FLORIDA DEPARTMENT OF ECONOMIC OPPORTUNITY

PERFORMANCE FUNDING MODEL

WORKFORCE INNOVATION FUND

FINAL EVALUATION REPORT

SEPTEMBER 2019

Submitted by:
The Policy & Research Group
www.policyandresearch.com

8434 Oak St.
New Orleans, LA 70118

1411 4th Ave., Suite 1000
Seattle, WA 98101

This report was funded with federal funds under a grant awarded by the U.S. Department of Labor's Employment and Training Administration. The content of this publication does not necessarily reflect the views or the policies of the U.S. Department of Labor, nor does mention of trade names, commercial products, or organizations imply any endorsement of same by the U.S. government.

EXECUTIVE SUMMARY

The Florida Department of Economic Opportunity (DEO) and CareerSource Florida (CSF) received a Workforce Innovation Fund (WIF) grant in 2014 to implement and evaluate a statewide Performance Funding Model (PFM). As part of the WIF grant, the PFM team contracted with The Policy & Research Group (PRG) to evaluate the implementation and program cost, and to measure the effect of this statewide innovation on labor market outcomes.

The PFM is a resource-distribution strategy used to reward local workforce development boards (LWDBs) for their performance relative to seven performance metrics. In implementing the PFM, CSF's aim was to incentivize change and motivate local board leadership to increase efficiency and effectiveness. Ultimately, the PFM team hypothesized that this increased efficiency and effectiveness at the local board level would result in better outcomes for individual clients (increased employment, increased quarterly wages, and decreased time to employment). The study of the effectiveness of the program therefore aims to assess the difference in employment and wage outcomes exhibited by clients of the state workforce system before and after the statewide innovation. The Implementation Evaluation documents the rollout of the PFM; how that implementation deviated from the plan; what external contextual factors influenced the rollout; the challenges and successes of the project; and lessons learned during implementation. The cost study calculates PFM costs for three phases of the project: development (July 2014–June 2015), startup implementation (July 2015– June 2017), and ongoing implementation (July 2017–June 2018). It is the hope of the project staff and evaluation team that the findings from these three studies can be used in the development of future pay-for-performance models and the lessons learned can be leveraged in the design of future projects.

KEY RESEARCH FINDINGS

The introduction of the PFM was an ambitious undertaking, and one in which the PFM team was ultimately successful, despite the complexity of the project, challenges that arose, and external factors that impacted implementation.

Key deviations from the planned rollout of the PFM included delays in the initial round of awards, delays in making the web application data available, and turnover of key staff. Staff turnover included the departure of the original architect of the PFM in the first year of implementation and the departure of the original PFM project manager in the second implementation year.

Presented below are lessons learned through implementation that can inform future iterations of the PFM, as well as state-level pay-for-performance models outside of Florida:

- Successful rollout of the model requires more than just technical staff. Having staff aligned to the key functional requirements of the PFM was essential to the successful implementation. During the grant period, CSF and DEO expanded the PFM team to include the following core roles: project manager, senior software engineer, communications specialist/project coordinator, chief economist, and performance and analytics specialist, as well as a technical support position and an administrative support position provided by DEO.

- Multiple communication strategies are necessary to help boards understand how the metrics are calculated; in particular, one-on-one technical assistance proved to be particularly effective.

- It would be ideal to have the ability to make changes to the metrics based on economic changes and changes to state- or local-level goals; however, effective refinement of metrics can be difficult, as consistency is important for comprehension and buy-in at the local level.
- The lag in certified wage data is a hinderance to the PFM model. Though the lag affects all states, and therefore was outside the control of the PFM team, the PFM theory of change dictates that boards change their behavior based on performance feedback. In the case of wage data, feedback arrives too late to impact operational decision making.
- Having a web application up and running as soon as possible is critical. Providing boards with access to up-to-date data and the ability to drill down in those data was key for both board understanding of the PFM and the metrics, as well as for the motivation to respond to data and enact change.
- Data accuracy is essential in building trust with local boards. An adjustment to metric calculations in the first year impacted the perceived validity of PFM data. At CSF's request, an enhanced process was implemented to ensure a two-party review of the data. In this process, DEO pulled data relevant to the PFM from the *Employ Florida* case management system, verified the data per metric definitions, and securely transferred data tables to CSF for uploading to the PFM website.
- Boards need consistent help in understanding how to change their strategies and performance to influence metrics. This support was provided by CSF via written briefs, webinars, and one-on-one technical assistance throughout the grant period.
- Boards disliked the competitive aspect of the PFM model; in future iterations, PFM developers could consider how to reward collaborative efforts.
- The funding provided must be substantial enough to motivate participation for boards of all sizes to participate in and engage with the model. The PFM incentive awards varied significantly for each state fiscal year: $5.7 million was announced for 2015–2016; $11.5 million for 2016–2017; and $1.5 million for 2017–2018. Some boards, particularly those serving larger populations, indicated that incentive amounts were not substantial enough to influence priorities.
- Integrating the model with other requirements and metrics reduces conflict and demands less time of local boards. During implementation of the PFM, the Workforce Innovation and Opportunity Act (WIOA) was introduced, and boards had to learn and respond to new WIOA metrics at the same time as new PFM metrics. The PFM team found success in working with boards to illustrate how the two sets of metrics were related and could be complementary.
- Obtaining and carefully integrating stakeholder feedback from the beginning of the project facilitates the creation of a model that is relevant, affects decision making as intended, and is sustainable in design.

To assess the effect of the PFM, the evaluators employed a nonexperimental, natural experiment design with an off-year comparison group. We compared employment and wage outcomes for individuals who enrolled in LWDB services after the introduction of the PFM with those who enrolled in LWDB services before the PFM. We employed propensity score weights and statistical modeling to control for observed differences in group composition and for differences in contextual and labor market experience attributable to the off-year comparison. The purpose of the study is to produce nonexperimental estimates of the effect of a statewide innovation within the Florida workforce system on client labor market outcomes. The results are summarized below by research question.

Research Question 1: Effect on Probability of Employment

The preferred empirical model suggests that four quarters after enrollment, the PFM had a marginal positive effect on the employment outcomes of CSF clients included in the study. Estimates indicate that, for the typical or average client, "turning on" the PFM increases the probability of employment by 1.7% four quarters after enrolling in LWDB services.

Research Question 2: Effect on Quarterly Wages

The preferred empirical model finds that four quarters after enrollment, the PFM had a marginal negative effect on the wage outcomes of CSF clients included in the study. Estimates indicate that for the typical or average client, "turning on" the PFM decreases quarterly wages by $139 four quarters after enrolling in LWDB services.

Research Question 3: Effect on Time to Employment

Preferred model estimates indicate that CSF clients who were unemployed at the time of enrollment and exposed to the PFM became employed incrementally sooner than similar unemployed participants in the comparison group. The treatment effect is very small and negligible in consequence.

Discussion

According to the outcomes study, the PFM produced mixed results: a marginal positive effect on client employment outcomes, marginal negative effect on wage outcomes, and negligible but positive effects on employment for CSF clients who were unemployed at the time of enrollment. These modest and mixed results are not surprising given the complexity of the intervention being implemented, the systems-level nature of the intervention, turnover in key staff, and the challenges in implementation that resulted in the delay of some key programmatic components, such as the web portal that allowed LWDB leadership to see their progress on the metrics.

First, client employment and earnings are distal outcomes of the PFM. Several levers have to move in the hypothesized direction before the client would actually realize the hypothesized benefits – in an effort to meet performance targets, local board leadership needs to change policies and programming, management and frontline staff need to change their behavior based on these changes, and then LWDB clients need to experience these changes. A systems-level intervention, although having the potential to impact many more individuals and become a permanent change, is more complicated to implement and offers a less direct line to client impact than a direct-services program, such as an innovative job training program.

Second, as noted, the PFM team experienced significant turnover in staff over the project period, including the original architect of the PFM in the first year of the grant and the original project manager in the second year. Left with a very complex project that was only partially implemented, CSF and DEO assembled and restructured a team to forge forward with implementation. At the conclusion of the grant, LWDB executive directors praised the communication and technical assistance provided to them in the second half of the grant. Nearly all highlighted the critical change that occurred with the transition to the new team.

Third, there was a delay in rollout for the web portal, which is a key feedback mechanism for LWDB leadership to see how they are performing on the PFM metrics. At the start of the project, CSF laid out three objectives, one of which was the creation "of a comprehensive, easy-to-understand, web-based data portal to provide local workforce development boards with the data necessary to inform their decision-making processes." Without this functionality the PFM was not operating as intended.

Additionally, the evaluation period may have been too short to fully assess the full set of effects on labor market outcomes. Reporting lags for receiving certified wage data and the due date for the final report both effectively truncated the evaluators' ability to assess outcomes beyond four quarters. As described in detail in this report, the treatment group enrollment period ended as PFM was still in mid-implementation. A longer treatment period and a longer follow-up period may result in more clear-cut outcomes.

Despite all of this, the findings indicate that there is evidence of promise. After overcoming the initial implementation challenges, the PFM team and LWDB leadership have worked to make the relationship more productive and collaborative. Final interviews with the PFM team and local board leadership suggest a willingness to engage with the PFM (or a new iteration of the PFM) going forward. Whatever form the revised model takes, CSF and local board executive directors learned from implementation and assert the value of significant planning, supportive partnerships, and continuous improvement in any future efforts.

TABLE OF CONTENTS

INTRODUCTION

The Workforce Innovation Fund (WIF), which is administered by the U.S. Department of Labor Employment and Training Administration (USDOL ETA), is designed to contribute to the field of workforce development by supporting the implementation and evaluation of innovative programs. *Incentivizing Performance Outcomes in the Modern Workforce Environment: Designing and Implementing a Performance Funding Model in the CareerSource Florida Network* was funded as a Type A award in Round II (2014) to the Florida Department of Economic Opportunity (DEO) and CareerSource Florida (CSF) to implement a statewide systems-level change. Type A WIF projects support implementation and evaluation of new and untested ideas.[1] WIF grantees are required to contract with a third-party evaluator and in 2014, DEO contracted with The Policy & Research Group (PRG).[2]

THE INTERVENTION

The Performance Funding Model (PFM) was designed by the CSF Analytics Unit in 2014. CSF serves as the workforce system's policymaking board in the state of Florida. DEO, the state's workforce agency, serves as the WIF grantee and the administrative entity to CSF. Twenty-four locally controlled local workforce development boards (LWDBs) assist CSF in their goal to help Floridians enter, remain, and advance in the workforce. The PFM necessitated coordination and effective partnerships between DEO, CSF, and LWDBs throughout the state of Florida.[3]

THEORY OF CHANGE

The PFM provides financial rewards to LWDBs based on their performance in meeting a set of targets. The model is intended to incentivize local boards to improve their practices, and ultimately, the outcomes for the job seekers and businesses they serve. This is accomplished by distributing annual financial awards to local boards based on the extent to which they reach their targets relative to other boards. Within the model, boards are given additional performance credit for serving populations with barriers to employment and those more at risk of not connecting to the labor market, as identified by the Workforce Innovation and Opportunity Act (WIOA).

The seven performance metrics are divided into "short-horizon" (*Unemployed Placement Rate*; *Time to Earnings*; *Cost per Employed Exit*; *Business Engagement*) and "long-horizon" (*Earnings per Dollar Spent*; *Average Earnings*; *Customer Satisfaction*) metrics. Short-horizon metrics were assessed in all three award years; long-horizon metrics were assessed in Years Two and Three only.

The logic model defines a set of key mediators that are anticipated as a result of exposure to the PFM and are necessary for the PFM to have its desired effect of impacting workforce outcomes in the state:[4]

- Board leadership changes policies and programs in an effort to meet performance targets
- Management and frontline staff behavior changes in an effort to meet performance targets
- Board clients experience new policies and programs based on LWDB changes to meet performance targets

---

[1] Type B projects support implementation and evaluation of promising ideas. Type C projects support adaptation and/or scale-up of a proven idea.
[2] *Thomas P. Miller & Associates* (TPMA) was originally subcontracted by PRG to conduct the implementation study and cost study; in 2017, PRG assumed these responsibilities along with the outcomes study.
[3] CSF previously referred to LWDBs as regional workforce boards.
[4] The original logic model was included in the *Evaluation Design Report* (EDR) submitted to the *WIF National Evaluation Coordinator* (WIF NEC).

- Board clients experience new interactions with frontline staff based on workforce development changes to meet performance targets
- Employers experience new policies and programs based on workforce development board changes to meet performance targets
- Board clients receive more effective services

OBJECTIVES

According to CSF, a core principle of state-level workforce policy in Florida is one in which LWDBs are given the flexibility to serve local populations in a manner that best assists them in entering, remaining, and advancing in the workforce, while advancing core statewide objectives. The PFM provided CSF and DEO an opportunity to implement a model designed to incentivize improvements in statewide outcomes while providing diverse LWDBs with the flexibility necessary to serve their clients.

At the start of the project, three objectives for the PFM were defined by CSF:

1. The creation of a performance funding model that correctly incentivizes LWDBs to work toward common, identified goals
2. The expansion of current data collection systems and the integration of new data collection tools that capture the data necessary to measure progress toward the incentivized goals, and the integration of these tools into a web-based PFM status monitor
3. The creation of a comprehensive, easy-to-understand, web-based data portal to provide LWDBs with the data necessary to inform their decision-making processes; this allows them to benchmark and track their performance, which encourages collaboration to maximize the potential of shared resources and ensures clear and effective communication[5]

According to PFM developers, the key hypothesis to be tested over the life of the grant was that "financial rewards attached to clear performance metrics will result in system-wide performance improvement on those key metrics." The PFM team adopted a TIE (Target, Improve, Excel) rewards approach, in which local boards could achieve success by (a) meeting or exceeding the global performance target as established by the model; and/or (b) showing substantial improvements in their global performance score over an annual period relative to other participating boards; and/or (c) residing among the top performers relative to other participating boards. To be evaluated under the TIE approach, boards had to first meet both the negotiated goals for the WIOA *Primary Indicators of Performance* and the PFM minimum threshold requirements.

PROGRAMMATIC TARGET POPULATION

The PFM's target population is 24 LWDBs in the state of Florida, and ultimately, the recipients of services offered by each board. Though all local boards were offered the opportunity to participate in the PFM, 20 participated in the program during all three award years.[6, 7] According to CSF, the statewide network assisted "more than 210,000 job seekers secure jobs and more than 65,000 employers with recruiting, hiring and training needs" in the 2017–2018 fiscal year.[8]

---

[5] The web-based status monitor referenced in the second objective was later integrated into the web-based data portal.

[6] *CareerSource Central Florida* participated in Years One and Three; *CareerSource Broward* participated in Year One only. Local leadership cited several reasons for opting out of Years Two and Three, including onerous survey requirements for businesses, the small incentive amount relative to the board's annual budget, and the inability to see the operational advantages of the program.

[7] *CareerSource Pinellas* and *CareerSource Tampa Bay* withdrew themselves from eligibility for 2016–2017 and 2017–2018 incentive award consideration due to USDOL and Florida DEO investigations into board operations.

[8] Retrieved January 24, 2019, from https://careersourceflorida.com/wp-content/uploads/2018/11/2017-18_CSF_Annual_Report.pdf

Boards are provided additional credit in their performance scores for serving certain populations with barriers to employment. Throughout the grant period, this included participants who were enrolled in the *Welfare Transition Program*, as well as veterans, disabled individuals, formerly incarcerated individuals, and/or individuals experiencing homelessness. In the third year, boards were also provided credit for serving participants with mandatory work requirements from the *Reemployment Assistance* program in Florida and/or the *Supplemental Nutrition Assistance Program* (SNAP).

PERFORMANCE METRICS

Each performance metric measures a different dimension for which each local board is evaluated in order to determine if they receive PFM incentive funding. A global performance target and targets by metric for each board were established at the start of each fiscal year. Using the web application, raw data were pulled and validated by DEO, submitted to CSF for global performance score and metric calculations, then made available to boards via the web tool. Annual award allocations were then made based on these calculations.[9] PFM performance metrics are divided into three categories: *placement*, *exit*, and *business*.

PLACEMENT

The PFM placement metrics focus on job seekers who have gone long periods of time without a wage and are particularly in need of finding employment. The number of job seekers considered by the placement metrics is intentionally small relative to the total number of job seekers. The metrics were originally intended to complement the Workforce Investment Act's (WIA) *Entered Employment Rate* by focusing on the very hard to serve. An important difference between the placement metrics and the current WIOA *Primary Indicators of Performance* is that the PFM placement metrics consider currently open cases, whereas the *Primary Indicators of Performance* only measure outcomes after a job seeker leaves the CSF network.[10]

EXIT

The PFM exit metrics examine outcomes experienced by job seekers after they have received services and left the CSF network. For this reason, the exit metrics are similar to the WIOA *Primary Indicators of Performance*. However, the *Average Earnings metric* of the PFM considers how much all participants earn after they have exited, whereas the WIOA *Median Earnings metric* considers only job seekers who were employed when they exited.

BUSINESS

The business metrics assess the interactions between LWDBs and the businesses they serve and the satisfaction of those businesses. Measurement of performance on the *Customer Satisfaction metric* relies on the number of surveys sent to businesses served, the response rate to those surveys, and answers to survey questions.

---

[9] Prior to the launch of the web application in 2017, local boards were provided quarterly performance data via an Excel spreadsheet.
[10] WIOA was signed into law on July 22, 2014, and replaced WIA.

IMPLEMENTATION STUDY

RESEARCH QUESTIONS

We aim to answer the following research questions through the Implementation Evaluation of the PFM:

1. *How did the PFM implementation roll out?*
2. *How did the actual rollout differ from the planned rollout?*
3. *What external contextual factors occurred during the PFM implementation?*
4. *What are the lessons learned from the PFM process?*
5. *What were the challenges and the successes?*

Research questions were discussed by PRG, CSF, and DEO during an on-site visit to CSF offices on November 27, 2017. Questions were then submitted to the *WIF National Evaluation Coordinator* (WIF NEC) on December 20, 2017.[11]

DATA COLLECTION AND ANALYSIS

To qualitatively assess each research question, implementation of the PFM, and experiences of CSF, DEO, and participating local boards, stakeholder interviews and focus groups were conducted by PRG and *Thomas P. Miller & Associates* (TPMA) throughout the grant period; each was audio-recorded and transcribed.[12] To analyze qualitative data, PRG research analysts read through transcripts to identify the range of responses to interview questions at each implementation stage.

In the spring of 2018, soon after Implementation Evaluation responsibilities shifted from TPMA to PRG, a PRG senior research analyst reviewed all data collected by TPMA to identify the full range of external influences to the PFM already identified by stakeholders. This comprehensive list was then organized into thematic categories. A PRG senior research analyst interviewed CSF and DEO staff about each category during calls held in March and April 2018. During this time, CSF and DEO provided PRG with detailed context regarding any potential factors of influence.

In the fall of 2018, a PRG research analyst conducted final one-on-one interviews with 17 LWDBs and 9 employees from CSF and DEO who contributed to the project. Each interview was customized to the knowledge and expertise of the interviewee, but all were asked to summarize their involvement with the PFM and to reflect on perceived strengths and weaknesses of the program.

It should be noted that this report reflects the perceptions of CSF, DEO, and local board staff who elected to participate in focus groups and interviews; it is not a wholly objective account and should be considered limited in this respect. Table 1 provides a summary of primary qualitative data collection activities.

---

[11] Revisions were submitted by PRG to the WIF NEC due to the transition in Implementation Evaluation responsibilities from TPMA to PRG in 2017. TPMA's original 12 Implementation Evaluation research questions can be found in the *Phase I Final Report and Revised Plan for Phase II*, which was submitted by PRG to DEO in December 2015.

[12] Transcripts were not available for the spring 2016 and December 2016 phone calls. In these instances, PRG relied on notes taken by TPMA at the time of each call.

Table 1. Primary Qualitative Data Sources

Date	Participants	Interview Type	Conducted by
Spring 2016	**Executive Directors**[13] 21 local workforce development boards	One-on-one (phone call)	TPMA
September 2016	**PFM Program Director & Director of Research and Analytics** CareerSource Florida	One-on-one (in person)	TPMA
	Senior Management Analyst Supervisor Department of Economic Opportunity	One-on-one (in person)	TPMA
	Executive Directors 18 local workforce development boards	Focus Group (in person)	TPMA
December 2016	**Executive Directors** 20 local workforce development boards	One-on-one (phone call)	TPMA
October 2017	**Executive Directors** 2 local workforce development boards	Focus Group (in person)	TPMA
	Executive Directors 8 local workforce development boards	Focus Group (in person)	TPMA
	CareerSource Florida PFM team 4 participants	Focus Group (in person)	TPMA
Spring 2018	**CareerSource Florida & DEO PFM team** 7 participants	Focus group (phone call)	PRG
Fall 2018	**CareerSource Florida & DEO PFM team** 9 participants	One-on-one (phone call)	PRG
December 2018	**Executive Directors** 17 local workforce development boards	One-on-one (phone call)	PRG

To confirm and clarify qualitative data gathered through interviews and focus groups, PRG conducted regular evaluation calls with CSF and DEO throughout the grant period. CSF and DEO staff also provided ongoing implementation process updates to PRG during these calls.

KEY IMPLEMENTATION EVENTS

In this section, we examine the first two research questions: *How did the PFM implementation roll out?* and *How did the actual rollout differ from the planned rollout?* We have defined three distinct periods to be used for both the implementation and cost study: Phase I: Development, Phase II: Startup Implementation, and Phase III: Ongoing Implementation. A timeline of key events within each phase is described below.

PHASE I: DEVELOPMENT

JULY 2014–JUNE 2015

FORMATIVE RESEARCH & MODEL DEVELOPMENT

In the first year of the grant, the CSF Analytics Unit worked to develop and refine the PFM. The PFM developer traveled to all participating local boards to gather information to create profiles, refine

[13] In some phone calls, support staff participated either in lieu of or along with the executive directors.

targets, and collect feedback from representatives. By the end of the year, the funding model was finalized. All metrics and targets were released to boards by June 23, 2015.

STAFFING

During this year, other members of the CSF Analytics Unit assisted the PFM developer in minor support roles. In April 2015, a PFM project manager was hired by CSF. CSF executive leadership and the board of directors provided project oversight throughout the grant period.

PHASE II: STARTUP IMPLEMENTATION

JULY 2015–JUNE 2017

TECHNICAL ASSISTANCE

On July 28, 2015, the CSF Analytics Unit submitted a white paper to all local boards in which the model was explained.[14] Also submitted were plans for future modifications and technical assistance, such as monthly webinars conducted by CSF to explain various components of the model.

Beginning July 2016, the three long-horizon metrics (*Earnings per Dollar Spent*, *Average Earnings*, *Customer Satisfaction*) were integrated into the model, as was planned during the PFM's original design period. The definition of participants for the placement metrics was expanded to include individuals regardless of employment status upon entering the workforce system. The *Time to Earnings* metric was revised to measure the length of time participants had been without a job (the number of quarters since the participant either entered the system or last earned a wage). The minimum threshold methodology was more clearly defined so that boards were evaluated under the TIE approach if they achieved minimum thresholds of 50 to 100 points per metric totaling 525 or more when summed together.

WEB APPLICATION

The web application was originally slated to launch by the end of 2015. Due to a change in scope and overhaul of the site, rollout was delayed until May 2017.

