and associated costs/benefits are observed directly from the Department of Human Services using data matching techniques.	DHS Administrative Data
Change in Unemployment Insurance Payments	Changes in benefit levels are observed directly from the Minnesota Unemployment Insurance Program using data matching techniques.	UI Benefits Data
Change in Incarceration Costs	Changes in costs are derived from the difference in recidivism rates between treatment and comparison groups. For individuals with prior history in a correctional facility, recidivism can be determined from Department of Corrections data using data matching techniques. Cost per inmate equals the average length of stay in a correctional facility multiplied by the marginal per diem cost per inmate.	Department of Corrections Administrative Data (for felony-level offenses) Bureau of Criminal Apprehension Administrative Data (for lesser offenses) Notes: Data on Drug Courts and Lesser Offenses collected by the Bureau of Criminal Apprehension (BCA) may be difficult to collect due to challenges linking individual records to BCA data, which do not use SSN).

Certain program effects are intentionally not included in the SROI framework (or at least not in this initial version). These effects are either speculative or hard to quantify without the use of major assumptions:

- Economic Multipliers

- Changes in Mental and Physical Health
- Changes in Worker Productivity
- Extrapolation of Costs/Benefits Beyond Observed Timeframes

Other effects may be technically quantifiable using administrative data, but are not included because of difficulty linking data and/or reliably tracking costs, minimal expected impact or relevance to workforce program participants, and a general preference to remain conservative when estimating benefits. The following monetary effects are therefore not included:

- Changes in childcare assistance benefits
- Changes in subsidized housing benefits
- Changes in Prescription Drug Program benefits
- Changes in child support payments
- Changes in General Assistance payments
- Change in Work Benefit Program benefits

Moreover, the costs to government associated with collecting taxes and administering public benefits programs, including the costs of determining eligibility, investigating fraud, making payments, et cetera, are not included. Such costs are difficult to quantify and, in the interest of keeping estimates conservative, are acknowledged but not monetized.

Costs Associated with Workforce Programs

Estimating the costs associated with workforce program participation seems straightforward at first but is quite difficult.

In light of these challenges, which are outlined below, the SROI framework takes a very straightforward approach, which comes with a major caveat. For the sake of simplicity and feasibility, estimates of program- and service-related costs should be based on existing state and federal accounting and reporting guidelines, procedures, and timeframes. The benefit of this approach is that it builds on existing practices and conventions. The caveat, of course, is that these practices and conventions vary significantly between programs (and sometimes within them). For this reason, SROI estimates should be interpreted with caution.

Challenges Regarding Cost Accounting

Measuring costs consistently, completely, and transparently is difficult given the differing methods used to account for costs, different accounting periods used, different units of measurement, and differing guidelines and practices regarding how costs are associated to specific activities and services. Consistently discerning the difference between marginal and fixed (or administrative) costs is also a challenge.

Ideally, we could determine for a given program participant a unique program cost estimate based on the program(s) the individual participated in, the dates of their enrollment, and the specific services they received (e.g. training, support services, other employment services). At this time, such an approach is not feasible given the challenges outlined above and because it would involve gathering data from stand-alone spreadsheets and, sometimes, paper financial reports.

Moreover, many publicly-funded workforce service providers braid together funding from other sources. Ideally, these “braided funds” should be measured and incorporated into benefit-cost analyses for two reasons:

- Most workforce programs are funded by a variety of sources, some of which may be private dollars or in-kind donations. If we do not track these costs consistently and accurately, total costs may be understated, and SROI overstated.
- Since a program's SROI can be overstated by the omission of braided funds, programs may have an incentive to underreport these sources of funding.

That said, it is not currently practical to measure braided costs dependably. First, information on braided costs is often costly and difficult to collect accurately, as it is buried in audits requiring time to retrieve, peruse, extract, and interpret the necessary data. More limiting is the fact that data on braided costs is incomplete across programs and over time. The incompleteness of the data and the difficulty of accessing and interpreting it make it infeasible to incorporate braided costs into analyses at this time.

Most workforce programs also refer customers to other services and resources that may contribute to positive outcomes. In order to achieve appropriate attribution of impacts, ideally we could understand the full cost structure of vended services and their providers, as well as any sub-vendors that may be operating. However, this level of sophistication is not currently feasible.

APPENDIX V: Technical Limitations and Caveats

It is important to understand the limitations associated with any performance evaluation approach, as well incentives created by such evaluations. The GWDC Net Impact Advisory Group has identified the following limitations, caveats, potential unintended incentives.

General Caveats

General Equilibrium Effects

The main concern in this area is that the training programs do not necessarily increase the supply of jobs and therefore any job filled by a participant of a labor market program is filling a job that may have been filled by someone else. In other words, people coming out of training programs may be displacing other workers.

This study does not directly address these concerns, but there are several reasons why these concerns may not be necessarily as serious as sometimes believed. The first is that this argument depends on one's perception of the labor market matching process. Employers often cite the inability to find skilled workers as one of the most significant concerns they have about labor markets. By providing a link between workers and jobs, labor market programs may increase the efficiency of the labor market by reducing the time it would have taken for employers to find a good match. In other words, it is not clear that a given job would have been filled as quickly. Second, the labor market program may provide the specific skills that employers prefer, again suggesting that the programs improve the matching of the labor market.

Validity of Comparison Groups

We take several approaches to address concerns about the validity of the comparison group. One point that is important to keep in mind is that the comparison group helps control for the effects of common economic conditions, which is especially relevant for the time period covered by this pilot. In particular, our analysis covers the time of the financial crisis and the subsequent high unemployment that followed. Since our comparison group comes from the same labor markets as our treatment group, we can control for these economy-wide effects on our outcome variables.

General Limitations of Return on Investment

ROI analysis is an essential tool for understanding certain dimensions of programs (namely cost effectiveness), but should not be considered the full measure of a program's worth. ROI does not take into account all the value programs create, and some programs' benefits are more difficult to monetize than others. ROI analysis does not address other considerations policy makers must make (e.g. moral considerations, fairness and equity, "fit" to the state's needs). Policy makers should use ROI as one tool in their toolbox for evaluating programs, but should also consider other important dimensions that ROI does not address.

Apples-to-Apples Comparisons

Workforce programs have varying missions and serve a variety of different customers. Programs are also subject to local factors outside their control, including economic factors, resource limitations, and so forth. We have endeavored to build a single standardized methodology and to incorporate performance benchmarks that take these specific aspects of programs into account. That said, it may be inappropriate to compare certain programs side-to-side when they serve vastly different customers, have different missions, or face different local conditions.

Data Inconsistencies and Gaps

The way data are reported by service providers can vary. For instance, providers track which services each customer receives via WorkforceOne, but how providers define these services varies. Additionally, what may be considered an administrative cost may vary from provider to provider. Another challenge is data integrity; incomplete data result in some individuals being dropped from the analysis. Improving data quality would increase sample sizes and allow the framework to be applied more broadly.

Return on Investment-Related Limitations and Likely Effects

LIMITATION

“Below the Line” Benefits

Many of the benefits (to participants and taxpayers alike) that may be attributable to a program are not quantified in the ROI model. These benefits include but are not limited to:

- Reductions in public expenditures associated with assistance programs not included in the model (General Assistance, SSI, Childcare Assistance, emergency services, subsidized housing, etc.)
- Increased worker productivity, which benefits firms and Minnesota’s overall economic outlook
- Local economic multipliers due to increased individual spending power
- Reduced social costs related to criminal activity and administrative costs to the criminal justice system
- Benefits to mental and physical health
- Benefits to the children and families of customers

Persistence of Benefits

The benefits of training and employment programs may persist for years, even decades. We have taken a conservative approach by estimating benefits in the near term (three years) using actual data; we do not attempt to project benefits beyond that point.

“Braided” Costs

Most workforce programs are funded by a variety of sources, some of which may be private dollars or in-kind donations. We are not able to track all these costs, so cost totals may be understated.

Unintended Incentives. Since a program’s ROI can be overstated by the omission of non-public funding, programs may have an incentive to underreport other sources of funding.

Referred and Vended Services

Most workforce programs refer customers to other services and resources that may contribute to positive outcomes. We are not able to track these referrals, so the benefits they may create may be improperly attributed.

Unintended Incentives. Since a program’s ROI can be boosted by increasing the use of referrals or vended services, programs may have an incentive to reduce the direct services they provide.

Determining Administrative Costs

Our ability to identify all administrative cost associated with a program is limited. This may understate total costs.

Marginal Administrative Costs

Administrative costs are more likely to be fixed, so it is likely that additional investments in

LIKELY EFFECT

Understates ROI

Understates ROI

Overstates ROI

Overstates ROI

Overstates ROI

Understates

programs would contribute more to service-related costs. A backward-looking ROI analysis that includes all costs (both fixed and marginal costs) would therefore understate the returns to additional investment.