ANNUAL FUNDING & TARGETS

Year One targets were revised on July 28, and again on August 21, 2015.[15] In July, boards were informed that $6.15 million was available for the first award year. Preliminary performance rankings for Year One were announced on February 22, 2017. LWDBs were offered a two-week period to review results. During that time, calculation adjustments were made to the *Cost Per Employed Exit* metric. Year One incentives were then distributed to local boards in March and April 2017.

The $11.5 million allocated for Year Two awards, as well as targets for the year, were announced to boards by June 2016; award funds were distributed in June 2018. The $1.5 million allocated for Year Three awards, as well as targets for the year, were announced to boards by June 2017; according to the PFM team, award funds were distributed in July 2019.[16]

[14] A detailed explanation of the model, reflective of updates made to metrics and methodology during the grant period, can be found here: https://pfm.careersourceflorida.com/

[15] PFM award years coincide with the Florida fiscal year, which begins on July 1 of each calendar year. Year One ran from July 2015 to June 2016; Year Two ran from July 2016 to June 2017; Year Three ran from July 2017 to June 2018.

[16] The time period to award funds was based on the receipt of certified wage data, which is delayed due to the time required for the *U.S. Department of Revenue* to certify the data.

STAFFING

The PFM developer and original project manager both left the project during Phase II. With their departures, CSF assigned a group of four people to work on the project in the following specialized roles: project manager, communications specialist/project coordinator, chief economist, and senior software engineer. Additional expertise in the areas of external affairs, technology, data, and performance analytics was also provided on an as-needed basis.

DEO contributions also changed during Phase II. DEO's senior management analyst supervisor, who was involved in early stages of the PFM, left the agency. To enhance data validation processes, DEO's supervisor of data collection and analytics assumed a heightened role in the PFM. DEO's workforce administrator retained administrative and oversight responsibilities throughout Phase II.

Apart from the PFM, leadership at CSF underwent a significant restructuring during this time. A timeline of key staffing changes within the PFM team, as well as changes in CSF leadership that occurred during PFM implementation, are provided in Figure 1 below.

PHASE III: ONGOING IMPLEMENTATION

JULY 2017–JUNE 2018

TECHNICAL ASSISTANCE

Once the new PFM team was in place, consistent technical support was provided to LWDBs. The new team organized regular webinars, workshops at annual professional development summits, and a technical review committee with local board representatives. CSF's chief economist and communications specialist/project coordinator provided boards with frequent one-on-one technical assistance during this time.

Due to feedback from local boards, CSF modified the *Customer Satisfaction* metric calculations to better accommodate the capacity of boards that serve small-, medium-, and large-sized geographic areas. In Year Three, CSF also designated additional populations for which local boards could receive credit in metric calculations. This change was intended to further incentivize boards to pursue reduced welfare dependency and to increase alignment with WIOA.

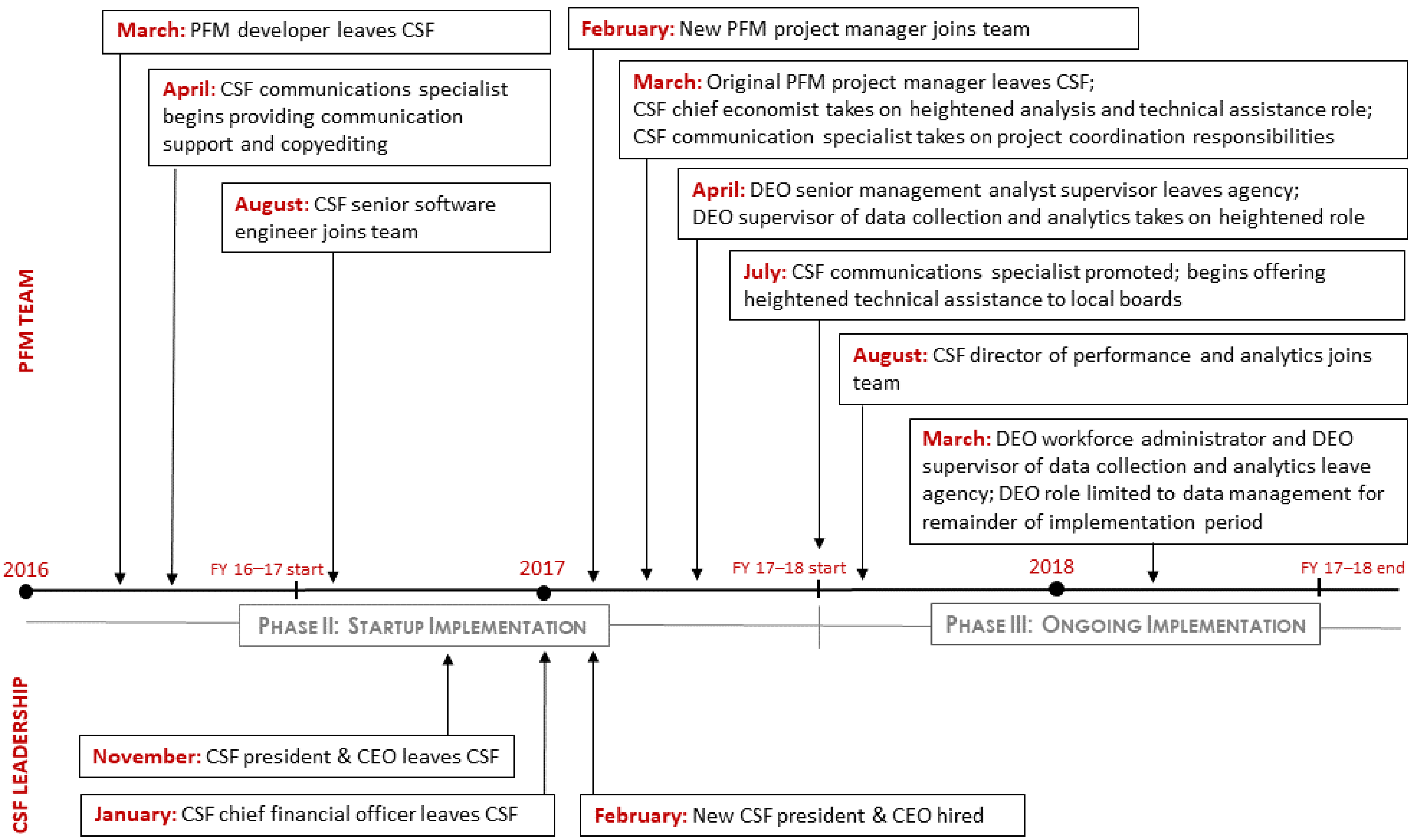
WEB APPLICATION

The PFM project manager, communications specialist/project coordinator, and chief economist met regularly to discuss communication strategies and how best to translate those strategies onto the new site. The team plans to continue refining the site and data available therein after the project period.

STAFFING

DEO's two original PFM team members remained on the project through March 2018. Remaining CSF and DEO team members continued on the project through the end of the 2018–2019 state fiscal year.

Figure 1. Key Staffing Changes During PFM Implementation Period



CHALLENGES, SUCCESSES, AND LESSONS LEARNED

OVERVIEW

In this section, we outline the findings for two research questions: *What are the lessons learned from the PFM process?* and *What were the challenges and the successes?*

In interviews, local board executive directors described several reasons for participating in the PFM. Most perceived the PFM to be a low-risk venture ("a no-brainer" or "win-win") as they would be eligible for additional annual funds and not be penalized for failure to meet PFM metrics.

Several issues were raised by stakeholders regarding the relevance and utility of PFM metrics and data. Although local boards appreciated the opportunity to offer feedback in the early stages of the project and its intended purpose of tailoring metrics to their local areas, most felt that the model fell short in capturing success at the local level. In initial interviews, directors expressed a desire for staff to focus on quality of service, not hitting metrics – to "do a good job and the numbers will follow." Focus groups and interviews suggest that the PFM made little impact on board-level operational decisions until the release of the web application. Comprehension of the PFM metrics and value of the PFM data appear to have increased dramatically for boards once teams were able to access the web tool and interact with their data.

Some design challenges persisted throughout the grant period, though PFM team members and local board staff offered a wealth of specific, actionable considerations for future program iterations to mitigate these issues. Below, we provide a more detailed analysis of successes, challenges, and lessons learned as they relate to key domains of the PFM's design and intended behavioral impact. We have categorized these findings under either Staffing & Leadership, PFM Metrics, PFM Data, or Board Engagement. A discussion of external factors that impacted implementation is included in the section that follows.

STAFFING & LEADERSHIP

THE CHALLENGE

Turnover of key staff is a potential challenge for any multiyear project, and the PFM team experienced a fairly high level of turnover. Throughout the project period, there were many changes in critical project-related staff and state-level leadership at both CSF and DEO, and these changes impacted PFM implementation. In February 2017, project leadership implemented a significant restructuring of the organization of the project; embedded in the restructuring was a shift in data management responsibilities from CSF to DEO.

STRATEGIES & ADJUSTMENTS

The PFM project manager reports that the 2017 restructuring to a core seven-person team with specialized, clearly defined roles enabled staff members with areas of expertise to oversee specific operational, communication, and analytical project components. At the conclusion of the grant, LWDB executive directors appear to have positive perceptions about support provided to them by the seven-person PFM team and nearly all highlighted the critical change that occurred with the transition. One said that "CareerSource Florida staff over the past, I want to say year or two, was just absolutely exceptional – with the process, with the feedback, and also I think very open to listening to the suggestions and the concerns, far more so than was there originally." Another executive director noted that they "always had the ability to reach out and get assistance if we needed it, whether it was through a phone call, an email, whatever – we've always had someone there ready to help."

PFM METRICS

COMPLEXITY

THE CHALLENGE

The complexity of PFM metrics frustrated stakeholders at all levels. Nearly all executive directors who participated in interviews pointed to issues of complexity and confusion in the first two years of implementation, such as "it was overly complicated," "definitions change abruptly and for, seemingly, no reason," or "anything that takes over 100 pages to explain is too difficult to deal with." In turn, executive directors expressed frustration that they were not equipped to educate frontline staff about the PFM. In a 2016 interview, the original PFM project manager said that "finding effective ways to communicate the details of what we're doing to local boards" remained a challenge and that, even with a trained economist on the CSF staff who could effectively communicate how metrics were evolving and how boards could use the web application to strategize, "there [wasn't] enough time because we're a small team and because there are 24 boards and because they're on such different levels of understanding."

STRATEGIES AND ADJUSTMENTS

During final interviews, one PFM team member agreed that initial communication "was a little too technical" but said improvements were made and "the more you talk about something, the better you get at talking about it. The more you explain something to someone, you always find a better way of doing it." Final interviews with executive directors also highlighted a marked change in comprehension with the introduction of a new PFM team, communication strategy, and heightened technical assistance in 2017. Nearly everyone that accessed one-on-one technical assistance provided in the second and third year of implementation recognized it as critical in understanding metrics and how to engage with the program. One executive director said that "the staff worked really well at getting down to the nitty gritty of the terminology that was used and the equation" and putting the information into "language we're used to . . . formulas that we're used to seeing."

VALIDITY

THE CHALLENGE

In initial interviews, many respondents expressed concern regarding the validity of the PFM metrics and a lack of enthusiasm for metric-oriented program objectives. One director described the PFM as misaligned with board values, a "reverse incentive" that "pushes us towards the easiest [people] to place . . . instead of the people who need help the most." Others said that the PFM is "trying to drive a quality-driven system into a quantifiable measure" and, in reference to the original business engagement metrics, that "you can't really capture the value of a relationship" in a survey. Many perceived the program to be rewarding processes, such as survey completion, rather than outcomes, such as long-term relationships with employers, particularly in the first year of implementation.

In final interviews, many executive directors expressed concern over the *Time to Earnings* calculation, stressing that training is often warranted. One said that "Trying to rush a placement can . . . hurt the region. It can also hurt the individual to say that person might really have needed training or additional services, where with a performance funding model, you're really trying to rush the placement, trying to get them employed as fast as possible." Another emphasized, "For us, the goal is to help you get a better job, whatever better means to you. And that goes counter to the measuring we use."

STRATEGIES & ADJUSTMENTS

In the second year of PFM implementation, CSF introduced the long-horizon metrics, which included the *Customer Satisfaction* metric. The performance score for this metric is determined by the survey response rate of businesses served by the LWDB, the expected response rate, and the average response to the first two questions on the survey. At the start of the fiscal year, the expected response rate was set at 70% for all LWDBs. Based on input from local boards, the expected response rate was reevaluated and modified to include a tiered structure of an expected response rate of 25% for large boards, 45% for medium boards, and 70% for small boards to provide an equitable distribution of survey responses needed. According to PFM team members, without this adjustment "it would have really hurt chances for some of the larger boards to participate in the PFM."

Although the PFM team was not able to adjust the model fully during the grant period, they continually discussed concerns, and are eager to improve upon the limitations of the model in future program iterations. For example, one PFM team member offered, "If there had been a way to weight [the metrics] a little differently depending on what the goals of the state are for that particular year because of economic changes, then I think that might have made a stronger model." Another noted that when the PFM was developed, "education and training services were not as high on people's list or high in terms of workforce development as obtaining a job" and that although "metrics were well aligned during that time in 2013–2014 when this concept came about – given the economic conditions, given the political conditions," the current low unemployment rate in Florida warrants heightened emphasis on training within the model and could potentially offer a more meaningful indicator of performance than job placement.

PFM DATA

WAGE DATA

THE CHALLENGE

In early interviews, executive directors expressed frustration with data accessibility and timeliness. The lag in receiving complete certified wage data was identified as problematic by several respondents – "half a year will be over before we figure out how we did in the previous year." One said that, by the time the data are available "we're in a new economy." Data-related delays have also made it difficult for boards to understand how to change operations to positively impact metrics. One respondent said, "when you don't know where the data is coming from, and you have to wait a year to look at it, it's too late to affect change to the positive."

STRATEGIES & ADJUSTMENTS

In final interviews, the lag in certified wage data was mentioned by several respondents as a persistent problem, though all recognized that the lag was outside the control of the PFM team.[17] PFM team members suggested future project iterations might manage expectations related to wage data, explore the use of "a different data set" altogether, or develop new predictive approaches that might get closer to "quasi-real-time reporting of wages."

WEB APPLICATION

THE CHALLENGE

Due to a change in scope and rebuild of most elements of the web application in the second implementation year, boards only began to engage with the web tool toward the end of 2017.[18] In interviews, some boards indicated that the delay caused board leadership and frontline staff to lose

[17] *Unemployment Insurance* (UI) wage data are used as the source for calculating PFM metrics. State verification processes for UI wage data can take several months. As such, delivery timelines for these data are outside the control of individual states within the United States.

[18] Development delays caused the PFM web application to be released in May 2017. In October 2017 interviews, after training and technical assistance were provided, boards indicated they were just beginning to engage with the platform.

interest in engaging with or prioritizing the PFM. One executive director said, "I think at the very beginning, I wish the website and the actual data was there like it was toward the end . . . the only thing you could really control or see or gauge to see where you were at was the business engagement piece."

STRATEGIES & ADJUSTMENTS

A PFM team member said that getting the web application up and running sooner was their top recommendation for what should have been done differently during implementation. Final interviews suggest that once the web application was released, it was user-friendly and adequate training and technical assistance were provided by the PFM team to end users. One executive director echoed this assessment and noted, "Once the portal was available and we could get out there and we could drill down . . . then that helped us tremendously to be able to really put the pieces together and understand what PFM really was." Others said that they "used the portal as a teaching tool to our managers and staff to show them the metrics and expectations and where they were coming from," and noted that the web application "caused us to look at data that we hadn't looked at before."

INACCURACY

THE CHALLENGE

Adjustments made to the first round of PFM award calculations (2015–2016) led to mistrust among some boards regarding data integrity. As a result, one respondent reported in initial interviews that their board maintained records and calculations, independent of Salesforce and the PFM web application, to ensure data accuracy.

STRATEGIES & ADJUSTMENTS

At the request of CSF, DEO assumed data management responsibilities in February 2017. The PFM project manager reports that this shift had a positive impact on implementation, as it provided greater accountability. Data moved from *Employ Florida*, to DEO, to CSF to ensure adequate validation prior to all PFM web application uploads. No recalculations in performance were required in the following year. In final interviews, no boards expressed concern over data inaccuracy, suggesting that strategic communications and access to the web tool successfully alleviated these concerns at the local level.

BOARD ENGAGEMENT

OPERATIONS

THE CHALLENGE

Interviews and focus groups suggest the PFM's impact on board operations to date is relatively limited. One said that the PFM is "more of a trend [tool] . . . than a tool to help us adjust and modify in real time. That was a concern that our board brought up. And thus, now, it's not a duplication of effort but we have to use other tools in order to try to capture things that are real time." Moreover, while comprehension and perceived utility of the model improved over time, many executive directors did not understand the metrics early on, and were therefore not capable of explaining the model to frontline staff or making operational changes based on PFM data during the grant period.

STRATEGIES & ADJUSTMENTS

Ultimately, some local boards perceived a greater level of engagement with businesses and special populations as a result of the PFM design and the targeted technical assistance that was provided to them. For instance, boards noted that the PFM "drives our contact with the employers," helps to build "rapport with the business community," and expands "communication with our employers." One executive director noted that a PFM team member provided calls to inform them when they were "getting close to [a] particular measure" and offered assistance and specific suggestions to the board. Another said, "I'd always just give them a call, even if I just had a simple question." Using PFM data, boards have also increased class offerings and modified practices related to skill set-matching in order to streamline and accelerate the reemployment process for clients.

COMPETITION

THE CHALLENGE

Though the PFM was not developed to garner competition among boards, stakeholders at all levels recognized that competition is inherent to the model. Many respondents said that competing with other boards for a finite pool of resources discourages collaboration and sharing of best practices. The boards have longstanding relationships with one another and continue to work together; however, some would prefer to be incentivized to build upon existing collaborative efforts that benefit the state workforce system.

STRATEGIES & ADJUSTMENTS

In the final interview, some PFM team members and local board representatives suggested that stratification of small, medium, and large boards could be of value beyond the business engagement calculations – with respect to equity, garnering buy-in from larger boards, and minimizing the perceived non-collaborative elements of the model. One executive director said that the model should be designed "so that incentives are shared more evenly between regions [local boards] based on reaching individual metrics."

INCENTIVE FUNDING

THE CHALLENGE

The CSF Board of Directors committed the following amounts to PFM awards for the 24 LWDBs for performance during each state fiscal year: $5.7 million in 2015–2016; $11.5 million in 2016–2017; and $1.5 million in 2017–2018.[19] According to the PFM project manager, the funding commitment in the first two years represented the board belief that "the PFM would be and continue to be a strong performance tool for the future." In the third year, a budget shortfall at the state level constrained the PFM and caused the CSF board to prioritize "sustainability" and, thus, substantially decrease the amount of funding allocated to the project. Potential award amounts that were insignificant relative to annual operating budgets, as well as the decreased funding allocations for 2017–2018, caused some local boards to lose enthusiasm for the project.

STRATEGIES & ADJUSTMENTS

The PFM project manager indicates that, theoretically, the Year Three (2017–2018) PFM award allocation represents a baseline level of annual award funding that local boards could reliably expect from CSF year after year. Final interviews suggest that though interest was heightened in years with higher award allocations, nearly all local boards were prepared to participate in the program in all three implementation years, as it could potentially expand their annual budget with no risk to the board.

EXTERNAL FACTORS

The purpose of this section is to elucidate factors that influenced implementation of the PFM, apart from its intended design and rollout. The data presented here offer a better understanding and documentation of the environment in which the PFM was implemented, from start to finish. In this section, we address

[19] Ultimately, $6.15 million was distributed to 23 local boards in Year One, $10.29 million was distributed to 20 local boards in Year Two, and $2.21 million was distributed to 21 local boards in Year Three.

the final research question: *What external contextual factors occurred during the PFM implementation?* The two external factors detailed below are the introduction of WIOA and natural disasters. Although each of these factors impacted operations in the state, the PFM team reports that none were significant enough to curtail the PFM model or halt PFM implementation efforts.

WORKFORCE INNOVATION AND OPPORTUNITY ACT

THE CHALLENGE

WIOA took effect on July 1, 2015. This new legislation required local boards to adjust to significant and mandatory changes in federal reporting by July 1, 2016. In initial interviews, concerns were raised about the complexity and burden of directing staff to focus on PFM metrics and WIOA *Primary Indicators of Performance* simultaneously. A PFM team member recognized this conflict and noted that learning the "different definitions and different purposes" of two unique sets of metrics placed an unintended burden on boards. The team reports that rollout of WIOA is the only external policy or initiative that had a significant impact on the PFM.[20]

STRATEGIES & ADJUSTMENTS

When local boards were balancing WIOA and PFM metrics simultaneously, project staff initiated one-on-one technical assistance to facilitate successful implementation of the PFM. Though greater alignment with WIOA is certainly a priority, it does not appear that the PFM team or local board staff desire a model that restricts itself to full replication of either state- or federally mandated measures. For example, one PFM team member said that, upon viewing the web application data, "certain local boards were very shocked . . . that individuals were in their system so long" and noted that this information doesn't "get picked up in federal measures, but was something that the performance funding model was able to see and show them from their operations perspective."

NATURAL DISASTERS

THE CHALLENGE

Florida has experienced three separate hurricanes during the PFM implementation period: Hermine (2016), Matthew (2016), and Irma (2017); Maria (2017) also impacted the state due to displaced persons moving to Florida from Puerto Rico. Recovery from these events required a significant time commitment from CSF and DEO staff, from daily phone calls to temporary relocations during Irma and Maria to bring expertise to impacted areas. The PFM team reports that infrastructure damages and road closures in the aftermath of these storms prevented staff from accessing work sites. In final interviews, one local board pointed to the lack of dynamism with the PFM, specifically with regard to adjusting targets in the aftermath of hurricane damage.

STRATEGIES & ADJUSTMENTS

After Hurricane Maria, changes in the labor market due to the influx of displaced persons caused DEO to develop new processes to quickly match incoming persons to jobs. CSF offered greater flexibility and heightened assistance to affected local boards in PFM data submissions during these times. Despite the significant impact these events had on the state of Florida, the team reports minimal disruption to PFM implementation activities.

---

[20] In interviews, CSF also cited the following initiatives that occurred during the grant period but were insignificant in their impact on PFM implementation: *Governor's Reemployment Challenge*; *Florida Job Placement Report*; *Florida Talent Prosperity Dashboard*; *Florida Chamber Foundation's Prosperity Initiative*; *The Florida Scorecard*; *Community Service Block Grant Program*; and *Employment First Florida*.

OUTCOMES STUDY

The purpose of the outcomes study is to determine the effect of a statewide innovation within the Florida workforce system on client labor market outcomes. The PFM is a resource-distribution strategy used to reward LWDBs for their performance relative to seven performance metrics. In implementing the PFM, CSF's aim was to incentivize change and motivate local board leadership to increase efficiency and effectiveness. Ultimately, CSF hypothesized that this increased efficiency and effectiveness at the board level would result in better outcomes for individual clients (increased employment, increased quarterly wages, and decreased time to employment). The study therefore aims to assess the difference in employment and wage outcomes exhibited by clients of the state workforce system before and after policy change (we refer to the PFM as a policy herein as it was proposed by CSF and adopted statewide).

RESEARCH QUESTIONS

The study assesses three research questions identified by the program's theory of change.

1. ***Employment.*** *Do clients exposed to the PFM demonstrate greater gain in the likelihood of being employed from pre- to postprogram than a matched comparison group of clients who were not exposed to the innovation?*
2. ***Wages.*** *Do clients exposed to the PFM demonstrate greater gain in quarterly wages than a matched comparison group of clients who were not exposed to the innovation?*
3. ***Time to Employment.*** *Do clients exposed to the PFM become employed sooner than a matched comparison group of clients who were not exposed to the innovation?*

DESIGN

NATURAL EXPERIMENT WITH OFF-YEAR COMPARISON GROUP

The key challenge of a nonexperimental evaluation such as this is to identify how an intervention has affected participants without conflating it with alternative factors. Client employment outcomes, after all, are not merely a function of services and/or training received, but also by the individuals' background experiences, current circumstances, and personal attributes. The most effective way to ensure that these factors do not bias estimates in a study is by random assignment of participants to treatment and comparison groups. With random assignment, all relevant characteristics and experiences are balanced in expectation across the treatment and comparison groups, thus ensuring that that treatment assignment is independent of outcomes (and estimates remain unbiased). In the absence of random assignment, however, researchers must rely on nonexperimental techniques to establish a more qualified type of independence. For any nonexperimental study, the independence (i.e., unbiased estimate of) treatment on outcome is conditional on the statistical efforts to (actually) balance the two groups and the assumptions imposed. All nonexperimental work has this limitation.

PRG's approach was to assess the effect of the PFM by means of a nonexperimental design (NED). The design employs a natural experiment to compare client outcomes before and after the PFM was introduced in July 2015. The policy change, therefore, "assigns" participants to treatment and comparison groups in a way that approximates random assignment. In other words, because the time at which the policy change is introduced is arbitrary with respect to client characteristics, individual clients who enrolled in services either before or after the change should be similar in expectation. We use statistical modeling and propensity score weighting to adjust for the nonequivalence in labor markets experienced by the treatment (PFM) and comparison groups and any remaining nonequivalence in baseline characteristics. Detailed information about the merits and limitations of this design can be found in the Analytic Methods section.

POPULATION OF INTEREST

The population of interest was all LWDB clients in Florida who enrolled to receive services during the study period.[21] The analytic samples approximate this population closely, as shown in Tables 3 and 4. All clients who were included in the PFM participant data set, had complete outcome data, had data on relevant background characteristics (i.e., covariates), and enrolled in services during the study period were included in our analysis.[22, 23] The comparison study window began on July 1, 2012 and went through June 30, 2015. The treatment study window began on July 1, 2015 and went through June 30, 2018. For each analytic sample, enrollment into the study occurred during a portion of the study window. For Research Questions 1 and 2, enrollments in the first 8 quarters of each 12-quarter study window are included in the analytic sample. For Research Question 3, enrollments in the first 11 quarters of each 12-quarter study window are included in the analytic sample.

The selection mechanism for assignment into the treatment or comparison group is the "turning on" of the PFM. It was PRG's expectation that, putting aside their respective labor market experiences, this assignment mechanism would create two groups – a treatment group that received services while the PFM was operating and a comparison group that received services prior to introduction of the PFM – that are equivalent. Balance statistics presented in Tables C.1 and C.2 in Appendix C largely validate this expectation. The treatment and comparison study periods are of equal length, three years apart, and begin and end at the same time in the calendar year. Additional details on the study period can be found in Appendix E.[24]

TREATMENT CONTRAST

The comparison condition can be understood as a business-as-usual (BAU) contrast. The comparison group did experience the same workforce system (with a few notable exceptions that were outside the control of the evaluation) as the treatment group, minus the PFM.[25] Therefore, we are explicitly comparing labor market outcomes (wages and employment) for individuals who experience LWDB services with and without the PFM being active. The advantage of this approach is that it does not confound programmatic or unobservable local differences with treatment assignment, which would have been the case with any external (i.e., outside CSF's LWDB client population who would hypothetically benefit from the PFM) comparison group. The comparison sample was exposed to LWDB services from

[21] We laid out a sampling strategy in the initial analysis plan. However, because we were able to receive complete employment and covariate data from the state and in an effort to increase the generalizability of our results to the entirety of Florida's workforce clients, the benchmark analysis uses all available, complete, and within acceptable overlap data for all clients of LWDBs in Florida during the study period.

[22] Although named the PFM participant data set, this data set includes individuals interacting with LWDBs through WIA- and Wagner-Peyser-funded services during the comparison period as well.

[23] By "complete outcome data," we are referring to UI wage data that were considered certified by DEO. For all individuals enrolled during the treatment or comparison period, UI wage data were certified at the fourth quarter after enrollment. At the direction of DEO, any missing data in a certified UI wage data set should be interpreted as earning zero wages for that quarter.

[24] Initially, we proposed to observe outcomes three quarters post-exit from services, which meant that participants had to enroll *and exit* within the first nine quarters in their respective windows to be eligible for inclusion in the study. Although outcomes would be observed for any given client three quarters post-exit from services, the treatment and comparison windows consisted of a maximum of nine quarters in which the participants could potentially enroll and exit, plus three quarters to observe outcomes for clients who exited services in the ninth quarter. In a memo dated April 2018, we proposed to simplify the benchmark study period for Research Questions 1 and 2 by observing outcomes for all participants in the fourth quarter post-*enrollment*. To be included, participants need only enroll within the enrollment period and outcomes are observed four full quarters following the enrollment quarter.

[25] WIOA took effect on July 1, 2015. This new legislation required local boards to adjust to significant and mandatory changes in federal reporting by July 1, 2016.

July 1, 2012 to June 30, 2015, prior to implementation of the PFM. During this period, workforce services, on the whole, would be the same as they were during the treatment group study period (July 1, 2015 to June 30, 2018).[26]

The study, therefore, aims to assess the difference in employment and wage outcomes exhibited by clients of the state workforce system before and after the statewide innovation. The parameter of interest for each research question is a single coefficient that represents the average effect of the PFM on probability of employment, quarterly wages, and time to employment.

DATA SOURCES AND COLLECTION

No original data were collected for the study. Extensive individual-level outcome data, covariate data, and contextual/local economic data, used for propensity score weighting and analytical modeling, were collected from DEO and the U.S. Bureau of Labor Statistics (BLS). The evaluation required receipt of pre- and post-exposure individual-level Unemployment Insurance (UI) wage data (i.e., quarterly wages) for individuals who enrolled in services at one of the 24 local boards during the treatment or comparison enrollment period.

For each individual who engaged with a local board during the comparison or treatment enrollment period, we retained eight quarters of pre-exposure (i.e., baseline) wage data and up to eight quarters of data following their enrollment date.[27] Our benchmark analysis is based on the employment outcomes at the fourth full-quarter post-enrollment; due to the lag time in receiving certified wage data, data for all participants were complete at this time point.[28] The data received from DEO consist of four individual-level text data files. One file contains basic demographic information such as gender and race – we refer to this as the participant data set. Two of the files contain information specific to an individual's enrollment(s) within WIOA and/or Wagner-Peyser, such as entry and exit date, as well as demographic data collected at entry into those service programs.[29] Finally, a certified UI wage data file includes quarter, year, and the wages associated with each time period. All four data sets contain the same unique identifier that is used to link enrollees after receipt.[30]

In addition to data submitted by DEO, we accessed publicly available data from the USDOL Bureau of Labor Statistics (BLS). Data made available by BLS include the following contextual economic indicators: unemployment rate, total labor force, total employed, and total unemployed (all from the *Local Area Unemployment Statistics* [LAUS] program), and average weekly wages (from the *Quarterly Census of Employment and Wages* [QCEW]). These data were later used to calculate the workforce local board-level

[26] In order to allow for observation of outcomes for all participants, for Research Questions 1 and 2, we do not count enrollments in the third year of each group; therefore, enrollment in the treatment group took place from July 1, 2012 to June 30, 2014, and enrollment in the comparison group took place from July 1, 2015 to June 30, 2017.
[27] In the approved EDR, we outlined a plan using five quarters of pre-intervention wage/employment data prior to enrollment for each participant. Because we were able to obtain more retrospective data, we included eight quarters of pre-intervention data. Along with relevant and available covariates, these data were used for matching/weighting, pre-intervention trend diagnostics, baseline equivalence testing, and to construct the pre-program outcome observations in the analytic models.
[28] For participants who enrolled earlier in the comparison and treatment windows, we conducted a secondary analysis to analyze outcomes at the eighth full-quarter post-enrollment. The last two quarters of enrollees for the treatment group are dropped for this secondary analysis due to a lack of certified UI wage data at the eighth quarter post-enrollment. To balance the two groups, we also dropped the last two quarters of enrollees from the comparison group.
[29] For the sake of simplicity, any mention of WIOA programming or funding streams with regard to the outcomes study implies the WIA equivalent when considering data prior to the mandate to report using WIOA measures in July 2016.
[30] This ID was created by DEO, and all data deemed personally identifiable by DEO were removed prior to data transmission.

variables used in the analysis.[31] A more detailed account of the data sources, data collection, and data management procedures can be found in Appendix D.

VARIABLES

We use the administrative data described above in the analysis as individual-level covariates, time-variant and fixed economic and contextual variables, and dependent variables. All variables are defined in detail in Appendix D; in Table 2, we provide definitions of dependent (i.e., outcome) variables, as well as a list of all covariate and contextual variables that were used in analysis.

Table 2. Dependent and Independent Variables

Dependent Variables

Employment outcomes are assessed with three measures: employment status (i.e., whether one was employed during a given quarter), quarterly wages (i.e., the total wages earned in a quarter), and time to employment. We outline how these measures are constructed below.[32]

Employment Status	An individual is considered to be employed in any given quarter if he/she has earned $100 or more in the quarter being measured. Dummy variables are created for each quarter where 1 means that an individual is employed (according to the above definition), and 0 means the individual is not employed.
Quarterly Wages	This is a continuous variable measuring the amount of certified, employer-reported wages reported in the wage data set. All earnings variables are adjusted for inflation using the Consumer Price Index for all Urban Consumers (CPI-U) to the beginning of the treatment group outcome period – which is July 2017.
Time to Employment	Time to employment is measured using the same employment definition as Research Question 1; an individual who earns $100 or more in a quarter is considered to be employed.

Independent Variables

Here, we list the individual-level covariate variables, fixed contextual variables, and time-variant economic and contextual variables that we have used to describe our sample, and that have been considered for inclusion in propensity score estimating models in the analytic models.

Individual-level covariates	• Age at enrollment • Gender • Race/ethnicity • Region ID • Education level • Veteran status • Disability status • Reemployment Assistance claim paid	• Homeless • Offender • Low-income • Limited English proficiency • Single parent • Average wages prior to entry • Employment status prior to entry
Fixed contextual variables	• Quarter of entry • Month of entry • Number of days from beginning of study window to enrollment date • Number of days from beginning of quarter to enrollment date	• Cumulative number of enrollments in LWDB services • Funding stream (WIA/WIOA or WP) • Receipt of WIA/WIOA training services
Time-variant economic and contextual variables[33]	• Unemployment rate • Number in the labor force • Number employed	• Number unemployed • Average weekly wages

[31] The county-to-workforce board key was obtained here on December 10, 2018: https://careersourceflorida.com/your-local-team/

[32] We use the UI wage data as our data source for these variables. As such, we are limited in that we only have data on earnings that are required to be reported. UI wage data do not include earnings on some groups, including federal employees and self-employed individuals.

[33] Time-variant economic and contextual controls are included at the LWDB level for relevant quarters pre- or post-enrollment.

ANALYTIC METHODS

The only study design that permits for unequivocal determination of causation is a randomized controlled trial (RCT). In this case, PRG was unable to employ an RCT to assess the effect of the PFM because all LWDBs throughout the state were offered the chance to participate, which meant that random assignment of boards to control and treatment conditions was not possible. In the absence of an RCT, PRG adopted a design that is variously called an observational or nonexperimental design (NED). These designs attempt to isolate a causal relationship between an intervention and observed outcomes, but they are distinguished from RCTs in that none of them can do this with certainty.

In this study, we attempt to identify the unbiased effect of PFM on client employment and wage outcomes by comparing individuals exposed to LWDB services under the PFM policy (treatment group) with those who experienced LWDB services prior to the introduction of the model (comparison group). The principal identification assumptions of this design are that the assignment process of the natural experiment is arguably exogenous (so that both groups are equivalent in observed and unobserved characteristics) and that the statistical conditioning procedures (propensity score weighting, covariate balancing, and statistical modeling) are sufficient to remove bias in the estimate of treatment effect. We address these considerations concisely here, leaving a more detailed discussion of the methods to Appendices F, G, and I through K.

The validity of the natural experiment relies on the argument that all observed and unobserved factors are exogenous to treatment assignment and therefore individual participants are assigned to treatment and comparison groups in a quasi-random manner. This design employs a "natural experiment" to compare client outcomes before and after the PFM was implemented by CSF in July 2015. In this case, the PFM policy change "assigns" participants to treatment and comparison groups in a way that approximates random assignment. We say that it approximates random assignment because we expect that the assignment mechanism should isolate the treatment assignment from any systematic background differences in the two groups. In other words, because the time at which the policy change is introduced is arbitrary with respect to client characteristics, individual clients who enroll in services either before or after the change should be similar in expectation and well balanced across an array of characteristics.

For every NED, the claim of conditional independence ultimately rests on an assumption – because one cannot observe the validity of the claim on the (very important) unobserved characteristics that (more than likely) motivate the outcomes under study. Nevertheless, we do not simply make the assertion that our "natural" assignment mechanism has produced equivalent treatment and comparison groups. We present empirical evidence in the form of baseline equivalence statistics for our raw (i.e., unweighted sample) in Appendix C. The statistics demonstrate that baseline characteristics of both groups are well balanced and distributed in a way that is consistent with random assignment.

The principal source of conditioning uncertainty with this design is that the treatment and comparison groups are exposed to differing economic/labor market conditions at different times.[34] This imbalance in the experience of the two groups is apparent in the graphics that display the comparative historical labor market conditions for both groups (see Appendix B), in the average labor market conditions at time of

[34] Empirical descriptive data presented in Appendix B demonstrate that the two groups enrolled in LWDB services during dissimilar labor market environments. The PFM window was different from the comparison period because unemployment was lower, wages were higher, and labor market participation was higher. We can observe this difference, as well, in the graphic that plots the average individual-level quarterly wages (unweighted) across eight quarters prior to enrolling in LWDB services for both groups (in Appendix C). The treatment group has higher average wages, and it improves over time, prior to the final quarter, at a higher rate.

enrollment summarized in the unweighted baseline equivalence statistics (located in Appendix C), and in the scatterplot graphics of average wage earnings for both groups prior to enrolling in LWDB services (also in Appendix C).[35] The uncertainty arises from the question of whether or not the statistical methods used in this analysis are enough to control for these factors and remove bias from the estimates.[36]

This question is analogous to the selection issue that exists in any comparative study that compares participants who have not been randomly assigned to a program and relies on a set of observable conditioning characteristics to justify a conditional independence assumption. Such a study may rely on a variety of methods, but the identification strategy ultimately depends on the assumption that the conditioning set is enough to account for the unobserved selection process. Especially in the situation where individuals have selected into (and out of) a program (such as receiving or not receiving specific services) a confound exists because the motivational factors that predispose selection into those services (and therefore treatment assignment) are also likely related to the outcome and that (being unobserved) is confounded with the treatment status. Nevertheless, the argument is generally made that the conditional independence assumption can be satisfied by conditioning on prior wages and select background characteristics.

In particular, at least in labor market studies that examine wages and employment as outcomes, the argument is often made that realized pre-program labor market outcomes serve as a sufficient proxy for important missing unobserved characteristics (such as motivation or ability) and thus allow the conditional independence assumption to hold.[37] We think this makes sense and believe that the argument holds for varying economic conditions (occurring at different times) as well as it does for unobserved characteristics (selection confound). Specifically, those individuals who are exposed to a better job market will earn more and will earn increasingly more, so their wages prior to exposure will be higher and will increase at a greater rate than those who are exposed to a less vibrant job market.

The argument suggests that statistically adjusting for differences in labor market performance (i.e., modeling preenrollment trends) should be enough to identify the unbiased treatment effect and for the conditional independence assumption to hold. This is what we do, in any case. We describe our statistical modeling procedures below (and in more detail in Appendices G and I through K), but our approach is to first maximize the equivalence of the two groups on these important baseline characteristics (propensity score weighting). We then empirically control for differences in labor market conditions (and any remaining imbalance in individual characteristics) by statistically adjusting for pre-program wage and employment trends, time-variant average differences in labor market characteristics, potentially influential policy changes, and any remaining time-invariant individual characteristics.[38]

[35] Balance statistics demonstrate that the treatment and comparison groups are very similar across a broad array of observed individual-level characteristics, even without weighting. They also demonstrate that there is a substantial difference in the external labor market conditions for both groups.

[36] To the extent that we can successfully control for the actual (observed and unobserved) differences with the observed data (i.e., by modeling prior earnings trends and controlling for average labor market conditions and material policy change) we mitigate the so-called time confound. One point of emphasis: to call it a time confound distorts the identification issue at hand. Time is undoubtedly inextricable from (and therefore confounded with) treatment assignment, but it is not time itself that is expected to bias our estimates of the effect of PFM. Rather, it is the economic, labor market, and policy conditions that happen during those periods that may do so. And because we have data on those factors, we can control for them in our modeling and – at least arguably – satisfy the conditional independence assumption without a confound.

[37] Andersson, F., Holzer, H., Lane, J., Rosenblum, D., & Smith, J. (2013). *Does federally funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms*, NBER Working Paper No. 19446. Heinrich, C., Mueser, P., Troske, K., Jeon, K., & Kahvecioglu, D. (2013). Do public employment and training programs work? *IZA Journal of Labor Economics*, *2*, Article 6.

[38] In addition to the direct ways that differing labor market conditions (resulting from the asynchronous treatment and comparison periods) might insinuate bias into the estimates, there is also concern that a change in the labor market could insinuate selection effects such that people with differing characteristics (observed and unobserved) would differentially select into receiving services before and after the PFM was implemented.

We test this argument empirically by conducting a sensitivity analysis that contrasts wage and employment outcomes for the treatment group with a contemporaneous comparison group.[39] Details of this analysis are included in Appendix L. We use the same analytic procedures to compare employment outcomes for the treatment group with those for a subgroup of the comparison group that did not receive LWBD services at the time when the PFM was active (but did receive those services in the comparison period). We observe employment outcomes contemporaneously for both groups. Findings are consistent with our benchmark results.

BALANCING PROCEDURES

After study participants have been selected into treatment and comparison groups via the natural experiment mechanism outlined above, we generate propensity score weights to be used in our analytic models. Propensity scores seek to empirically optimize the balance of the two groups on an array of background characteristics and contextual factors (that have been observed in the data). Propensity scores predict the probability of being selected into the treatment group, based on an array of variables that are theoretically or empirically predictive of the outcome of interest. For each analytic sample, although the treatment and comparison groups were well balanced to start, we maximized the apparent balance of our sample using propensity score weights. We used the Generalized Boosted Model (GBM) to produce propensity scores that were included as Inverse Probability Weights in the analytic models to balance the two groups. Summary statistics of the baseline data after weighting suggest that the analysis compared groups that are very similar. We then used modeling procedures (with these weights included) to estimate the effect of the PFM on participants' wage and employment outcomes. A detailed account of the propensity score procedures, including variables included in the propensity score model, and weighting equation are provided in Appendix F.

Although propensity scores are widely used in applied research and are generally accepted as a legitimate method to satisfy the conditional independence assumption, the technique can be misused or misapplied.[40] In the case of this study, the efficacy of the balancing procedures can be (partially) observed by a comparison of the unweighted and weighted baseline equivalence graphics (located in Appendices C and M for unweighted and weighted, respectively) and by comparing the unweighted and weighted scatterplot graphs of mean wage and employment trends prior to enrolling in the study (again in Appendices C and M). The baseline equivalence dot graphs (Figures C.1, C.4, M.1, and M.5) show how weighting has improved the overall equivalence of the two groups on baseline characteristics by minimizing the standardized mean difference (SMD) between the two groups. The scatterplot graphics

---

Our design addresses this potential source of bias with propensity score weighting, which seeks to maximize the equivalence of the two groups on observed characteristics.

[39] In response to many external factors that changed since filing the EDR with WIF NEC, PRG also reassessed methodological decisions and examined the benchmark approach. During the proposed study period, the labor market in Florida changed and was exposed to transformative events. Unemployment rates declined from 11 to 4%, and the labor force expanded by over 1 million people. In the midst of the treatment window, the federal WIA was superseded by the WIOA. Two major hurricanes hit Florida and Puerto Rico in 2017. Significant events and trends such as these present challenges to the internal validity of the proposed benchmark design. With these in mind, alternative designs were considered but ultimately rejected. Since the PFM is a statewide policy, directed toward all 24 LWDBs within the state, none of the boards were excluded from the policy change. Two of the alternative designs that were considered were (1) to contrast the group exposed to the PFM policy change to an out-of-state group that was not exposed to the PFM; or (2) to compare one group of Floridian clients with others that may have received more (or less) exposure (i.e., in a dose-response relationship). The first option was excluded because data could not be obtained. The second was abandoned because the nature of what was being estimated was uncertain and undefined and was, therefore, deemed unsuitable as a benchmark approach. In the end, PRG decided to retain the original approach but augment it with additional analyses that will be used to test the robustness of those benchmark estimates. In addition, minor changes have been made to the study period, which are detailed in Appendix E.

[40] King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis.* Retrieved August 22, 2019, from http://j.mp/2ovYGsW. Heckman, J., Ichimura, J., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*(5), 1017–1098. Smith, J., & Todd, P. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1–2), 305–353.

(Figures C.2, C.3, M.2, and M.3) illustrate how weighting has minimized (but not completely removed) the difference in the location and slope of the wage trends for both groups and removed any difference that existed in the employment trends of both groups. This suggests that the analytic comparison (achieved by weighting) is between two groups that are more alike in their labor market performance and experience.

To empirically assess the question of whether or not application of propensity scores has increased bias, to respond to the criticism in the literature that propensity score methods may exacerbate selection issues if the required propensity score assumptions have not been met, and to determine the robustness of our preferred estimate, we conduct a sensitivity study that removes the propensity score weights from the preferred analytic models. We report these findings in the Effect of PFM on Employment and Wages section in this report and in detail in Appendices N and O.

MODELING

RESEARCH QUESTIONS 1 (EFFECT ON PROBABILITY OF EMPLOYMENT) AND 2 (EFFECT ON QUARTERLY WAGES)

For Research Questions 1 and 2, we used a comparative short interrupted time series (CSITS) design to estimate the effect of the PFM on the probability of employment and quarterly earnings. The CSITS is an analytic differencing approach that is a special case of the difference-in-difference (DID) design but, because it allows for statistical adjustment of differing trends, it is more appropriate in situations where the contrasted groups do not demonstrate common trends. We selected the CSITS based on an analysis of graphical scatterplots of mean wage and employment outcomes for eight quarters prior to enrollment in the study for both the treatment and comparison groups. An analysis of these graphs (reproduced in Appendix G) determined that the two groups did not evince parallel trends, which violates the key identifying assumption of the DID model. In contrast, the CSITS design explicitly accounts for nonparallel program trends (in addition to their relative mean values) by modeling them as separate linear functions.