(prospective)
ROI

Incentives to Keep in Mind

Participation timelines, i.e. when a participant is exited. Programs may exit participants at specific points to increase their observed net impact or return on investment, for instance by waiting to exit a participant until a good wage is established for a full quarter. To address, we base our follow-up analysis timeframes on the date of program entrance.

Selective Enrollment: Programs may select participants who are easier to serve with regard to the outcomes being measured. The use of well-matched control groups helps to offset this, as does the disaggregation of results and the use of contextualized performance goals.

APPENDIX VI: GWDC and OLA Recommendations on Net Impact Evaluation

Both the Governor's Workforce Development Council (GWDC) and the Minnesota Office of the Legislative Auditor (OLA) have made recommendations in recent years about the use of net impact evaluation in understanding and improving Minnesota's workforce programs.

GWDC Recommendations

These recommendations can be found in the GWDC's 2014 policy advisory, *Building Partnerships to Overcome Barriers*, pages 19-20, available at <http://www.gwdc.org/docs/publications/GWDC-Building-Partnerships-2014.pdf>.

Net Impact Recommendations

Goal: Minnesota uses a standardized net impact evaluation framework to guide workforce strategy and aid in the continuous improvement of workforce education and training programs toward better outcomes for participants.

Methodology

Recommendation 1: Standardized Net Impact Evaluation Design

Evaluation design should follow the work of the GWDC pilot project and the technical specifications agreed to by the ROI Initiative Advisory Group.³² These specifications should be reviewed, updated, and built upon as necessary via the framework's oversight function (see recommendation seven).

Recommendation 2: Contextualized Net Impact Goals

Net impacts estimated for specific programs and providers should be interpreted in their appropriate context, taking into account the population served, local economic conditions, and other factors outside the control of the program itself. Accordingly, statistical techniques should be used to develop net impact targets that are adjusted to account for program-specific and provider-specific factors. These targets should be used to identify useful benchmarks, encourage appropriate comparisons, and understand programs and providers in context.

Recommendation 3: Leading Indicators

Statistical techniques should be used to develop leading indicators that are statistically predictive of longer-term net impacts. These leading indicators could include near-term participant outcomes (such as employment or the attainment of a certain wage) or programmatic progress points (such as completion of a training module or a score on a particular assessment).

Data Sharing and Infrastructure

Recommendation 4: Statutory Support for Net Impact Data Sharing

The legislature should consider changes to state statute to permit the ongoing sharing of individual-level data between state agencies specifically for the purposes of the framework, barring other state or federal data privacy restrictions. These changes should be made in as narrow a fashion as possible, and with as many safeguards for individual data privacy as possible.

Recommendation 5: Better Consistency Across State Data Systems

State data systems should be improved through the creation of more consistent definitions and data collection practices across systems.³³ In particular, greater consistency and completeness would be useful with regard to:

- How entrance into and exit from programs is defined.
- How data on participant characteristics are defined.
- How various activities/services offered to participants are defined and tracked, including referred and vended services.
- How progress points and other potential leading indicators are defined and tracked.
- How costs are defined, categorized, and reported. In particular, greater clarity and alignment of cost categories to programmatic services and activities data (including administrative or fixed costs) is needed.
- Other costs associated with participant outcomes, including braided funds from other sources and costs associated with referred and vended services.

Recommendation 6: Development of a Data Warehouse

To fully utilize the net impact framework, the state should continue the development of longitudinal data systems including a data warehouse that integrates data across state agencies and programs to the greatest extent possible. The warehouse should also integrate financial data on program costs and make it easier for third-party service providers to provide data in a timely fashion.

Evaluation Integrity and Reporting

Recommendation 7: Oversight Function

The GWDC should continue to act as an oversight board to ensure a highly credible, transparent, and standardized net impact framework. The GWDC should oversee the net impact framework, make recommendations on the use of data and results, and solicit guidance from advisory committees made up of additional stakeholders and representatives from the research and evaluation community.

Recommendation 8: Evaluation Integrity

To mitigate conflicts of interest, net impact and cost-benefit analyses should be either: (1) conducted by entity(ies) or organization(s) that do not make policy recommendations nor fund, administer, or operate workforce development programs or (2) include external audits of results, data, and methodologies.

Recommendation 9: Reporting

For the net impact framework to be useful, its results and insights should be readily accessible, intuitive to explore, straightforward to understand, and packaged with key audiences in mind. Results should be presented with a high level of context and guidance for proper interpretation and use. An online, interactive dashboard tool that meets high standards for usability should be developed to ensure transparency and accessibility.