Others have argued that selection bias is removed by conditioning on earlier labor market outcomes.[41] We argue that this same logic can be applied to market (history) bias. Highly motivated individuals will tend to do better than less motivated individuals in the same market. A corollary of this is that similarly motivated individuals will tend to earn more and earn at higher rates in a robust market than they will in a weak market. In other words, a labor market will tend to have consistent effects on similar individuals. Pre-program labor market outcomes are strongly predictive of post-program market outcomes both because they serve as proxies for motivation and ability, but also as proxies for the relative health of the labor market to which the participant is exposed. If so, modeling (conditioning on) this variation in relative pre-program wages and employment status – and the rate at which those outcomes improve over time – should help us to identify the effect of the PFM without the confounding influence of the varying economic conditions. A detailed discussion on the procedures for determining the appropriate model, model selection and specification, as well as empirical evidence is provided in Appendices G, I, and J. We test our modeling strategy specifically by way of a sensitivity study that employs a contemporaneous comparison group. Details of this study are described in Appendix L and results are provided in Appendices N and O.

In addition to propensity score weights and the CSITS modeling of pre-program outcome trends, we also included an array of variables to control for variations in local economic conditions, contextual policy change, local district effects, and residual covariate imbalance. Including measures of local labor market conditions as time-variant independent variables in the estimating equation was intended to aid in the

[41] See footnote 37.

identification of the effect of the PFM by controlling for any residual systematic differences that remained after controlling for pre-treatment trends.[42] The remaining time-invariant variables were intended to improve precision and adjust for baseline variation that remained after propensity score weighting. Although each of the variables has a theoretical justification for inclusion, we adopt an empirical selection process for the inclusion of independent variables in the estimating model. This involves an iterative model-fitting procedure, using goodness-of-fit statistics including log likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) to determine the best fitting model. Our preferred model in both cases includes the full set of independent variables. A detailed description of all variables can be found in Appendix D, and analytic model specifications, model selection procedures, and goodness-of-fit statistics can be found in Appendices I (Research Question 1) and J (Research Question 2).

RESEARCH QUESTION 3 (EFFECT ON TIME TO EMPLOYMENT)

Research Question 3 asks if clients exposed to the PFM become employed sooner than an equivalent group of clients who were not exposed to the PFM. We rely on the same natural experimental assignment mechanism and employ the same propensity score weighting procedures outlined above, but this research question requires that we use a different sample and empirical model to estimate an answer to the question. The sample – described below – is the subset of clients who enrolled in LWDB services within the study window, but who were unemployed at the time that they enrolled. And because the outcome of interest is the occurrence of an event (employment), we adopt a slightly different analytic approach.

When dealing with the analysis of event occurrence, standard regression techniques fall short because they must impose restrictions on cases that do not experience the event during the period of data collection.[43] These restrictions are either to put aside (exclude from analysis) or impose artificial outcomes (that the event did or did not occur). Both approaches are unsatisfactory because they impose assumptions that most likely will attenuate or inflate (bias) the estimated treatment effect. This is known as censoring, and it creates problems for the analysis of time to employment in a workforce training system because some clients will have 11 quarters to realize employment (if they enroll early in the study window), whereas others will have only one or two quarters to realize the outcome (if they enroll late in the study window). A conventional analysis would require that those who enrolled later in the study window be excluded from the study (perhaps attenuating the treatment effect because PFM is hypothesized to be working more effectively later in the study), or we would have to artificially impute one outcome or the other (employed or not employed).

A discrete-time hazard model avoids this by structuring the data into discrete periods in which the event (employment) may or may not occur. It sidesteps the censoring issue because it estimates risk (of becoming employed) not based on whether someone becomes employed or not during the study period, but whether or not someone becomes employed within each discrete time period (in this case a quarter) at which they are observed. Individuals enter the study when they enroll in LWDB services and are evaluated for that “risk” during each quarter. They either experience employment or they do not. When

[42] For our analytic models, we used LWDB-level BLS data on average weekly wages, the number of employed individuals, and the total number of individuals in the labor force. BLS data were downloaded at the county level and averages were created to represent the 24 LWDBs.

[43] Singer, J. D., Davidson, S., Graham, S., & Davidson, H. S. (1998). Physician retention in community and migrant health centers: Who stays and for how long? *Medical Care*, *38*, 1198–1213. Singer, J. D., & Willett, J. B. (1993). It’s about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, *18*, 155–195. Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

they exit the study (because they enrolled late in the study window) without experiencing the event, they are simply no longer part of the risk set.[44]

We use this statistical model to estimate whether implementation of the PFM reduces the amount of time someone remains unemployed after seeking assistance from a LWDB in Florida.[45] The model is fit by way of a maximum likelihood logistic regression model that estimates the likelihood of employment as a function of the same independent variables included in the previous models – group assignment (PFM or comparison), time period (each discrete quarter since enrolling in the study), individual-level covariates, blocking variables, and time-variant labor market conditions.[46] The coefficients produced by the logit model are in the form of conditional log odds ratios, which are difficult to interpret directly (but are reproduced in Table P.1 in Appendix P).[47] We convert these estimates to probabilities, which are easier to understand (see Table P.2 in Appendix P). Our principal mode of interpretation and explanation is through the production of tables and graphical illustrations that map the conditional probability of employment and the cumulative conditional probability of remaining unemployed for both the PFM and comparison groups over the duration of the study period. These are model-based estimates that illustrate the effect of the PFM for each discrete quarter during the study period. We also produce a variety of descriptive statistics, tables, and graphics (e.g., life tables and median "lifetimes") that aid in the interpretation of the inferential estimates; these are included in Appendix P.

As with the first two research questions, independent variables are operationalized and selected for inclusion based on theory, but we also adopt an empirical selection process for the specification of time and for the inclusion of independent variables in the estimating model. This involves an iterative model-fitting procedure, using goodness-of-fit statistics including log likelihood, AIC, and BIC to determine the best fitting model. Our preferred model includes the full set of independent variables and specifies time generally as a system of 11 quarterly dummy indicators. A detailed description of all variables can be found in Appendix D, and analytic model specifications, model selection procedures, and goodness-of-fit statistics can be found in Appendix K.

DESIGN MODIFICATIONS

After submitting the *Evaluation Design Report* (EDR) to the *WIF National Evaluation Coordinator* (WIF NEC) and after receiving final outcome data, PRG was compelled to make several alterations to analysis of the data.[48] Two changes were dictated by the size of the analytical data set, which contained 2.3 million

---

[44] Our observation period for the treatment and comparison groups is 11 quarters; however, due to the rolling nature of enrollment into workforce services (and as a result, the outcomes study), not all individuals in our analytic sample are examined for the full 11 quarters. In risk modeling language, the event occurrence is employment (or earning more than $100 dollars in the quarter), the beginning of time is the quarter that the individual enrolls in the study (by enrolling in services), and the discrete time period is the UI wage data reporting quarter.

[45] The sample for Research Question 3 included all unemployed individuals who enrolled in Wagner-Peyser or WIA/WIOA services at a LWDB during the first 11 quarters of each enrollment period (treatment and comparison). We did not include individuals enrolled in the last quarter of each enrollment period because their enrollment quarter was also the last quarter of observation, and an individual could not experience the event (employment) during their quarter of enrollment.

[46] In this model, we do not model prior labor market performance (as quarterly wages or employment) as we do in the analytic models for Research Questions 1 and 2. As such, identification of the treatment effect (and conditional independence assumption) relies on the time-variant statistical adjustment for local labor market conditions included as controls in the statistical model and, to a lesser extent, on the propensity score weights and natural experiment assignment.

[47] The coefficient of specific interest is the treatment group indicator that estimates the differential in conditional risk of employment for the PFM group (in log odds). Converted to probability, this value represents the effect of PFM on the probability of employment for any given discrete time period. Since the logit model is only linear in log odds, the magnitude of the effect varies across each time period, but it does so in a way that is conditional on the magnitude of the baseline risk (or risk of the comparison group).

[48] We are referring to the approved EDR as well as the memos titled "Evaluation Update Response to *WIF National Evaluation Coordinator* WIF Round 2 *Evaluation Design Report* Review" submitted September 30, 2015, and "Revisions to Impact Evaluation" submitted April 17, 2018.

observations (i.e., enrollments) initiated by 1.8 million unique individuals.[49] In the face of an analysis sample this size, our proposed propensity score-fitting procedures and multilevel specification of the first two analytic models were computationally intractable. Our analytic models would not converge and propensity score procedures produced unsatisfactory results. The final change was motivated by the distribution of outcomes for Research Question 1.

In addition to the changes outlined here that enable us to retain the full sample, we considered several alternative strategies that retained the initial analytic approaches but reduced the sample size – subsampling with multilevel models, and repeated subsampling with multilevel models and averaging parameter estimates (akin to bootstrap sampling). Ultimately, we settled on a tractable approach that retained the full sample because it offered the most reliable, valid, and transparent point estimate of the effect of providing PFM on labor market outcomes. An overview of these changes is provided below; details on the alternative strategies, and methods employed here can be found in Appendix H.

In the EDR, we initially proposed logit regression to estimate propensity scores. However, we had difficulty generating weights that improved the overall balance on observed pre-program characteristics. This may have been because the sample was already well balanced in participant characteristics (see Tables C.1 and C.2 and Figures C.1 and C.4 in Appendix C for unweighted baseline equivalence statistics); however, it was also certainly complicated by the size of the data set. The GBM is recommended by Guo and Fraser (2015) as an alternative method to seeking the best-fitting logistic regression model.[50] We found that the GBM was much more efficient and proved more effective at producing propensity scores that balanced the two groups; as a result, we used this approach in our analysis.

A second modification is our use of single-level analytic models to estimate the effect of the PFM on employment and wages. In the EDR, we proposed using multilevel models, but it was computationally infeasible to run multilevel models with weights on the full sample. Single-level models were the only practicable solution to retain the full sample. The main disadvantage of using a single-level model when a multilevel model is appropriate (as it is here with longitudinal observations) is that the standard errors will likely be downward-biased. The estimate of interest itself (i.e., the coefficient for the treatment variable) will be similar to that produced by the multilevel model, but any inference of statistical significance will be biased by a standard error that is erroneously too small. However, given that the sample size in Research Questions 1 and 2 is exceedingly large, this becomes less of a practical concern. This is so because the conventions of statistical significance have been established for samples of thousands – not millions, and (because of that) in samples this size almost every coefficient is significant.

Finally, in the EDR, we propose the use of a linear model to estimate the effect of the PFM on quarterly wages. Instead, we have selected a count model as our benchmark approach. We believe this decision is the most defensible, given the data. One methodological school says that a researcher should estimate parameters for an underlying probability distribution that best represents their data. Wage data are count data; they take on integer values that are nonnegative, and the distribution is patently right skewed. Moreover, errors are not normally distributed, which violates a key assumption of ordinary least squares (OLS) regression.

[49] This represents the size of the data set for Research Questions 1 and 2. For Research Question 3, the data set contained 1.2 million unique enrollments represented by about 1 million unique individuals.
[50] Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Another school of thought uses the Gauss–Markov and Central Limit Theorems to argue that, in the case of very large samples (such as this), OLS should perform better (i.e., produce the best linear unbiased estimate of the treatment effect) even if assumptions have been violated.[51] Some statisticians find this argument less convincing; however, in situations including long-tailed errors with correlation and nonconstant variance (such as this), we believe the prudent approach is to be cautious, and we adopt the count model as our benchmark approach. We test this analytic decision by including estimates from the OLS model alongside the preferred model estimates as a sensitivity study. Results are not identical but substantively alike. This decision is described in detail in Appendix J. Sensitivity results are presented in the narrative below as well as in Appendices N and O.

STUDY SAMPLE

COMPARATIVE LABOR MARKET CONDITIONS

As we have mentioned, a consequence of the NED with an off-year comparison group is that the treatment and comparison groups are exposed to different economic conditions. Here we provide a discussion of what the labor market statistics tell us, concretely, about the difference in labor market conditions experienced by the two groups.

We provide figures in Appendix B that plot statewide economic conditions during the treatment and comparison study periods. The comparison study period reflected in the figures is from July 2010 through June 2015, while the treatment study period is July 2013 through June 2018. Data were obtained from the BLS and cover the eight quarters prior to the enrollment period, the eight quarters of the enrollment period, and the four quarters following the enrollment period for the treatment and comparison periods.[52, 53] We present line graphs (to compare trends) and box plots (to assess overlap) for the number of individuals unemployed, employed, and in the labor force, the unemployment rate, the Florida minimum wage, average weekly hours worked, and average hourly earnings.[54]

Overall, Florida experienced a period of economic growth during the time from July 2012 through June 2018. The treatment group was generally exposed to a stronger labor market than the comparison group during its quarters of enrollment and outcome observation.[55] This is made evident by the line graphs, which indicate overall improvement in labor market outcomes across time and also show a consistent gap in outcomes between the treatment and comparison study periods. For example, Figure B.7 depicts the unemployment rate, which decreased at a consistent rate from 11.3 at the comparison group's earliest preenrollment quarter to 3.6 at the treatment group's final quarter of observation. The labor market had more participation as well – from the first quarter of fiscal year 2012 to the final quarter of fiscal year 2017, the number in the labor force grew from 9.3 to 10.3 million (see Figure B.1), and the number employed increased from around 8.3 to 9.9 million individuals (Figure B.3). Average hourly earnings

---

[51] Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*, 151–169.

[52] Although we present statewide figures in the report, it is important to note that the control variables used in our analytic models were specific to the local workforce board. To construct these variables, we downloaded county-level BLS data and calculated local board-level statistics based on which counties were represented by each board. For more on this process, see Appendix D. We also created graphics for each local board, but due to the number of figures produced, elected not to include them in our report.

[53] Graphics are associated with the two-year enrollment period for Research Questions 1 and 2.

[54] The federal minimum wage did not change throughout the entire study period. We include the Florida minimum wage for descriptive purposes, but we did not include this in our analytic process. The final labor market variables used in our analytic models included the number employed, number in the labor force, and average weekly wages.

[55] The eight quarters used as preenrollment observation quarters for the treatment group (from the first quarter of 2013 to the fourth quarter of 2014) are the same as the final eight quarters of the comparison study period.

increase for both groups over the course of the study, but the rate of increase is noticeably greater for the treatment group (see Figure B.13). In Figure B.11, showing average hours worked per week, the trendlines are not linear and cross over each other multiple times; on average, hours worked are between 34 and 36 hours per week. As described in detail above, we adjust for these differences in our statistical model primarily by trend adjustment and the use of time-variant covariates.

While finalizing our analytic models, we explored other policies, programs, or other external factors that may have affected local board operations during the implementation of the PFM.[56] The one policy that we ended up including as a control was the nationwide implementation of the WIOA, which required local boards to adjust to significant and mandatory changes in federal reporting by July 1, 2016.[57] In implementation evaluation interviews, CSF and local board directors also cited several other workforce initiatives that occurred during the grant period, but project leadership stressed that these programs were insignificant in their impact on PFM implementation; as such, these were not controlled for in the analysis.[58]

Finally, Florida experienced three separate hurricanes during the PFM implementation period: Hermine (2016), Matthew (2016), and Irma (2017); Maria (2017) also impacted the state due to displaced persons moving to Florida from Puerto Rico. Recovery from these events required a significant time commitment from CSF and DEO staff, but despite the significant impact these events had on the state of Florida, the PFM team reported minimal disruption to PFM implementation activities. In addition, the inclusion of the labor market variables discussed above should help to adjust for these natural disasters because the metrics will naturally respond to the influence that they have on the labor market.[59]

PARTICIPANT CHARACTERISTICS

RESEARCH QUESTIONS 1 AND 2

Table 3 presents background characteristics for the analytic sample for Research Questions 1 and 2. The table reports descriptive statistics on background characteristics for two distinct groups – individual clients served by LWDBs during the study period (right column), and unique cases included in the analysis (left column). The right column describes the sample of clients served by CSF during the study period and is structured such that the client is the unit of analysis. The left column describes the analytical sample when the client case becomes the unit of analysis. The numbers differ because the right column includes a census of all unduplicated clients, whereas the left includes a count of all cases including duplicated enrollments by the same individual. Because the eligibility requirements of the study allow for the possibility of a client enrolling in the study on multiple occasions, it is the left column that enumerates the analytic sample.[60] Background characteristic data were overall complete.[61, 62]

---

[56] We are aware that individuals in our treatment group may have been exposed more heavily to job opportunities where wages are not reported to the *U.S. Department of Revenue*, such as "gig" jobs, which have experienced a growth over the last several years. We believe that the labor market variables we have chosen are sufficient to control for any of the contextual factors that were different for the treatment and comparison groups over time.

[57] This is controlled for by way of a dummy variable that is a 1 for anyone who enrolled after July 1, 2016, and a 0 for individuals enrolled before that date.

[58] These programs include *Governor's Reemployment Challenge*; *Florida Job Placement Report*; *Florida Talent Prosperity Dashboard*; *Florida Chamber Foundation's Prosperity Initiative*; *The Florida Scorecard*; *Community Service Block Grant Program*; and *Employment First Florida*.

[59] As a reminder, the labor market control variables were incorporated into the analytic models at the local board level, so if one or more local areas was affected more heavily by these hurricanes, the statistics would reflect that level of detail.

[60] See Appendix E for details on the identification of the treatment and comparison groups.

[61] Thirty-five individuals were dropped due to invalid or out-of-range ages.

[62] The source of the majority of these data is the PFM participant data set; however, a few variables are available for WIA/WIOA participants only.

We requested data on all individuals served by Wagner-Peyser, WIA, or WIOA during the study period from DEO. The study period for Research Questions 1 and 2 ranges from July 1, 2012 through June 30, 2014 for the comparison group, and from July 1, 2015 through June 30, 2017 for the treatment group. The sample, therefore, should represent the entire population of clients that accessed (i.e., enrolled in) LWDB services funded by one of these programs during the study period.[63]

[63] The switch from WIA to WIOA took place on July 1, 2016; therefore, we have clients served by both programs in our treatment group. The comparison group does not contain any WIOA participants, only WIA.

Table 3. Full Benchmark Analytic Sample, RQ 1 & 2

Characteristic	Unique Enrollments (Analytic Sample) Percent/Mean	Unique Individuals Percent/Mean
	(*n* = 2,323,339)	(*n* = 1,776,476)
Mean age at enrollment	39.3	39.5
Education level		
Did not complete high school	8.8%	8.9%
Obtained high school diploma or higher	91.2%	91.1%
Veteran status		
Yes	6.2%	6.1%
Disability status		
Yes	5.0%	4.8%
Welfare Transition Program participant		
Yes	2.9%	2.6%
Supplemental Nutrition Assistance Program participant		
Yes	4.2%	4.0%
Reemployment Assistance claim paid		
Yes	20.0%	21.1%
Employment status at enrollment		
Unemployed	82.4%	83.6%
Employed	16.6%	15.5%
Employed with termination or military separation	0.9%	0.9%
Gender		
Female	52.7%	51.8%
Male	47.3%	48.2%
Race/ethnicity		
Hispanic/Latino	25.8%	26.3%
Haitian	2.0%	1.9%
American Indian/Alaska Native	1.3%	1.3%
Black/African American	29.7%	27.2%
Native Hawaiian/Other Pacific Islander	0.5%	0.5%
White	54.3%	55.9%
Asian	1.4%	1.4%
Other race	0.0%	0.0%
	Reduced sample: WIOA enrollees only	
	(*n* = 98,466)	(*n* = 58,496)
Homeless		
Yes	1.8%	1.7%
Offender		
Yes	10.0%	8.9%
Low income		
Yes	100%	100%
Limited English proficiency		
Yes	2.1%	2.0%
Single parent		
Yes	14.9%	13.5%
Received WIA/WIOA training service		
Yes	70.9%	73.8%

As shown in Table 3, our analytic sample for Research Questions 1 and 2 comprises 2.3 million unique enrollments that were initiated by 1.8 million unique individuals during our study period. The mean age for both samples is just over 39 years, and the majority identify as female.[64] Just over one half identify as white, around 30% as Black or African American, and one quarter as Hispanic/Latino. Over 90% are high school graduates. Roughly 20% of unique enrollments had a Reemployment Assistance claim paid while receiving services. Very few enrollees are veterans, disabled, SNAP recipients, or *Welfare Transition Program* participants. Finally, all WIOA enrollments in our analytic sample are classified as low-income, 71% have received a training service, and a much smaller percentage are single parents (15%), offenders (10%), or have limited English proficiency or are homeless (both 2%).

RESEARCH QUESTION 3

We now present the background characteristics for the analytic sample for Research Question 3. We report Research Question 3 separately because the sample for this analysis is reduced as the analysis is focused on employment effects for those who were unemployed at the quarter of their enrollment in workforce services.[65] The study period for Research Questions 3 ranges from July 1, 2012 through March 31, 2015 for the comparison group, and from July 1, 2015 through March 31, 2018 for the treatment group. The sample, therefore, should represent the entire population of unemployed clients that accessed (i.e., enrolled in) LWDB services funded by one of these programs during the study period.[66] We again report in two groups: unique individuals and unique enrollments, the latter of which is our analytic sample.

[64] In the initial data set we received from DEO, which covered enrollments from January 2006 through May 2019, the majority were male. However, once we created our treatment and comparison groups and the data set reduced in size, the majority became female.
[65] We define unemployment as less than $100 in wages in the quarter of enrollment.
[66] See footnote 63.

Table 4. Full Benchmark Analytic Sample, RQ 3

Characteristic	Unique Enrollments (Analytic Sample) Percent/Mean	Unique Individuals Percent/Mean
	(*n* = 1,214,269)	(*n* = 999,007)
Mean age at enrollment	39.2	39.3
Education level		
Did not complete high school	10.4%	10.3%
Obtained high school diploma or higher	89.6%	89.7%
Veteran status		
Yes	7.3%	7.1%
Disability Status		
Yes	6.9%	6.4%
Welfare Transition Program participant		
Yes	5.2%	4.7%
Supplemental Nutrition Assistance Program participant		
Yes	5.8%	5.6%
Reemployment Assistance claim paid		
Yes	10.4%	11.1%
Employment status at enrollment[67]		
Unemployed	91.1%	91.2%
Employed	8.3%	8.3%
Employed with termination or military separation	0.5%	0.6%
Gender		
Female	53.3%	52.5%
Male	46.7%	47.5%
Race/ethnicity		
Hispanic/Latino	25.6%	26.2%
Haitian	2.0%	1.9%
American Indian/Alaska Native	1.5%	1.5%
Black/African American	30.7%	28.7%
Native Hawaiian/Other Pacific Islander	0.5%	0.5%
White	52.6%	53.7%
Asian	1.5%	1.5%
Other race	0.0%	0.1%
	Reduced sample: WIOA enrollees only	
	(*n* = 55,636)	(*n* = 36,582)
Homeless		
Yes	2.5%	2.3%
Offender		
Yes	13.8%	13.0%
Low income		
Yes	100%	100%
Limited English proficiency		
Yes	1.9%	1.8%
Single parent		
Yes	15.1%	14.0%
Received WIA/WIOA training service		
Yes	69.2%	71.3%

[67] An eligibility requirement for Research Question 3 was that an individual was unemployed at enrollment in LWDB services. Employment status was determined using the UI wage data for the individual's quarter of enrollment. There is some variation between that variable and the administrative employment status variable presented here, which is to be expected when pulling data from two distinct administrative sources.

Table 4 shows that, again, the analytic sample for Research Question 3 is similar to the unduplicated sample. Included are 1.2 million unique enrollments represented by about 1 million unique individuals. For the analytic sample, we see that the mean age is 39 years and 90% have a high school diploma. The majority are female and White, a quarter identify as Hispanic/Latino, and around 30% identify as Black or African American.

Background characteristics of the unique enrollment and unique individual groups are mostly consistent across samples for Research Questions 1 through 3. However, the sample for Research Question 3 is slightly more disadvantaged. This is evidenced by a higher proportion of individuals who have a disability (7% vs. 5%), who are *Welfare Transition Program* participants (5% vs. 3%), who are SNAP recipients (6% vs. 4%), and who are offenders (14% vs. 10%). These differences are likely driven by the eligibility criteria for Research Question 3 – that an individual is unemployed at enrollment – as these qualities identify individuals who often face barriers to employment.

BASELINE EQUIVALENCE

We describe the equivalence of the PFM and comparison groups by means of a metric that quantifies the difference of the two groups in units that are standardized by the variation that is exhibited by the sample. This difference is called the standardized mean difference (SMD), and we report this difference across a range of baseline characteristics that have been observed in the data.[68] The SMD allows one to gauge the similarity of the PFM and comparison group in a way that is insensitive to sample size and raw metric.[69] The equivalence of the two groups is an important consideration for assessing the validity of the analysis because it helps to establish evidence that selection effects are negligible in any subsequent estimate. That estimate becomes more valid to the extent that the two contrasted groups are equivalent – or nearly so – and to the extent that the range of baseline characteristics that are used for comparison are a convincing conditioning set. Figures 2 and 3 present the baseline balance diagnostics for the PFM and comparison groups in the form of standardized differences, first for Research Questions 1 and 2 and then for Research Question 3.[70] The vertical gray dashed lines within each figure indicate standardized differences that are equal to or less than 0.1 and 0.25 standard deviations. Although there is no universal rule, it has become conventional in some disciplines to consider a SMD less than 0.10 as well balanced and less than 0.25 as an acceptable level of balance.[71]

[68] For continuous variables, the SMD is represented by Hedges' *g*. This is defined as the difference between the regression-adjusted means for the treatment and comparison groups, divided by the pooled within-group standard deviation of the baseline measure. For dichotomous variables, the difference in group means is calculated as the difference in the probability of the occurrence of an event. The effect size measure of choice for dichotomous variables is the Cox Index, which yields effect size values similar to the values of Hedges' *g* that one would obtain if group means, standard deviations, and sample sizes were available, assuming the dichotomous outcome measure is based on an underlying normal distribution. See Appendix F for formulas detailing these calculations. What Works Clearinghouse. (2017). *Procedures Handbook Version 4.0*. Retrieved July 25, 2019, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

[69] Significance testing is inappropriate for this diagnostic task because it conflates balance with statistical power. Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399–424. Austin, P. C. (2009). The relative ability of different propensity-score methods to balance measured covariates between treated and untreated subjects in observational studies. *Medical Decision Making*, *29*, 661–677. doi:10.1177/0272989X09341755. Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society*, *Series A*, *171*, 481–501. Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*, 1–21.

[70] Baseline equivalence statistics for the unweighted samples, which present a descriptive summary of the equivalence of both groups prior to propensity score weighting, can be found in Appendix C.

[71] Although there is no consensus on what value denotes balance, there is some agreement that regression adjustment is trustworthy when the SMD is .25 or less. For example, see Stuart, E. A. (2010) and Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169–188. The *What Works Clearinghouse* specifies that differences less than or equal to 0.05 standard deviations requires no statistical adjustment to be considered equivalent. For differences between 0.05 and 0.25 standard deviations, an analysis must include an acceptable statistical adjustment for the baseline characteristic to meet equivalence standards. Differences above 0.25 standard deviations in value are considered to be nonequivalent. What Works Clearinghouse. (2017). *Standards Handbook Version 4.0*. Retrieved July 25, 2019, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Figure 2. *Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2*

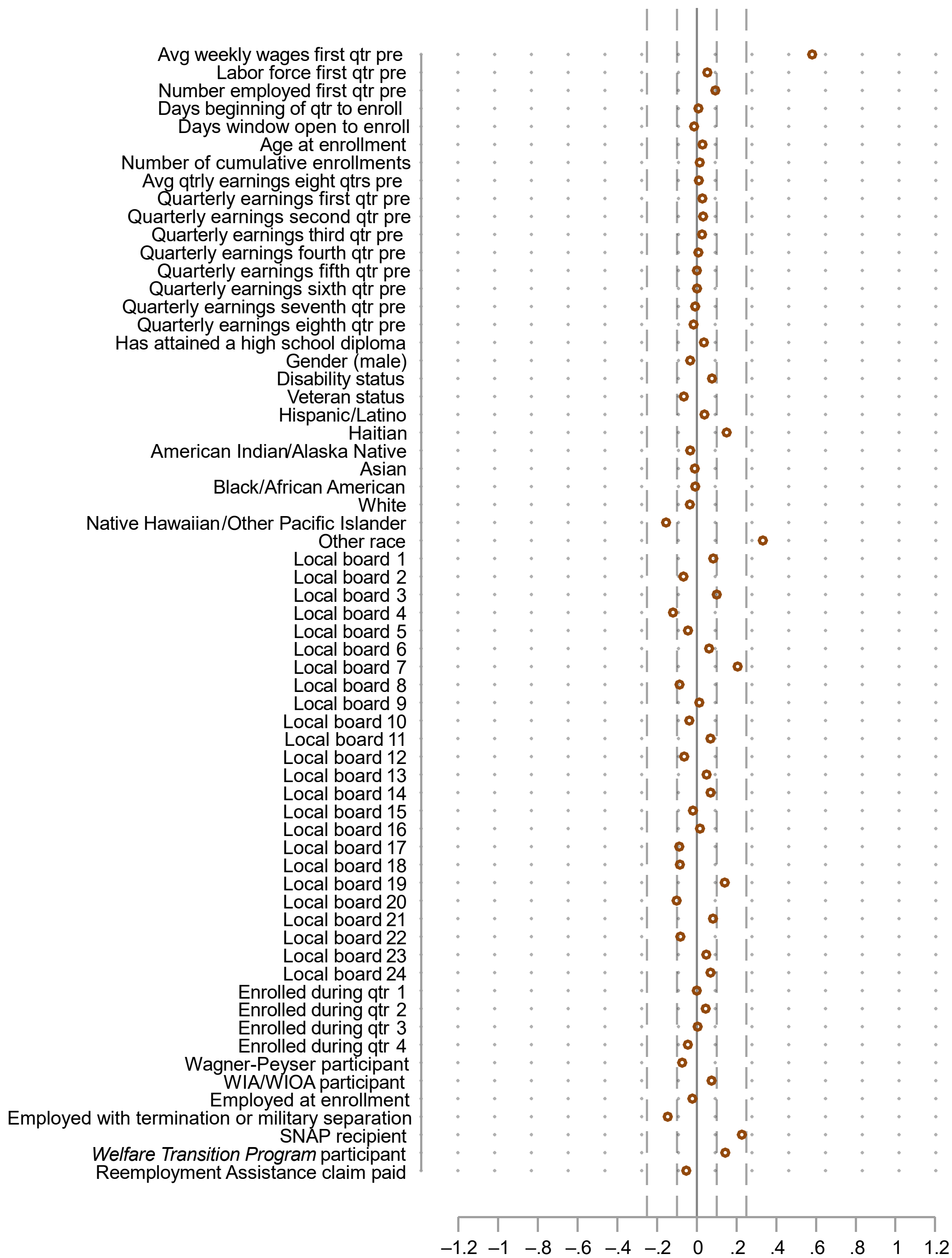
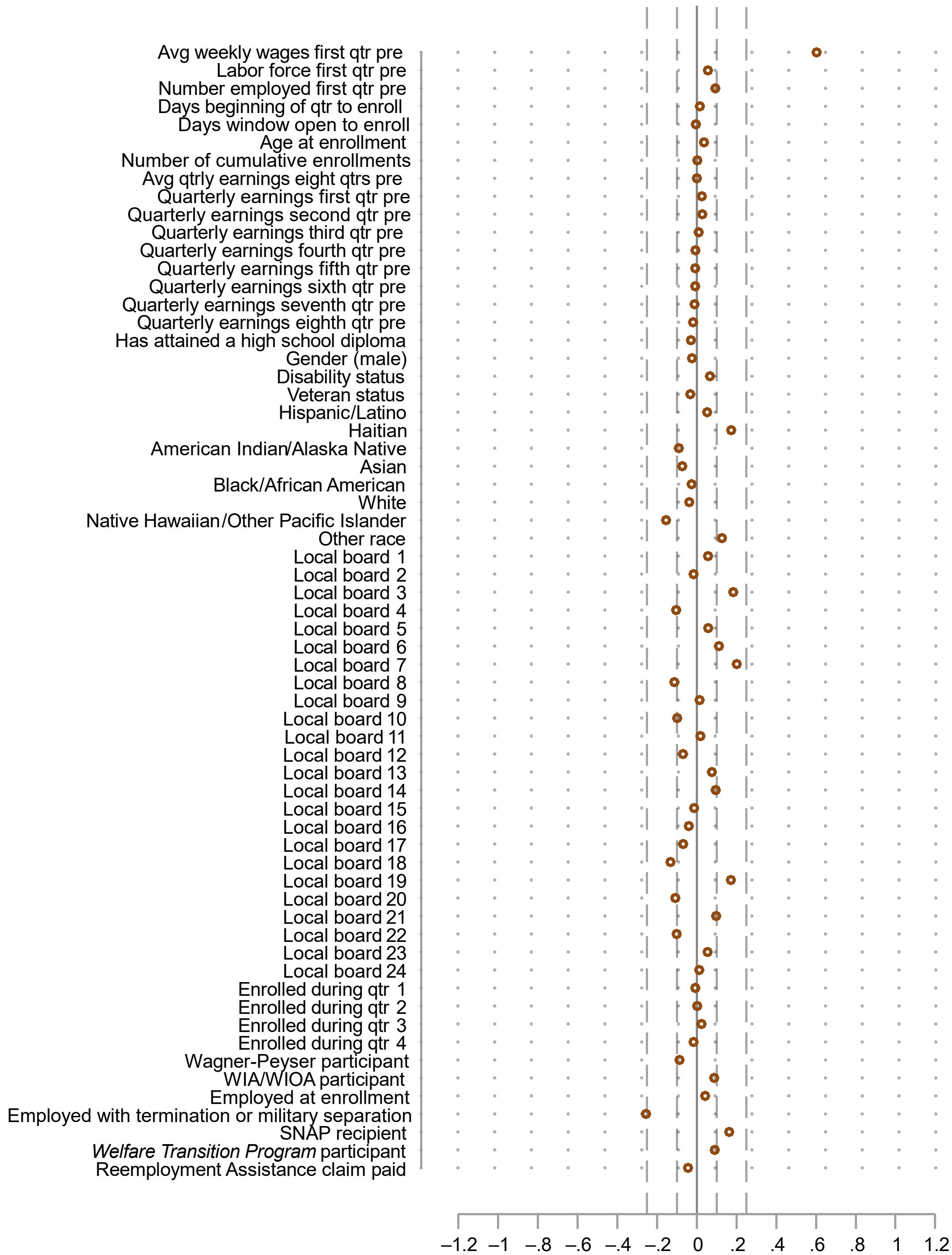


Figure 3. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3



Figures 2 and 3 illustrate that treatment and comparison groups are largely equivalent across the full range of the observed baseline characteristics that are available in the data. SMDs between the treatment and comparison groups are small in magnitude and are distributed in a way that appears unsystematic. In fact, there are only two variables in both samples that demonstrate a SMD that is greater than 0.25.

The percentage of cases that identify as Other race is much larger for the treatment group than the comparison group. This no doubt reflects some feature that differentiates the experience of the PFM and the comparison groups that remains unaccounted for by the propensity score weighting. However, given that less than 0.1% identify as Other race in the full sample, this difference in the SMD statistic is a function of small differences in small proportions in a large sample.

Baseline equivalence statistics also show that the two groups differ in terms of their average weekly wages in the first quarter prior to enrollment. Part of this discrepancy is a function of the difference in variability of the two measures.[72] However, consistent with what we have discussed above (regarding the comparative labor market conditions and earnings curves experienced by both groups), the treatment group experienced a labor market that was larger, more competitive, and more remunerative than that experienced by the comparison group. It was also growing in these features at a higher rate. What we see, in other words, are two groups that are remarkably similar in individual-level baseline characteristics but differ in their labor market experiences.[73]

As we explain elsewhere, we identify the treatment effect by controlling for these differences in the analytic model, which includes adjustments for the differing trends in pre-program outcomes evidenced by clients in both groups and with time-variant controls (using these same average local-level statistics) for the labor market experiences of each client. Full baseline equivalence statistics for all samples are presented in Appendix C (unweighted) and Appendix M (weighted).

EFFECT OF PFM ON EMPLOYMENT AND WAGES

KEY FINDINGS

RESEARCH QUESTION 1: EFFECT ON PROBABILITY OF EMPLOYMENT

Preferred model estimates (reported in Table N.1 in Appendix N) indicate that four quarters after enrollment, the PFM had a marginal positive effect on the employment outcomes of CSF clients included in the study.[74, 75] Linear transformations of model coefficients suggest that, for the typical or average client, "turning on" the PFM increases the probability of employment by 1.7% one year after enrolling in LWDB services.[76] Figure 4 illustrates the model-adjusted preenrollment trends (solid lines), post-enrollment projected trends (dashed lines), and predicted average post-enrollment employment probabilities for the treatment and comparison groups.

[72] This contrast in individual-level similarity and average labor market experience even persists when we consider prior earnings (for as many as eight quarters prior to enrollment). How can two groups have a large SMD in average labor market conditions but very small SMD in prior earnings? The answer is largely a function of the variability of the two measures. The average measure does not vary as widely as the individual-level quarterly wages.

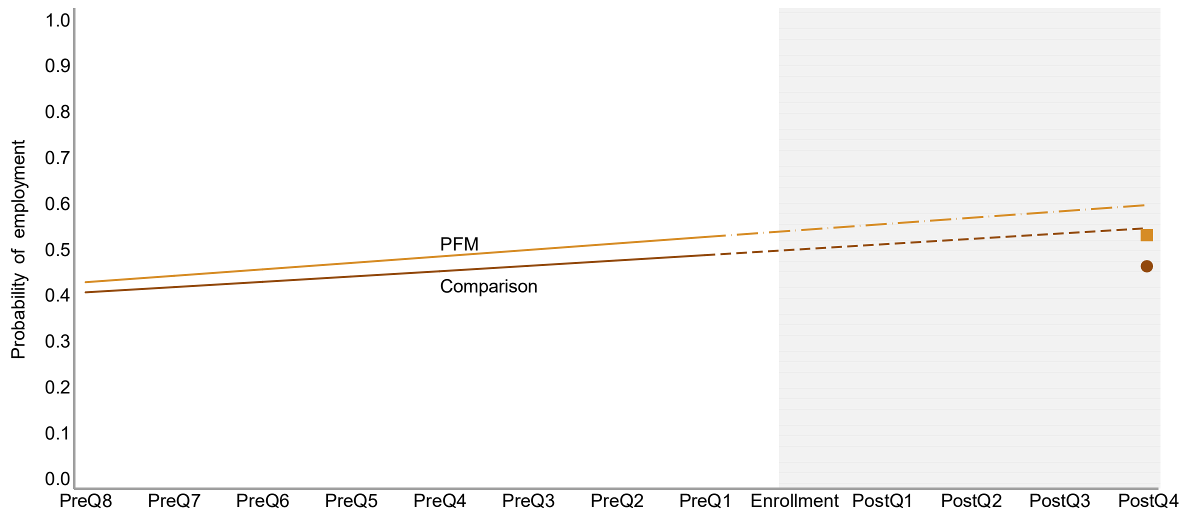
[73] This difference persists after propensity score matching because external labor market conditions have not been included in the propensity score model. When we included local board-level average labor market indicator variables in the propensity score model, they produced a perfect prediction of the propensity score and resulted in no overlap in scores for both groups.

[74] The preferred model is a single-level logistic regression that includes controls for individual-level covariates, time-variant economic and contextual variables, local board-level controls, and an additional contextual variable that controls for the policy switch from WIA to WIOA, a change that took place nationwide during the treatment group study window. For details on the preferred model specification for Research Question 1, see Appendix I.

[75] In Implementation Evaluation interviews, CSF also cited the following initiatives that occurred during the grant period but were insignificant in their impact on PFM implementation: *Governor's Reemployment Challenge*; *Florida Job Placement Report*; *Florida Talent Prosperity Dashboard*; *Florida Chamber Foundation's Prosperity Initiative*; *The Florida Scorecard*; *Community Service Block Grant Program*; and *Employment First Florida*. These were not controlled for in the outcomes analysis.

[76] The coefficients of logit model are not easily interpretable on their own because they represent values that are transformed by the logit link function. We calculate linear probability estimates of employment from model estimates by transposing the result of the regression equation at fixed values for the CSITS variables and all remaining control variables held at zero. Linear transformation of the coefficients with all control variables held at zero produces estimates of impact for the "typical" or average program participant because we have mean-centered the control variables.

Figure 4. Pre- to Post-Program Change in Probability of Employment



As shown in Figure 4, for the eight quarters prior to enrollment, the solid lines illustrate the model-adjusted average pre-program probability of employment. The lines indicate that, conditional on the model, both groups demonstrate a gradual but substantial improvement in mean employment outcomes across the eight quarters prior to enrollment.[77] For the comparison group, the average adjusted probability of employment improves from around 0.40 to 0.49. The treatment group, meanwhile, improved from 0.43 to 0.52.[78] Although it is not easily perceptible in the graphic, even after weighting and other statistical adjustments, the two trends are significantly divergent in slope, with the PFM group improving at a slightly higher rate than the comparison group.[79]

At the quarter prior to enrollment (PreQ1), the lines become dashed to illustrate that they are a projection of the baseline trends for the typical program participant in each group. If these trends were to continue, and excluding any difference due to program exposure, the model-based projections suggest that the comparison and treatment groups would be expected to realize employment probabilities of 0.54 and 0.59, respectively, four quarters post-enrollment in LWDB services. These estimated values represent the hypothetical employment outcomes for each group in the sense that they account for the pre-program differences in the two groups, but they exclude any post-program information. Estimates of post-program employment outcomes are plotted as points below the two lines: 0.46 for comparison and 0.53 for treatment.

The estimated employment outcomes for both the PFM and comparison groups are below the projected lines. This implies that four full quarters after enrolling in LWDB services, the difference in hypothesized and observed employment outcomes are negative.[80] It would, however, be an interpretive mistake to infer from this initial difference that service programs are underperforming or that clients' economic outcomes

[77] The linear trends illustrated in Figure 4 are a feature of the analytical model that incorporates distributional assumptions and adjustments for covariates and propensity score weights. Figure C.2 in Appendix C illustrates the unadjusted pre-program mean probabilities of employment for both groups.
[78] Model-based predicted probabilities are produced in Table N.2 in Appendix N.
[79] Coefficients for the treatment and trend interaction demonstrate this difference in log odds scale. See Table N.1 in Appendix N.
[80] Differences are –.08 (.46 – .54) for the comparison group and –.06 (.53 – .59) for the treatment group.

have diminished as a result of program exposure.[81] The negative difference in observed outcomes versus those implied by pre-program trends is a consequence of the employment interruption (unemployment or underemployment) that has confronted this population and the very reason that they have sought out the services provided by the LWDBs. In other words, observed employment outcomes are lesser than the projected line because of whatever motivated enrollment in LWDB services in the first place (e.g., job loss) and do not necessarily reflect the performance of either service platform.

Like the DID model, of which it is a special case, the CSITS model estimates program effect as the difference between each of the groups' projected and observed (mean) results. This estimate is nothing more (or less) than the difference in those first differences. In this case, the difference is represented by the relative – and barely perceptible – difference in distance of the points from the lines. For the "typical" program participant, the difference between the treatment group's mean projected and observed employment probability is 1.7% higher than that of the comparison group. In other words, the treatment group has an estimated 0.017 higher probability of employment post-enrollment than the comparison group.[82]

The estimate represents the single best point estimate of program effect, provided by the preferred model.[83] Estimates are statistically significant, but it is worth emphasizing that with sample sizes this large (more than 20 million observations), statistical significance is a poor gauge of substantive importance.[84] The magnitude of the estimated effect of PFM is conventionally considered small (effect size = 0.04).[85] With a few assumptions, this denotes that the relative "risk" of employment for the PFM group relative to the comparison group is 1.006, which means that for every 1,000 clients in the comparison group who find employment, we would expect 1,006 clients in the treatment group to find employment. In our analytic sample, the number of employed individuals in the comparison group is 764,803; using the higher probability of employment attributed to the treatment group, we would expect an additional 4,589 treatment group members to be employed than comparison group members.

Owing to the large sample size and the fact that the analytic sample is effectively the full population of clients who received Wagner-Peyser or WIOA services in Florida during the study window, we are generally confident that the preferred benchmark model has produced a reasonably accurate nonexperimental estimate of the effect of the PFM. The constraints and assumptions of the analysis do insinuate some uncertainty, however. That uncertainty is principally derived from (1) the time at which we observe the employment outcome, and (2) specific design and analytic choices that we have made. We discuss each, in turn.

First, we are estimating the effect of the program at one point, four quarters following enrollment in LWDB services. For the most part, this constraint is imposed externally by the implementation dates of the

[81] This is evidence of why the Interrupted Time Series (ITS) design would be inappropriate for this study. All effects would be negative – at least in the allowable scope of the evaluation/grant period.
[82] As previously mentioned, these estimated probabilities are calculated on the basis of all control variables in the analytic model being held at zero (mean value). Due to the nonlinear nature of the logit model, the predicted values will vary based on the value of pre-program characteristics and wages that are incorporated into the predictive algorithm.
[83] Model selection and specification procedures for Research Question 1, along with statistics and analysis of those statistics, are presented in Appendix I. The preferred model estimate is reported in log odds-ratio format as the coefficient for the interaction of the treatment and post-enrollment indicator term in (the fifth row) Table N.1 in Appendix N.
[84] Further, standard errors produced by the benchmark model are likely artificially too small (biased) because the single-level model does not account for the nonindependence of longitudinal (within-person) observations. This will not bias the coefficient itself, but it will tend to overstate significance levels. For details on the use of single-level models with this large sample, see the section on Design Modifications.
[85] Effect size is calculated as the Cox Index.

program, the length of the evaluation contract, and the necessity to measure outcomes for multiple years of program implementation. The literature on active labor market programs suggests that employment and wage outcomes tend to become manifest one year after the quarter of enrollment in LWDB services.[86] However, that literature also shows that outcomes can be highly variable over time. This means that the outcomes observed four quarters post-enrollment may capture the effect of the program, but that those outcomes may be highly transitory.

To test this, we ran a secondary analysis on a smaller sample using quarterly wages at the eighth quarter after enrollment.[87] This study supports the positive benchmark results, showing that the treatment group had a 6.5% higher probability of employment than the comparison group at eight quarters post. As shown in Figure N.1 in Appendix N, both groups' observed employment probabilities are lower than the projected probabilities had their baseline trends continued, but the decrease is greater for the comparison group.[88]

There is also uncertainty that derives from the design and analytic choices that we have made. This is a feature of all nonexperimental studies. In this case, we have selected as our benchmark approach a natural experiment with a non-contemporaneous comparison group that could confound differences in labor market and policy conditions with the treatment effect. The design itself cannot control for these differences, and so we rely on our statistical model to adjust for these differences (weighting, covariate, and trend adjustments). Although we believe that this strategy is as defensible as those used to control for selection bias in a nonequivalent, contemporaneous comparison group approach, we concede that it is difficult to confirm that we have effectively controlled for these differences.