OLA Recommendation

This recommendation can be found in the OLA's 2010 *Evaluation Report on Workforce Programs*, pages 33-37, available at <http://www.auditor.leg.state.mn.us/ped/pedrep/workforce.pdf>.

RECOMMENDATION: DEED should periodically perform a long-term statistical analysis comparing workforce program participants and non-participants and report its findings to the Legislature and the public.

Because it compares program participants with non-participants, the method we used provides a more rigorous assessment of the effectiveness of workforce programs than the current federal performance measures (described later in this chapter) or the various return-on-investment models that have been

offered by workforce service areas and nonprofit service providers. The analysis would be even more valuable if it were part of a series of similar analyses conducted at regular intervals, so that changes in program policy could be assessed. Minnesota should follow the lead of the state of Washington, which retained outside consultants to conduct similar comparison analyses in 1997, 2002, and 2006.³⁸

We believe that a comparison analysis has three key advantages over current evaluation methods. First, the analysis above compares program participants with similarly situated non-participants. By doing so, it places the performance of workforce program participants in the context of individual characteristics and current economic conditions. Current federal performance measures and local return-on-investment results (discussed later in this chapter) are subject to the vagaries of the economic climate. In a recession, these measures necessarily decline because unemployment increases for many sectors of the economy. In contrast, a model that compares participants to non-participants may show that participants are better situated in a difficult job market if the workforce programs provide real advantages. Conversely, a comparison model could indicate that strong program performance in a boom economy merely reflects the current economic situation and that non-participants are equally adept at finding jobs.

Second, the analysis allows for some examination of the differing impact of workforce programs on different populations. Our analysis found differences in program effects for men and women, a finding that would not be uncovered through examination of the standard performance measures. We hope that this finding will spur discussion among workforce program providers as they look for ways to improve their programs. However, it is important to note that evaluating program effects for subgroups becomes increasingly difficult for smaller groups. It may not be possible, for example, to obtain useful results when comparing disabled participants with disabled non-participants because the total number of individuals would not be large enough for appropriate statistical comparisons.

Third, our analysis looks at the longer-term impacts of workforce program participation by examining wage records up to four years following an individual's entry into the program. Current federal performance measures take only six to nine months of employment and earnings into account. We believe it is especially important to examine the longer-term effects when workers receive training, because workers that acquire new skills should become more employable and earn higher wages.

However, the analytical approach we used has its own drawbacks. It is logistically and methodologically complex. Our recommendation that DEED occasionally perform a similar study anticipates that the department will most likely hire an outside contractor to perform the study, which will incur costs. If the department retains an outside contractor to carry out our recommendation, it should consider retaining the contractor long before the analysis is undertaken. The contractor can then learn more about current data collection practices and possibly make suggestions to reduce logistical obstacles. Early communication could reduce the amount of time the contractor needs to prepare data for analysis and possibly save the department money in the long run.

Another drawback is that this analysis cannot entirely compensate for the effects of selective enrollment, when workforce center personnel enroll only the most promising individuals in workforce programs. The

³⁸ Footnote from the cited text: See Kevin Hollenbeck and Wei-Jang Huang, *Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State*, Upjohn Institute Technical Report No. TR06-020 (Kalamazoo, MI: W.E. Upjohn Institute, 2006); *Workforce Training Results 2002: An Evaluation of Washington State's Workforce Development System* (Olympia, WA: Washington State Workforce Training and Education Training Board, 2003); *Workforce Training Results: An Evaluation of Washington State's Workforce Training System, 1997, Second Edition* (Olympia, WA: Washington State Workforce Training and Education Training Board, 1997).

analysis can limit the influence of selective enrollment by matching participants with non-participants based on a number of factors that affect the likelihood of future employment, including previous employment history and various demographic characteristics. However, it cannot take into account various intangible factors that may affect a client's ability to find work, such as motivation, communication skills, punctuality, or other characteristics that may be readily apparent to an employment counselor.

A third drawback stems from the need to wait several years after program entry to assess the long-term performance of program participants. Our analysis, for example, does not include any participants that entered workforce programs during the current economic recession. Thus, we were unable to estimate program effects for a time period about which legislators and the public are particularly interested. Further, workforce programs change over time, so this analysis is to some extent out-of-date by design. Estimates of program effects for clients who started in 2005 are of limited usefulness in 2010 and certainly cannot be used as the sole means of evaluating current program performance. Nonetheless, we believe comparison analysis provides an important perspective on program outcomes that cannot be obtained using other measures.