We test this assertion empirically by conducting a sensitivity analysis that contrasts employment outcomes for the treatment group with a contemporaneous but nonequivalent comparison group. Details of this analysis are included in Appendix L (and results in Appendix N), but, in brief, we use the same analytic procedures to compare employment outcomes for the treatment group with those for a subgroup of the comparison group that did not receive LWBD services at the time when the PFM was active (but did receive those services in the comparison period). We observe employment outcomes contemporaneously for both groups. Estimates from the sensitivity analysis suggest that the effect of the PFM relative to the contemporaneous comparison group is essentially null.

This analysis suggests that our benchmark estimate may be very slightly overstating the effect of the PFM. One hypothesis is that the fixed-effects model and dichotomous outcome inhibit the efficacy of the statistical controls intended to adjust for the different labor market circumstances experienced by the PFM and comparison groups. However, it seems equally likely the difference in estimates may also be attributable to the dissimilar counterfactuals implicit within both studies. In the benchmark study, we are comparing two groups who have experienced an employment shock prior to enrolling in LWDB services and the study. In the sensitivity study, we are comparing the PFM participants to a comparison group that

[86] Research on active labor market programs suggest that job training programs tend to have small or negative effects on employment outcomes for periods of less than one year. For examples, see Heinrich, C., et al. (November 2009). *New estimates of public employment and training program net impacts: A nonexperimental evaluation of the Workforce Investment Act Program*, IZA Discussion Paper No. 4569. Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *Economic Journal*, *120*(548), F452–F477. Andersson, F., et al. (September 2013). *Does federally-funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms*, IZA Discussion Paper No. 7621.
[87] As we did not have eighth quarter outcome data for the final two quarters of enrollees for the treatment group; and in order to keep the balance between the two groups, we also dropped the final two quarters of enrollees from the comparison group.
[88] The projected probability of employment for the comparison group was 0.57; the observed probability was 0.45. The projected probability of employment for the treatment group was 0.63; the observed probability was 0.57.

has not experienced a similar shock (immediately prior to enrollment). As such, we might expect to see more muted effects for the PFM group produced by this sensitivity study.

Because our estimates may also be sensitive to a variety of other analytical decisions, we also include results from three additional studies. We conduct a sensitivity test of the preferred statistical model without propensity score weights included, the preferred statistical model without covariates included, and the preferred statistical model fit with OLS regression instead of a logistic analysis. For details on these studies, see Appendix L, and for the results, see Appendix N. Estimates for the benchmark analytic model and the four alternative models are graphically presented in Figure 5.

Figure 5. Estimated Effect of PFM on Probability of Employment for Benchmark and Alternative Models

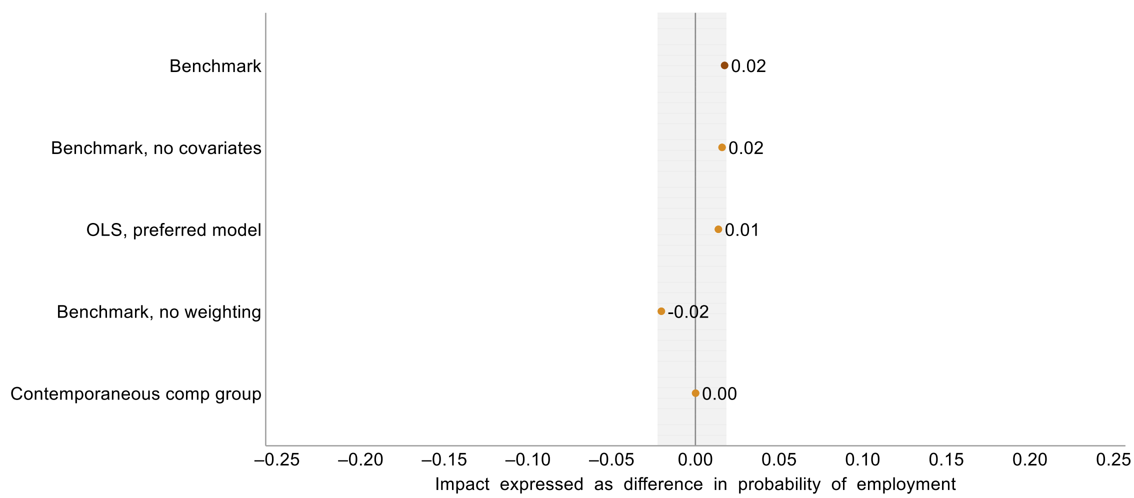


Figure 5 illustrates the relative consistency in treatment effect between each of the five models by plotting the relative predicted effect (expressed as the difference in probability of employment attributable to PFM) for each of the sensitivity studies and the preferred "benchmark" model. The figure also provides a graphical picture of the narrow domain of confidence for the estimated effect of the program. The shaded area delineates the range from the lowest estimate (benchmark model, no weighting = –2.0%) to the highest estimate (benchmark model = +1.7%). Two of the three sensitivity studies produce estimates that are virtually identical to the preferred model.[89] Although we believe that the preferred "benchmark" model provides the single best estimate of the program effect on employment, some of the sensitivity results temper confidence in that estimate. The range of estimates expressed by the dark shading is narrow and provides perhaps a more certain and conservative estimate of where the true effect of the program lies.[90, 91]

[89] There may be concerns about the number of covariates included in the benchmark model, but as shown in Figure 5, the estimates produced by the benchmark model and the benchmark model with no covariates are essentially the same.
[90] An additional layer of uncertainty is imposed by effect size estimates that are produced by the preferred and alternative models. As a result of the size of the analytic sample, we were unable to fit a multilevel logit model. This has the effect of (erroneously) reducing standard errors and increasing hypothesis test statistics (z scores). The sample size is so large that most z-tests are well above the conventional threshold for significance (for conventional samples). Nevertheless, estimates are small, and reported hypothesis tests are likely biased upward.
[91] Due to a USDOL investigation into the validity of data at local boards 14 and 15, we conducted a secondary analysis without those boards. The results of this analysis are consistent with benchmark findings.

RESEARCH QUESTION 2: EFFECT ON QUARTERLY WAGES

Preferred model estimates (reported in Table O.1 in Appendix O) indicate that four quarters after enrollment, the PFM had a marginal negative effect on the wage outcomes of CSF clients included in the study.[92, 93] Transformations of model coefficients suggest that for the typical or average client, turning on the PFM decreases quarterly wages by $139 four quarters after enrolling in LWDB services.[94] Figure 6 illustrates the model-adjusted preenrollment trends (solid lines), post-enrollment projected trends (dashed lines), and predicted average post-enrollment quarterly wages for the treatment and comparison groups.

Figure 6. Pre- to Post-Program Change in Quarterly Wages

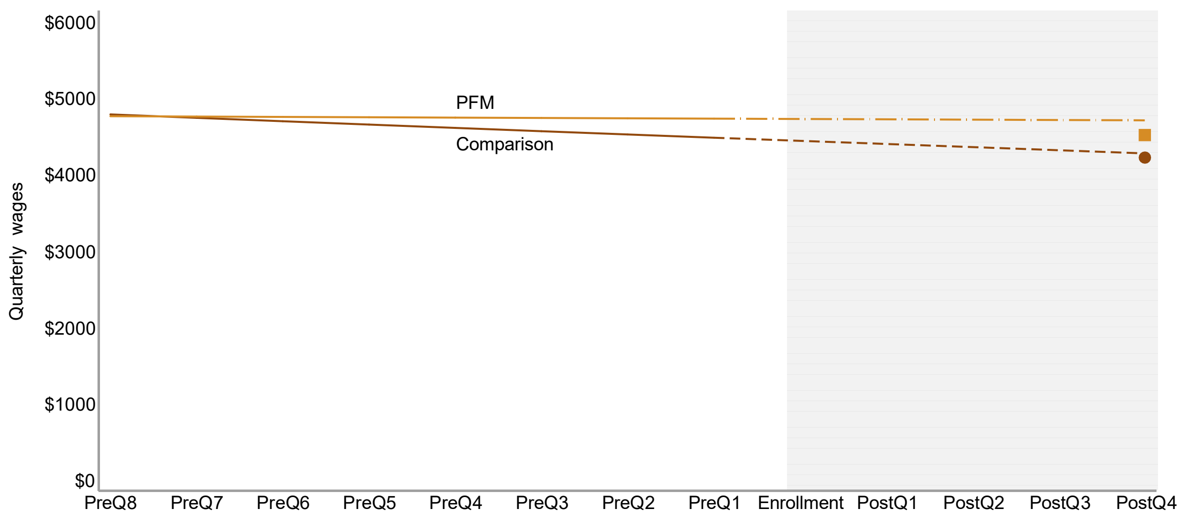


Figure 6 shows that, for the eight quarters prior to enrollment, the solid lines illustrate the model-adjusted average pre-program quarterly wages. The lines indicate that, conditional on the model, both groups demonstrate a gradual decline in quarterly wages across the eight quarters prior to enrollment.[95] For the comparison group, the average adjusted quarterly wages decline from around $4,789 to $4,483. The treatment group, meanwhile, declines from $4,765 to $4,733.[96] In this graphic, the relative declines in quarterly wages are perceptively different. After weighting and other statistical adjustments, the two trends are significantly divergent in slope, with the comparison group declining at a slightly higher rate than the PFM group.[97]

---

[92] The preferred model is a zero-inflated negative binomial (ZINB) single-level model that includes controls for individual-level covariates, time-variant economic and contextual variables, local board-level controls, and an additional contextual variable that controls for the policy switch from WIA to WIOA, a change that took place nationwide during the treatment group study window. For details on the preferred model specification for Research Question 2, see Appendix J.

[93] See footnote 75.

[94] The coefficients of the ZINB model are not easily interpretable on their own because they represent values that are transformed by a two-part model, which includes a logit component that estimates the incidence of inflated zeros separately (with a logit link), plus a negative binomial model that estimates the remaining count values with a log link. We calculate linear probability estimates of quarterly wages from ZINB estimates by transposing the result of the regression equation at fixed values for the CSITS variables and all remaining control variables held at zero. Linear transformation of the coefficients with all control variables held at zero produces estimates of effect for the "typical" or average program participant because we have mean-centered the control variables.

[95] The linear trends are a feature of the analytical model, which assumes a ZINB distribution, and adjustments for covariates and propensity score weights. Figure C.3 in Appendix C illustrates the unadjusted pre-program mean wages for both groups.

[96] Model-based predicted wages are produced in Table O.2 in Appendix O.

[97] Coefficients for the treatment and trend interaction demonstrate this difference in the untransformed log-link scale. See Table O.1 in Appendix O.

At the quarter prior to enrollment (PreQ1), the lines become dashed to illustrate that they are a projection of the baseline trends for the typical program participant in each group. Four quarters post-enrollment in LWDB services, if these trends were to continue, and excluding any difference due to program exposure, the model-based projections suggest that the comparison and treatment groups would be expected to realize quarterly wages of $4,276 and $4,711, respectively. These estimated values represent the hypothetical wage outcomes for each group in the sense that they account for the pre-program differences in the two groups, but they exclude any post-program information. Estimates of post-program wage outcomes are plotted as points below the two lines: $4,222 for comparison and $4,517 for treatment.

The estimated wage outcomes for both the PFM and comparison groups are below the projected lines, as they were with the employment outcomes. This implies that the difference in hypothesized and observed wage outcomes are negative four quarters after enrolling in LWDB services.[98] Again, this initial difference should not be interpreted as nonperformance in service programs, but the result of the employment interruption (unemployment or underemployment) that led that individual to seek out LWDB services.[99]

As with the employment outcomes, the CSITS model estimates program effect as the difference between each of the groups' projected and observed (mean) results. In this case, the difference is represented by the relative difference in distance of the points from the lines. The difference in this case is perceptible in the graphic, as the treatment square is noticeably further from its projected line than the comparison dot. For the "typical" program participant, the difference between the two groups' mean projected and observed quarterly wages is –$139 four quarters after engaging in LWDB services provided by CSF. In other words, conditional on pre-treatment differences in wage trends, covariates, and propensity scores, the PFM results in clients earning $139 less than the comparison group in the fourth quarter after enrollment.[100]

The estimate represents what we consider the most precise and unbiased single estimate of program effect, provided by our preferred model.[101] Estimates are statistically significant, but again, the sample size is very large and statistical significance is a poor gauge of substantive importance.[102] The magnitude of the estimated effect of PFM on wages is conventionally considered small (effect size = 0.03).[103] This translates, with a few assumptions, to approximately $2 less each day, or $11 less each week, or $556 less per year for the average PFM-exposed client.

We are generally confident that the estimates produced by our preferred benchmark model are an accurate reflection of the effect of the PFM. As with Research Question 1, however, the constraints and assumptions of the nonexperimental approach impose some uncertainty.

---

[98] Differences are $4,221.85–4,276.41 = –$54.56 for the comparison group and $4,517.28–4,711.04 = –$193.76 for the treatment group.

[99] And again, this is evidence of why an ITS design would be inappropriate to detect the short-term effects of the program.

[100] Estimated differences in quarterly wages are calculated on the basis of all control variables in the analytic model being held at zero (mean value). Model coefficients remain fixed, but the nonlinear nature of the ZINB model results in predicted dollar values that will vary based on the value of pre-program characteristics and wages that are incorporated into the predictive algorithm.

[101] Model selection and specification procedures for Research Question 2, along with statistics and analysis of those statistics, are presented in Appendix J. The preferred model estimate is reported in log odds-ratio format as the coefficient for the interaction of the treatment and post-enrollment indicator term in (the fifth row) of Table O.1 in Appendix O.

[102] And again, standard errors are underestimated by the single-level model.

[103] Effect size is calculated as Hedges' *g*.

Our benchmark analysis estimates the effect of the program four quarters post-enrollment in LWDB services. This is a constraint in the sense that it restricts evidence of the effect of the PFM to a single point in time. Although there is reason to expect that this time period should be adequate to realize program benefits, we also recognize that these outcomes can vary meaningfully over time.[104]

We again ran an additional analysis using quarterly wages at the eighth quarter after enrollment.[105] The results are consistent with those of our benchmark analysis with regard to the effect – the treatment group has significantly lower quarterly wages than the comparison group in the amount of –$175. However, the visual depiction (see Figure O.1 in Appendix O) shows that, unlike the benchmark results, observed eighth quarter wages for both groups exceed the projected wages, but this gain was greater for the comparison group.[106]

The design and analytic choices that we have made also introduce some uncertainty. This is a feature of all nonexperimental studies. In this case, we have selected as our benchmark approach a natural experiment with a non-contemporaneous comparison group that could confound differences in labor market and policy conditions with the treatment effect. The design cannot control for these differences, and we rely on our statistical model to adjust for these differences (weighting, covariate, and trend adjustments). Although we believe that this strategy is as defensible as those used to control for selection bias in a nonequivalent, contemporaneous comparison group approach, we concede that we do not know if we have effectively controlled for these differences.

We test this assertion empirically by conducting a sensitivity analysis that contrasts wage outcomes for the treatment group with a contemporaneous but nonequivalent comparison group. We use the same analytic procedures to compare wage outcomes for the treatment group with those for a subgroup of the comparison group that did not receive LWBD services at the time when the PFM was active (but did receive those services in the comparison period). We observe wage outcomes contemporaneously for both groups. Estimates from the sensitivity analysis suggest that the effect of the PFM relative to the contemporaneous comparison group is positive but statistically insignificant.

The estimates suggest that, contrary to what we saw in the employment outcome, our benchmark model may be underestimating the effect of the PFM. However, focusing on the difference between the two estimates probably misses the point. The difference between the two estimates is again very small (around $220 or 0.04 of a standard deviation), and the sensitivity estimates are statistically indistinguishable from zero, even in this exceptionally large sample. Further, they are distinct analytical samples with different counterfactuals. Some variation is to be expected. The more astute interpretation is probably one that accentuates the similarity of the two rather than the difference. Taken together, the two estimates suggest that the PFM is not having a meaningful effect on wages. We can be confident, in any case, that it is not having a meaningful positive effect on wages.

---

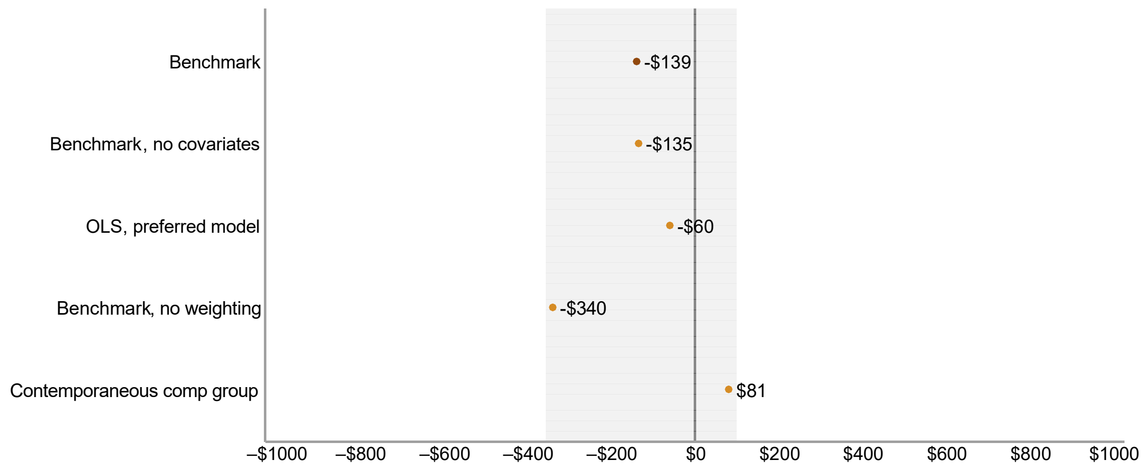
[104] For examples, see Heinrich, C. et al. (November 2009). *New estimates of public employment and training program net impacts: A nonexperimental evaluation of the Workforce Investment Act Program*, IZA Discussion Paper No. 4569. Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *Economic Journal*, *120*(548), F452–F477. Andersson, F., et al. (September 2013). *Does federally-funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms*, IZA Discussion Paper No. 7621.

[105] As we did not have eighth quarter outcome data for the final two quarters of enrollees for the treatment group; and in order to keep the balance between the two groups, we also dropped the final two quarters of enrollees from the comparison group.

[106] Projected wages for the comparison group were $3,898; observed wages were $4,419. Projected wages for the treatment group were $4,508; observed wages were $4,854.

As with the first research question, we conduct a set of sensitivity tests of the robustness of the preferred statistical model. We estimate PFM effect with variations of the preferred model – without propensity score weights included, without covariates included, and fitting the model with OLS regression instead of a ZINB analysis. Estimates for the benchmark analytic model and the four alternative models are graphically presented in Figure 7.

Figure 7. Estimated Effect of PFM on Quarterly Wages for Benchmark and Alternative Models



Estimates plotted in Figure 7 paint a largely consistent picture of a negligible PFM effect. Variations in modeling choices, comparison group, and design confound (time or selection) all produce estimates within $200 of the benchmark estimate.[107] Predicted values of effect on quarterly wages range from –$340 (no weighting) to +$81 (contemporaneous comparison group). We believe that the preferred "benchmark" model provides the single best estimate of the program effect on wages. This time it sits in the middle of the alternative estimates, which would seem to support that confidence. In any case, the range of estimates expressed by the gray shading is narrow and provides a more conservative bounding limit of where the true effect of the program lies.[108, 109]

RESEARCH QUESTION 3: EFFECT ON TIME TO EMPLOYMENT

Preferred model estimates (reported in Table P.1 in Appendix P) indicate that CSF clients who were unemployed at the time of enrollment and exposed to the PFM became employed incrementally sooner than similar unemployed participants in the comparison group.[110] The treatment effect is very small, and

---

[107] There may be concerns about the number of covariates included in the benchmark model, but as shown in Figure 7, the estimates produced by the benchmark model and the benchmark model with no covariates are essentially the same.

[108] An additional layer of uncertainty is imposed by effect size estimates that are produced by the preferred and alternative models. As a result of the size of the analytic sample, we were unable to fit a multilevel logit model. As we elaborate in Appendix H, this has the effect of (erroneously) reducing standard errors and increasing hypothesis test statistics (z scores). The sample size is so large that most z-tests are well above the conventional threshold for significance (for conventional samples). Nevertheless, estimates are small, and reported hypothesis tests are likely biased upward.

[109] See footnote 91.

[110] The preferred model is a discrete-time hazard model (fit with a logit regression) with dichotomous indicator variables for each discrete time period (quarter), a set of time-variant economic indicators (to improve identification of the treatment effect), a set of individual-level covariates, and local board-level blocking variables. Only individuals who had zero quarterly wages reported in their quarter of enrollment are included in the analytic sample. Details on the specification of this model are provided in Appendix K. The discrete-time hazard model is estimated with a logistic

results are statistically significant. However, the sample size is so large that conventional levels of statistical significance do not necessarily denote that the difference is meaningful.[111]

One way of illustrating the conditional estimates produced by the discrete-time hazard model is to calculate the predicted conditional probabilities of becoming employed over each quarter of the study period. In Figure 8, we present such a graphic by charting the estimated probability of becoming employed for 11 quarters following the enrollment quarter, for both the PFM and comparison groups. These conditional probabilities have been produced by our preferred model.[112] The line labeled PFM illustrates the quarter-to-quarter likelihood of the treatment group becoming employed, whereas the Comparison line illustrates the same for the comparison group. The vertical distance between the two lines is a visual representation of the effect of the PFM during each quarter post-enrollment.

Figure 8. Estimated Conditional Probability of Becoming Employed for PFM and Comparison Groups

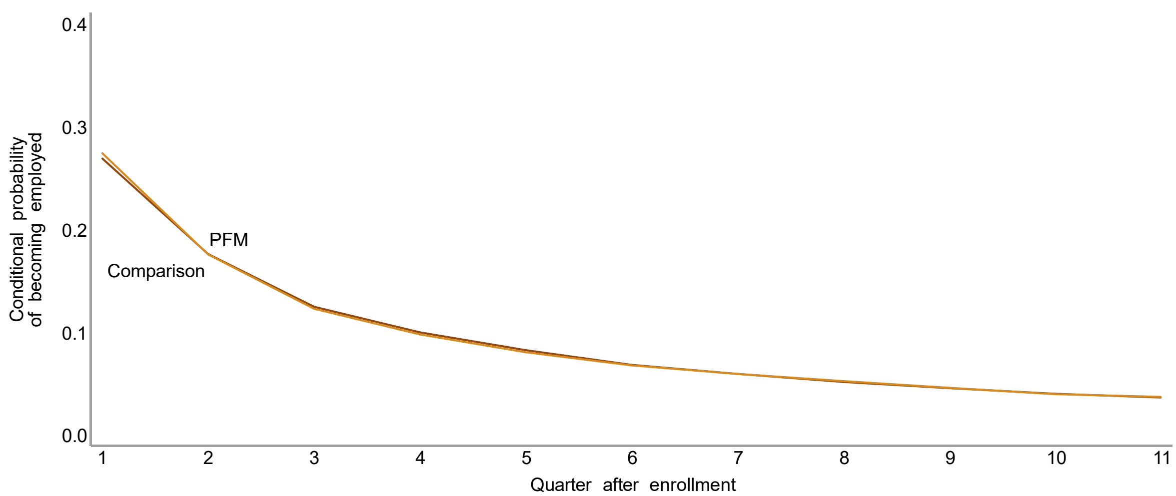


Figure 8 shows that the conditional probability of becoming employed is nominally higher for the treatment group in the first quarter and even that difference diminishes thereafter. In the first quarter post-enrollment, the PFM group has a 0.5% higher probability of becoming employed than the comparison group. After that, both groups converge. Table P.2 in Appendix P reports the conditional probability estimates for both treatment and comparison groups; by the second quarter, the two groups are becoming employed at virtually identical rates.

Figure 9 uses the same data to illustrate the cumulative effect of the PFM on time to employment. The graphic plots the proportion of PFM and comparison groups who remain unemployed at each quarter.

regression. Coefficients for the time indicators in this model estimate the conditional probability (as log odds ratios) of gaining employment for any given quarter and the treatment effect estimated is the conditional effect of PFM on that probability of employment (again expressed as a log odds ratio). The coefficient for the treatment effect is constant in log odds terms; its magnitude varies with magnitude of baseline risk (risk at time period t) when converted to probability. Probability estimates are calculated as the mean predicted probabilities (predict in Stata) for each group (i.e., PFM or comparison) at each discrete time point.

[111] The number of unique enrollments included in the analytic sample for Research Question 3 is 1,214,269.

[112] See footnote 110.

Figure 9: *Proportion of Individuals Who Remain Unemployed for PFM and Comparison Groups*

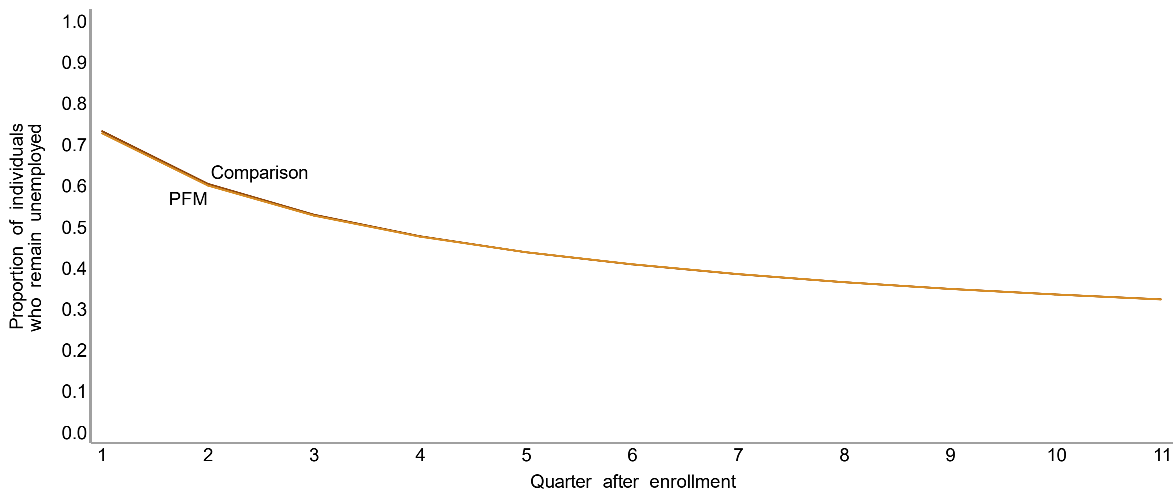


Figure 9 illustrates that though the model estimates show a positive and significant effect, the relative benefit of PFM is practically negligible. The lines represent the (model adjusted) proportion of each group that remains unemployed over time; any distance between the lines illustrates a difference in employment outcomes. In this case, the distance between PFM and comparison lines is imperceptible across 11 quarters. According to estimates presented in Table P.2 in Appendix P, at the first quarter post-enrollment, both groups remain unemployed at functionally similar rates (73%). By the 11th quarter, both the PFM and comparison programs have substantially reduced unemployment. But, the difference between the two groups has not changed at all; 32% of each group remain unemployed.

According to the findings, then, there is no meaningful difference in the time to employment for the PFM group. Differences in outcomes between the two programs is slight and fleeting. Nevertheless, under the PFM, proportionately more people find employment, and they do so sooner.[113]

As with the previous two research questions, estimates have been produced with nonexperimental methods. The validity of these estimates relies in part on assumptions – that may or may not hold – and statistical analyses – that may or may not adequately control for differences between the two groups.

LIMITATIONS

Given the constraints placed on the study, we believe we have produced the most rigorous nonexperimental analysis possible. Unweighted balance statistics reported in Tables C.1 and C.2 indicate that the treatment and comparison groups were broadly similar in baseline characteristics, and weighting procedures made the two groups even more aligned. Nevertheless, any nonexperimental design may reduce the ability to isolate the effects of programming from potential sources of bias.

[113] For the PFM group, the median time to employment (or the time at which half of the group achieves employment) is 3.55 quarters; whereas, the comparison group achieves that milestone at 3.79 quarters. In other words, the point at which half of the PFM group becomes employed occurs three weeks before the comparison group. These values are calculated using unadjusted data. Table P.3 in Appendix P provides additional descriptive statistics on the contrasted groups.

NEDs rely on observed features to mitigate the possibility that estimated treatment effects are conflated with other influences. NEDs cannot control for unobserved influences, and there is no way to be certain that the effect credited to the program is not also motivated by these unobserved factors.

There remains the question of whether our analytic model has adequately controlled for the differing labor market conditions that result from the off-year comparison group and produced an unbiased estimate of the effect of PFM. Any adjustments would likely be downward. The data show that – all other things being equal – wages and wage gains appear to be greater for those in the treatment group. Any objective improvements to the identification of the treatment effect (e.g., added conditioning variables or improved modeling) would seemingly only reduce the treatment effect further. Model selection procedures have evidenced this so far: estimates diminished with added statistical controls. It is also possible, but not likely, that the model has overadjusted for differences in the labor markets experienced by both groups. This would have to be a weakness in the CSITS design itself, for even the most parsimonious model produces estimates that are at or around zero.

An alternative hypothesis is that limitations in the data may have led to a downward biased estimate of the effect of PFM on labor market outcomes. The UI wage data that we used to measure earnings and employment status do not account for activity in the so-called gig economy. This issue was discussed at length by the evaluation team but because no reliable data exist, was not explicitly controlled for in our analytic models – other than the indirect influence that it would have on the available labor market metrics. Therefore, it is possible, given that the labor market for gig-economy wages had a greater importance in the treatment period (e.g., comparatively more treatment group members may have been gainfully employed partly or completely by jobs such as Uber driving) and these wages would have been unmeasured by the UI wage data, that employment and wage effects could be greater than what we find with our benchmark analysis. Sensitivity results, however, refute this hypothesis somewhat – especially those produced by the nonequivalent contemporaneous control group.[114]

DISCUSSION

According to the outcomes study, the PFM produced mixed results: a marginal positive effect on client employment outcomes, marginal negative effect on wage outcomes, and negligible but positive effects on employment for CSF clients who were unemployed at the time of enrollment. These modest and mixed results are not surprising given the complexity of the intervention being implemented, the systems-level nature of the intervention, turnover in key staff, and the challenges in implementation that resulted in the delay of some key programmatic components, such as the web portal that allowed LWDB leadership to see their progress on the metrics.

First, client employment and earnings are distal outcomes of the PFM. Several levers have to move in the hypothesized direction before the client would actually realize the hypothesized benefits – in an effort to meet performance targets, local board leadership needs to change policies and programming, management and frontline staff need to change their behavior based on these changes, and then LWDB clients need to experience these changes. A systems-level intervention, although having the potential to

[114] It is also possible that our statistical model has not adequately controlled for the implementation of WIOA and its associated requirements. To test this empirically, we conducted a sensitivity analysis that compared labor market outcomes for PFM participants exposed to WIA only (i.e., prior to the implementation of WIOA) with those exposed to WIA only (i.e., prior to PFM). Selection into treatment and comparison groups is identical to benchmark analysis except it is restricted to the last one-year period in the comparison enrollment period (July 2013 through June 2014) and the first one-year period in the treatment enrollment period (July 2015 through June 2016). Results do not differ substantively from our benchmark findings.

impact many more individuals and become a permanent change, is more complicated to implement and offers a less direct line to client impact than a direct-services program, such as an innovative job training program.

Second, as noted, the PFM team experienced significant turnover in staff over the project period, including the original architect of the PFM in the first year of the grant and the original project manager in the second year. Left with a very complex project that was only partially implemented, CSF and DEO assembled and restructured a team to forge forward with implementation. At the conclusion of the grant, LWDB executive directors praised the communication and technical assistance provided to them in the second half of the grant. Nearly all highlighted the critical change that occurred with the transition to the new team.

Third, there was a delay in rollout for the web portal, which is a key feedback mechanism for the LWDB leadership to see how they are performing on the PFM metrics. At the start of the project, CSF laid out three objectives, one of which was the creation "of a comprehensive, easy-to-understand, web-based data portal to provide local workforce development boards with the data necessary to inform their decision-making processes." Without this functionality the PFM was not operating as intended.

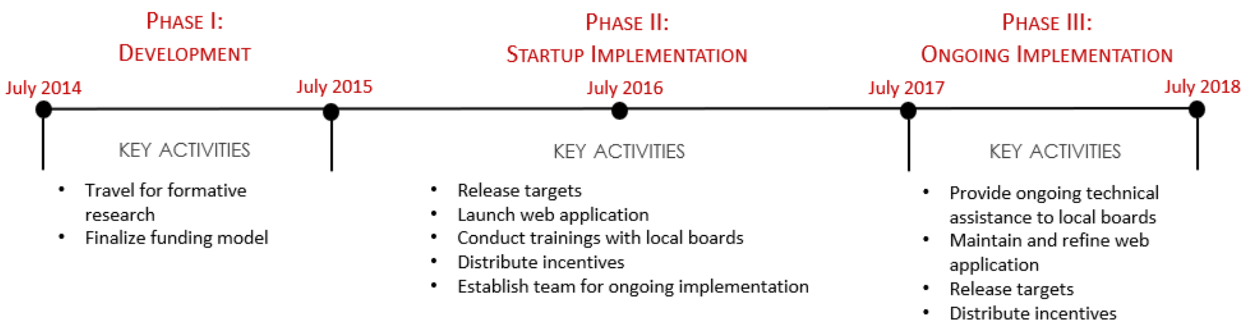
Additionally, the evaluation period may have been too short to fully assess the full set of effects on labor market outcomes. Reporting lags for receiving certified wage data and the due date for the final report both effectively truncated the evaluators' ability to assess outcomes beyond four quarters. As described in detail in this report, the treatment group enrollment period ended as PFM was still in mid-implementation. A longer treatment period and a longer follow-up period may result in more clear-cut outcomes.

Despite all of this, the findings indicate that there is evidence of promise. After overcoming the initial implementation challenges, the PFM team and LWBD leadership have worked to make the relationship more productive and collaborative. Final interviews with the PFM team and local board leadership suggest a willingness to engage with the PFM (or a new iteration of the PFM) going forward. Whatever form the revised model takes, CSF and local board executive directors learned from implementation and assert the value of significant planning, supportive partnerships, and continuous improvement in any future efforts.

APPENDIX A. COST STUDY

The study calculates costs for three distinct phases that occurred during implementation of the Performance Funding Model (PFM): development, startup implementation, and ongoing implementation. The delineation of these three phases has been informed by qualitative data gathered for the PFM Implementation Evaluation. Each phase is summarized in Figure A.1.

Figure A.1. PFM Implementation Phases



Phase I represents development efforts, including finalization of the PFM and travel for formative research of each local workforce development board (LWDB). Phase II represents startup implementation efforts, including refinement of technical assistance and training opportunities that were offered to LWDBs and the incremental transition to a consistent PFM team staffing structure. Phase III represents a steady state of ongoing implementation; in other words, anticipated annual costs if the model were to remain intact and stable.

METHODS

In August 2018, The Policy & Research Group (PRG) provided CareerSource Florida (CSF) and the Florida Department of Economic Opportunity (DEO) with data collection guidance for cost study tables to be populated for each quarter of the grant period. All preliminary cost study data were submitted by CSF to PRG in April 2019.

PHASE CALCULATIONS

The three implementation phases are defined as followed: Phase I: Development (July 1, 2014 through June 30, 2015); Phase II: Startup Implementation (July 1, 2015 through June 30, 2017); and Phase III: Ongoing Implementation (July 1, 2017 through June 30, 2018).

Total Costs per Phase *= Total Staff Costs + Materials + Travel + Consultative Services*

STAFF COSTS CALCULATION

Within each phase, the total staff costs will consist of the sum of each staff person's quarterly costs. For example, total Phase I staff costs = (staff person 1 Q1 + Q2 + Q3 + Q4) + (staff person 2 Q1 + Q2 + Q3 + Q4), etc., where staff person 1 Q1 = Hours Dedicated to Operating the PFM * Fully Loaded Hourly Rate.

The purpose of calculating staff costs within each phase is to account for peaks and valleys in staff effort based on specific project activities, as well as the incremental addition of new PFM team members.

Throughout the grant period, a total of ten unique CSF positions and five unique DEO positions dedicated time to the PFM. Table A.1 defines each component of the staff costs calculation.

Table A.1. Staff Costs Calculation Variables

Variable	Costs Required	Calculation
PFM team member Fully Loaded Hourly Rate	Base Labor Hourly Rate = annual salary/2,080 hours (total number of work hours in year) Fringe Rate = percentage that encompasses the employee benefits such as health, life, and disability insurance; retirement benefits; worker compensation insurance; unemployment insurance; employer's portion of the Social Security and Medicare taxes Indirect Rate = percentage used in federal grants to account for indirect costs such as rent and utilities, general and administrative expenses (accounting, human resources)	Fully Loaded Hourly Rate = (Base Hourly Rate) + (Base Hourly Rate * Fringe Rate) + ((Base Hourly Rate * Fringe Rate) * Indirect Rate)
Hours Dedicated to operating the PFM		Hours Dedicated to operating the PFM = percentage of full-time position dedicated to the PFM in a given quarter * 520 (total number of work hours in a quarter – 2,080/4)

NON-PERSONNEL COSTS CALCULATION

According to data provided by CSF and DEO, non-personnel costs of the PFM included the following: office space, furniture, equipment, office supplies, travel, and consultative services. These costs were totaled for both CSF and DEO within each phase.

RESULTS

Staff and non-personnel costs, as well as the total for each phase, are provided in Table A.2.

Table A.2. Total Costs

Phase	Variable		Cost
Phase I: Development	Staff costs		$328,661.73
	Non-personnel costs		$102,969.18
		Total Phase I:	**$431,630.91**
Phase II: Startup Implementation	Staff costs		$1,148,399.43
	Non-personnel costs		$173,012.08
		Total Phase II:	**$1,321,411.51**
Phase III: Ongoing Implementation	Staff costs		$364,861.20
	Non-personnel costs		$431,726.36
		Total Phase III:	**$796,587.56**

Based on these calculations, we find the total cost for one year of development to be $431,630.91; for two years of startup implementation to be $1,321,411.51; and for annual ongoing implementation to be $796,587.56.

Costs that are not included in our calculations are the actual cost of performance awards and costs provided by existing state infrastructure. Each of these costs is presented, by phase, in Table A.3.

Table A.3. Costs Omitted From Study Calculations

Phase	Variable	Cost
Phase I: Development	In-kind contribution:	
	customer relationship management software	$56,111.00
	Total Phase I:	**$56,111.00**
Phase II: Startup Implementation	Performance awards	$16,444,118.00
	In-kind contributions:	
	customer relationship management software	$694,078.00
	consulting & technical assistance	$92,000.00
	Total Phase II:	**$17,230,196.00**
Phase III: Ongoing Implementation	Performance awards (anticipated)	$2,205,882.00
	In-kind contribution:	
	customer relationship management software	$360,838.00
	Total Phase III:	**$2,566,720.00**

Costs that are specific to the grant (as opposed to the intervention) are also omitted from our calculations. These include grant administration, evaluation, and technical assistance costs.

Finally, it is important to note that these calculations are specific to the PFM as implemented in Florida and as part of a WIF grant cycle. Findings are limited in this respect and may not be representative of the cost of a similar initiative that is implemented in a different setting or that is supported by an alternate funding structure.

APPENDIX B. STATEWIDE COMPARATIVE LABOR MARKET CONDITIONS

The line graphs in this appendix show the trends in labor market conditions over the complete study window for each condition. The comparison study period reflected in the figures is from July 2010 through June 2015, while the treatment study period is July 2013 through June 2018. The box plots depict the distribution of these values over the same time period. In the box plots, the white line inside the solid box represents the median, the bottom of the box represents the first quartile, the top of the box represents the third quartile, and the top and bottom lines (or whiskers) represent the minimum and maximum values.

Figure B.1. Trends in the Number of Individuals in the Labor Force

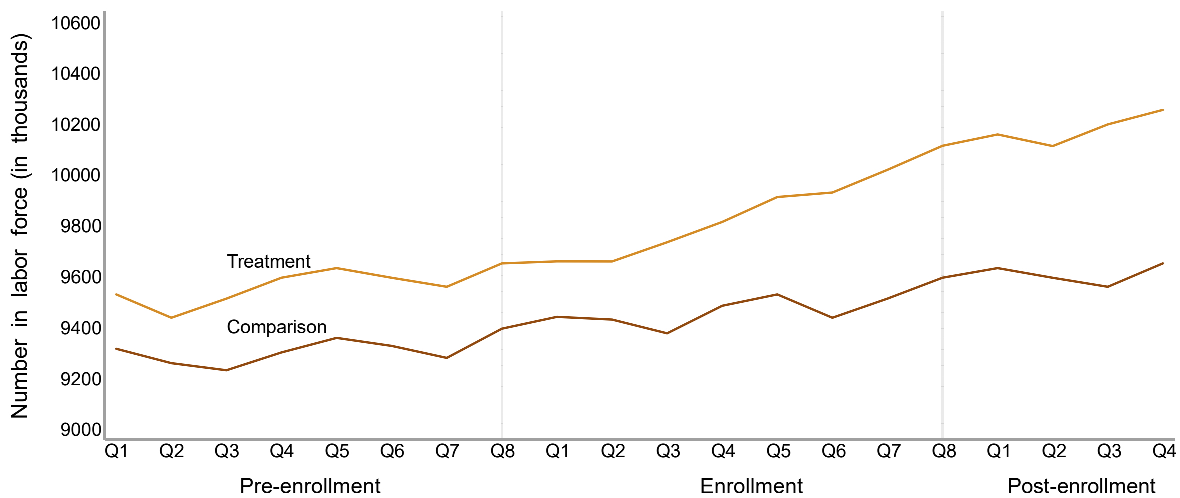


Figure B.2. Distribution of the Number of Individuals in the Labor Force

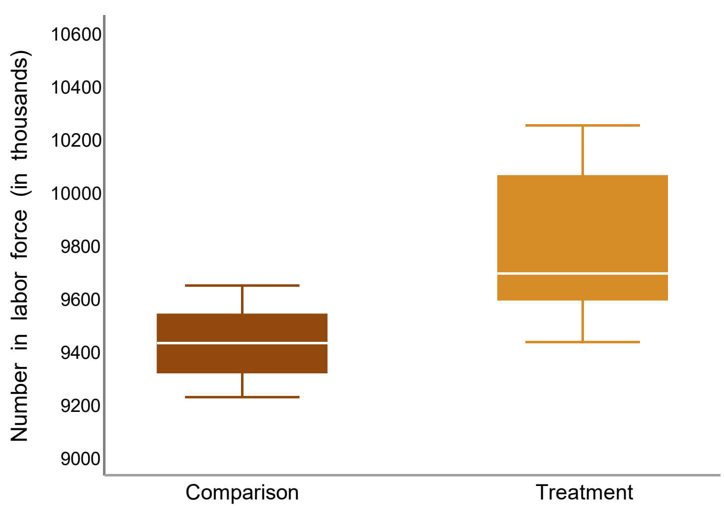


Figure B.3. *Trends in the Number of Employed Individuals*

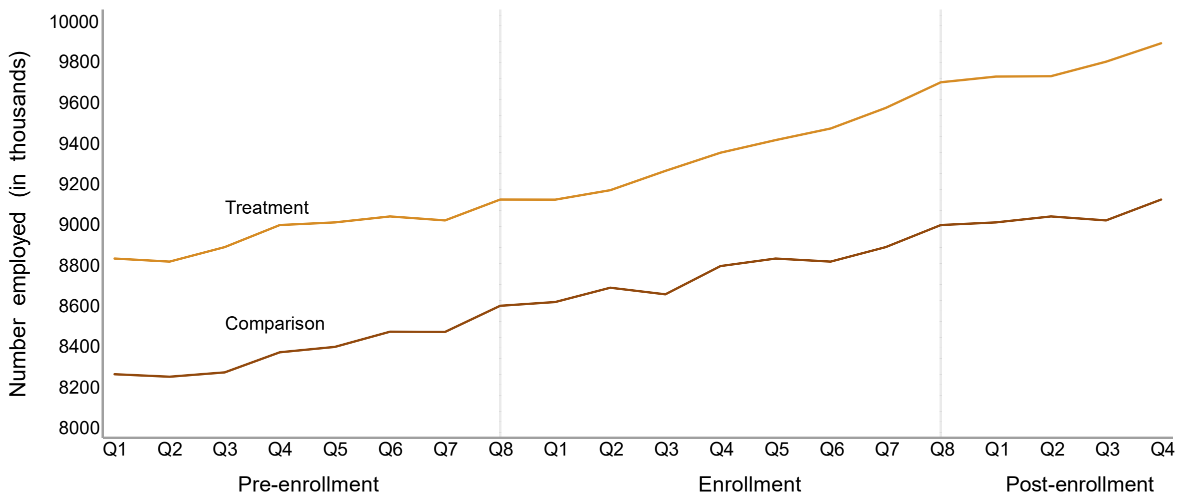


Figure B.4. *Distribution of the Number of Employed Individuals*

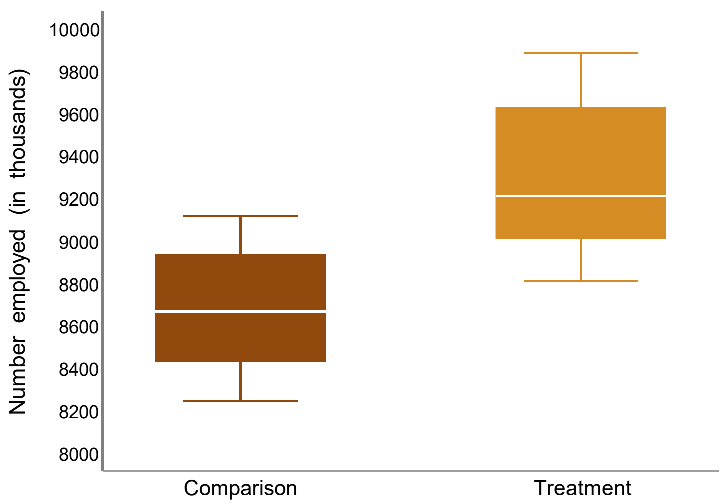


Figure B.5. Trends in the Number of Unemployed Individuals

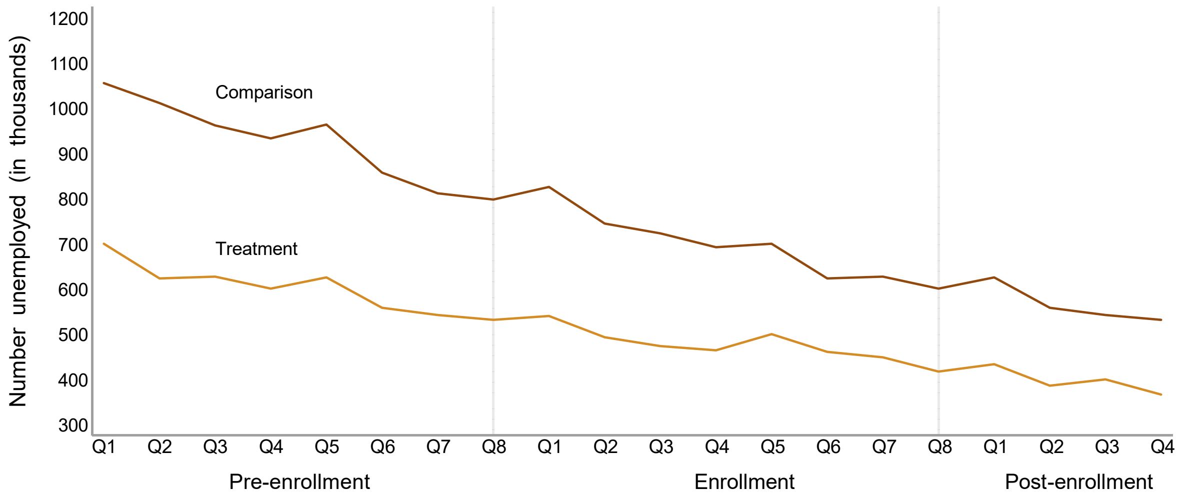


Figure B.6. Distribution of the Number of Unemployed Individuals

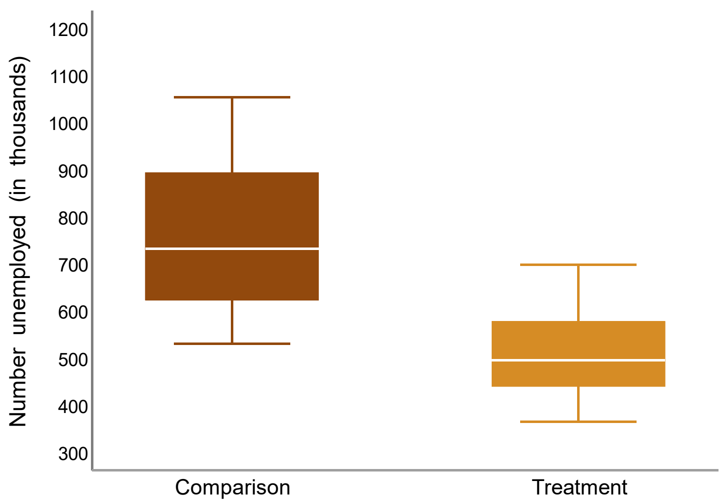


Figure B.7. Trends in the Unemployment Rate

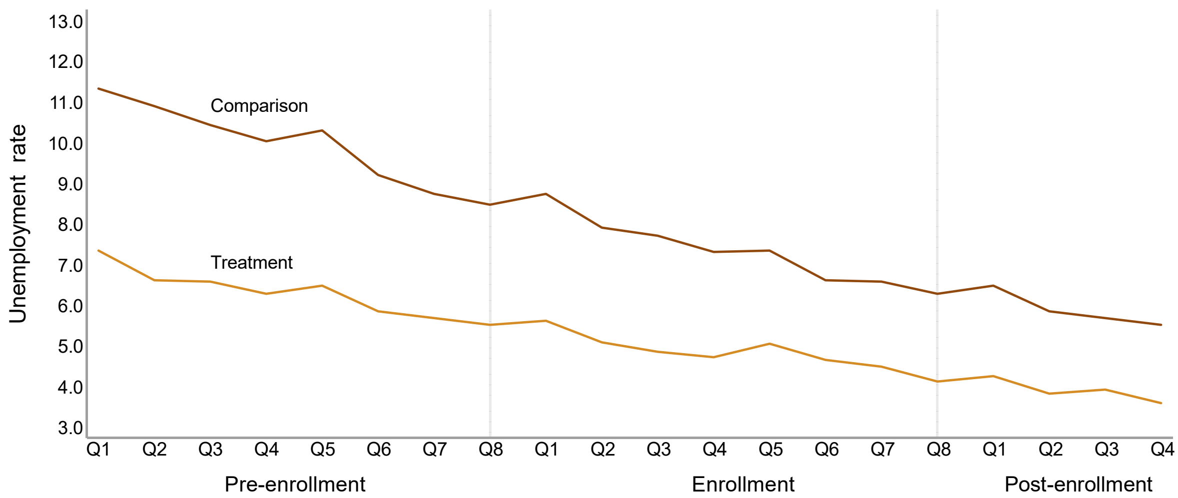


Figure B.8. Distribution of the Unemployment Rate

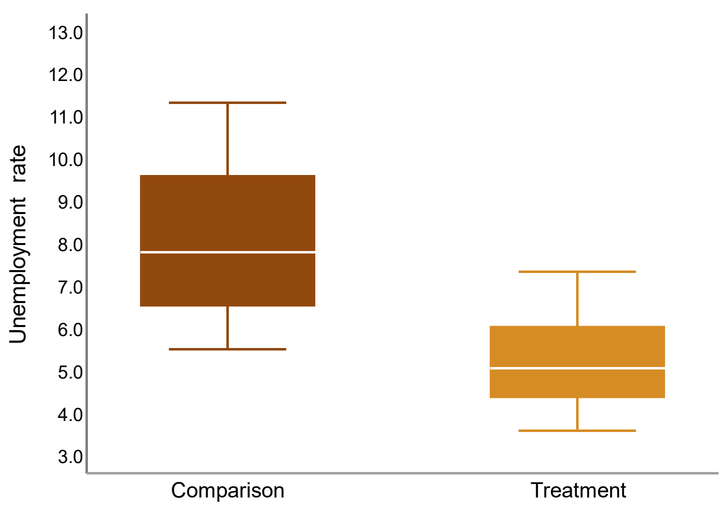


Figure B.9. Trends in the Florida Minimum Wage

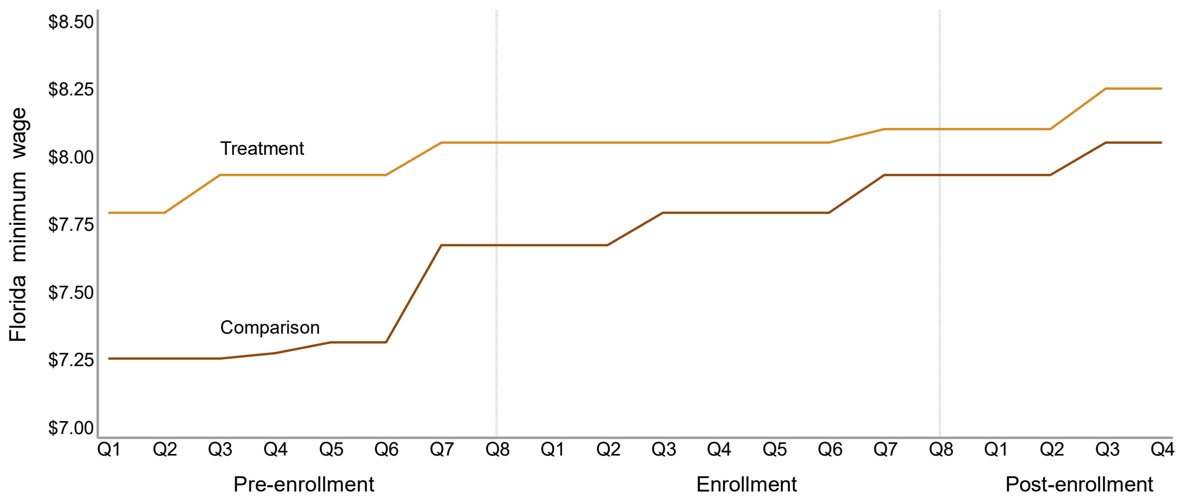


Figure B.10. Distribution of the Florida Minimum Wage

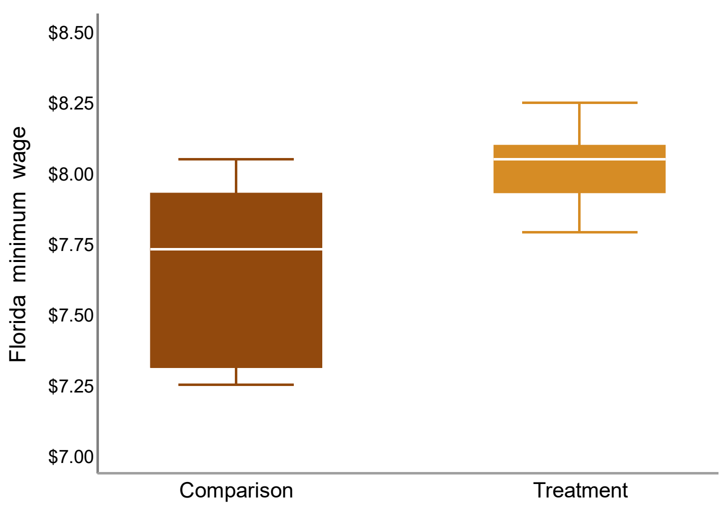


Figure B.11. Trends in Average Weekly Hours Worked

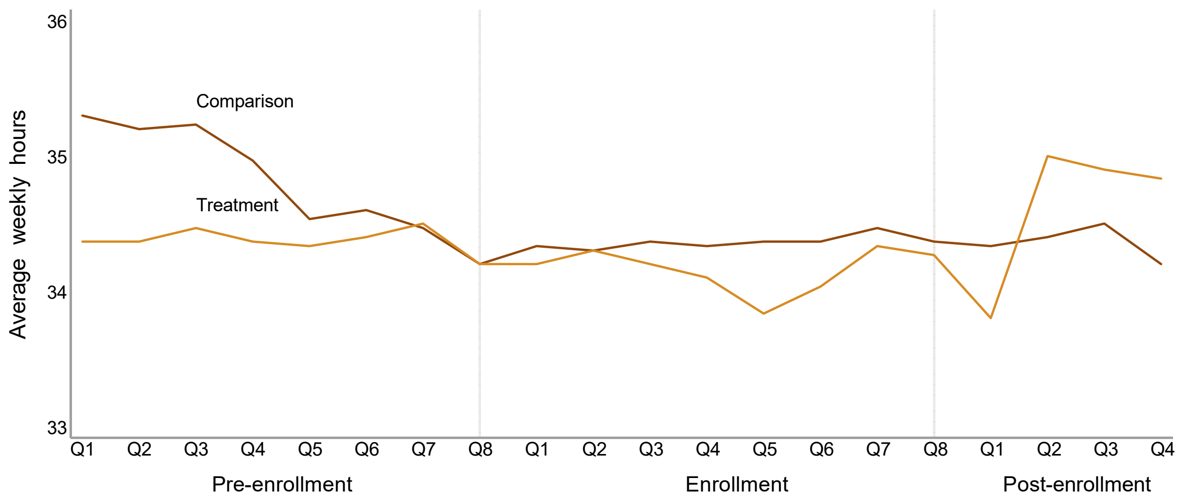


Figure B.12. Distribution of Average Weekly Hours Worked

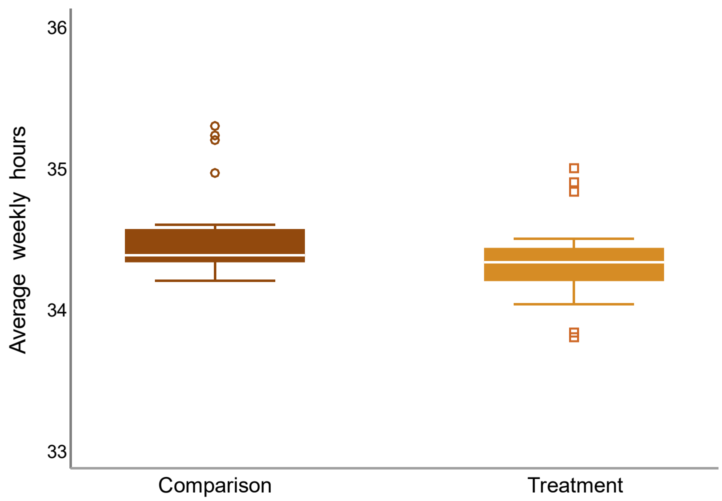


Figure B.13. Trends in Average Hourly Earnings

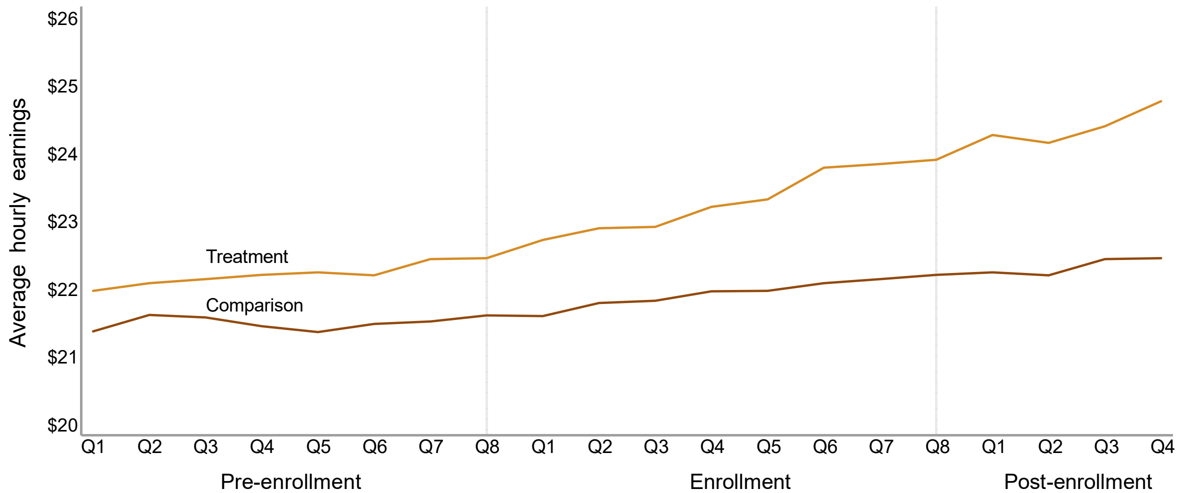
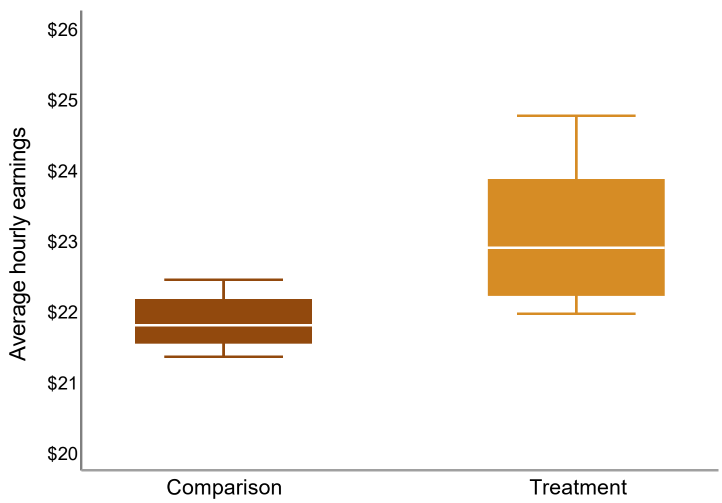


Figure B.14. Distribution of Average Hourly Earnings



APPENDIX C. EQUIVALENCE BEFORE PROPENSITY SCORE WEIGHTING

RESEARCH QUESTIONS 1 AND 2

Table C.1. Unweighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2

Variable	Treatment Mean (*n* = 994,578)	Comparison Mean (*n* = 1,328,761)	Standardized Mean Difference
Continuous Variables			
Average weekly wages 1st quarter prior to enrollment	$876.81	$804.22	0.68
Number in labor force 1st quarter prior to enrollment	707,268	644,938	0.14
Number employed 1st quarter prior to enrollment	671,659	596,407	0.18
Days from beginning of quarter to enrollment date	44.7	43.4	0.05
Days from beginning of study window to enrollment date	360.1	350.4	0.05
Age at enrollment	39.3	39.3	0.00
Number of cumulative enrollments in workforce services	2.1	2.1	0.02
Average quarterly earnings for eight quarters prior to enrollment	$3,934.59	$3,969.82	–0.01
Quarterly earnings prior to enrollment:			
1st quarter prior	$4,184.55	$3,846.24	0.05
2nd quarter prior	$4,227.99	$3,993.99	0.04
3rd quarter prior	$4,129.06	$4,020.21	0.02
4th quarter prior	$3,983.54	$4,046.31	–0.01
5th quarter prior	$3,920.44	$4,069.22	–0.02
6th quarter prior	$3,813.83	$3,984.11	–0.03
7th quarter prior	$3,665.19	$3,918.54	–0.04
8th quarter prior	$3,552.12	$3,879.94	–0.05
Dichotomous Variables			
Attained a high school diploma	91.3%	91.1%	0.01
Gender (male)	46.3%	48.1%	–0.04
Disability status	5.1%	4.8%	0.04
Veteran status	5.4%	6.9%	–0.16
Race/ethnicity:			
Hispanic/Latino	27.6%	24.4%	0.10
Haitian	3.0%	1.2%	0.56
American Indian/Alaska Native	1.3%	1.3%	–0.04
Asian	1.4%	1.4%	–0.01
Black/African American	30.2%	29.3%	0.03
White	52.1%	56.0%	–0.09
Native Hawaiian/Other Pacific Islander	0.5%	0.6%	–0.15
Other race	0.0%	0.0%	0.40

Table C.1. Unweighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2

Variable	Treatment Mean (*n* = 994,578)	Comparison Mean (*n* = 1,328,761)	Standardized Mean Difference
Dichotomous Variables (continued)			
Local board:			
1	3.2%	3.1%	0.02
2	1.2%	1.5%	–0.14
3	1.0%	0.9%	0.03
4	1.9%	2.6%	–0.17
5	1.9%	2.1%	–0.06
6	1.2%	1.3%	–0.01
7	0.9%	0.7%	0.17
8	3.7%	5.8%	–0.28
9	1.8%	1.9%	–0.01
10	2.5%	2.9%	–0.09
11	4.1%	3.9%	0.02
12	8.3%	10.5%	–0.16
13	4.0%	3.9%	0.01
14	6.2%	5.2%	0.10
15	9.3%	9.5%	–0.01
16	3.1%	3.1%	0.00
17	3.1%	3.9%	–0.14
18	3.4%	4.3%	–0.14
19	1.1%	0.9%	0.09
20	2.1%	3.2%	–0.28
21	7.0%	5.0%	0.22
22	5.1%	6.2%	–0.12
23	19.9%	13.9%	0.26
24	4.0%	3.8%	0.05
Enrolled during quarter 1	26.7%	27.8%	–0.03
Enrolled during quarter 2	24.0%	22.6%	0.05
Enrolled during quarter 3	24.8%	24.0%	0.03
Enrolled during quarter 4	24.5%	25.6%	–0.04
Wagner-Peyser participant	95.1%	96.2%	–0.16
WIA/WIOA participant	4.9%	3.8%	0.16
Employed at enrollment	16.6%	16.7%	0.00
Employed with termination notice or military separation at enrollment	0.7%	1.1%	–0.25
SNAP recipient	8.2%	1.2%	1.21
Welfare Transition Program participant	4.5%	1.8%	0.57
Reemployment Assistance claim paid	16.3%	22.8%	–0.25

Figure C.1. Unweighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2

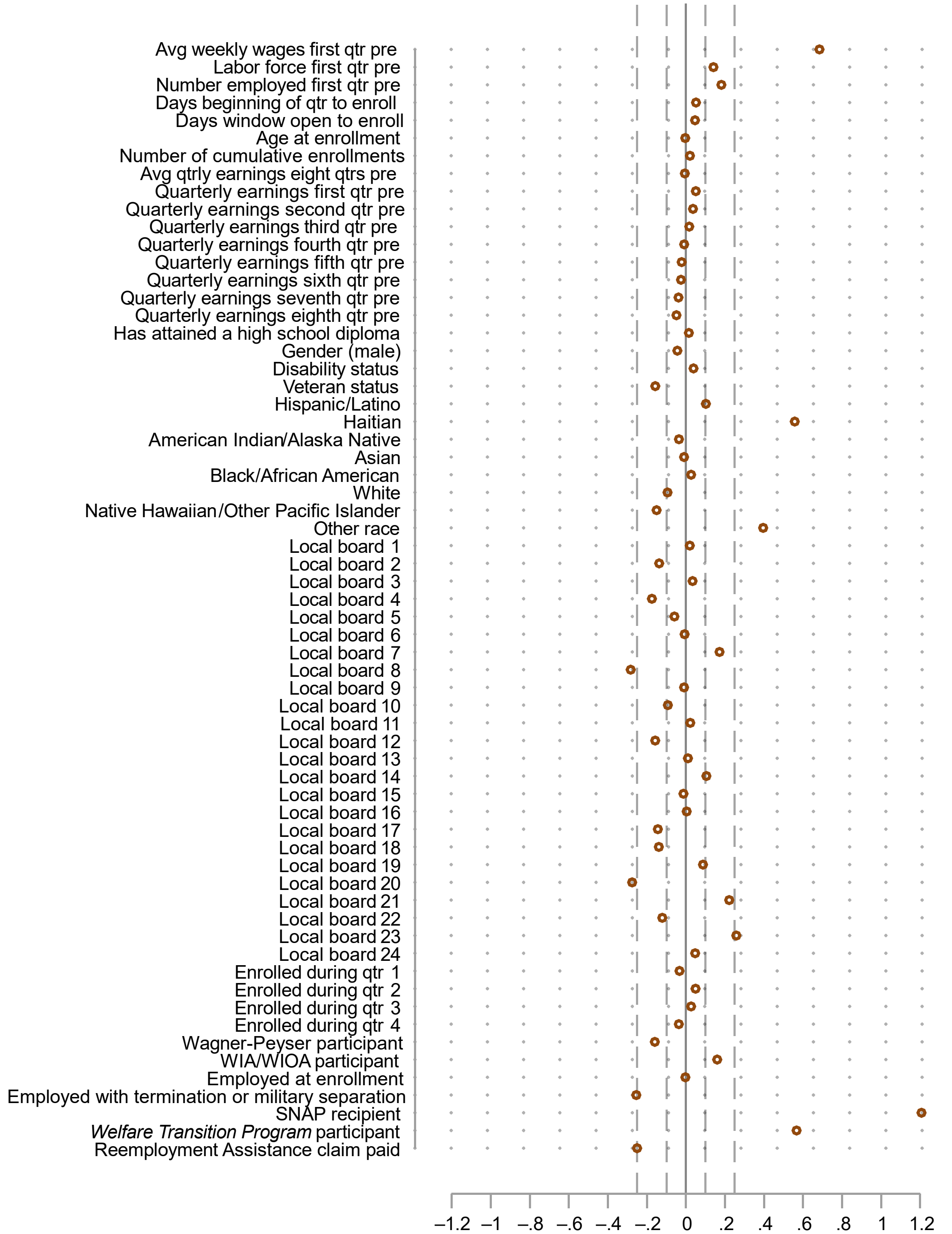


Figure C.2. Pre-Program Employment Trend, Unweighted, RQ 1

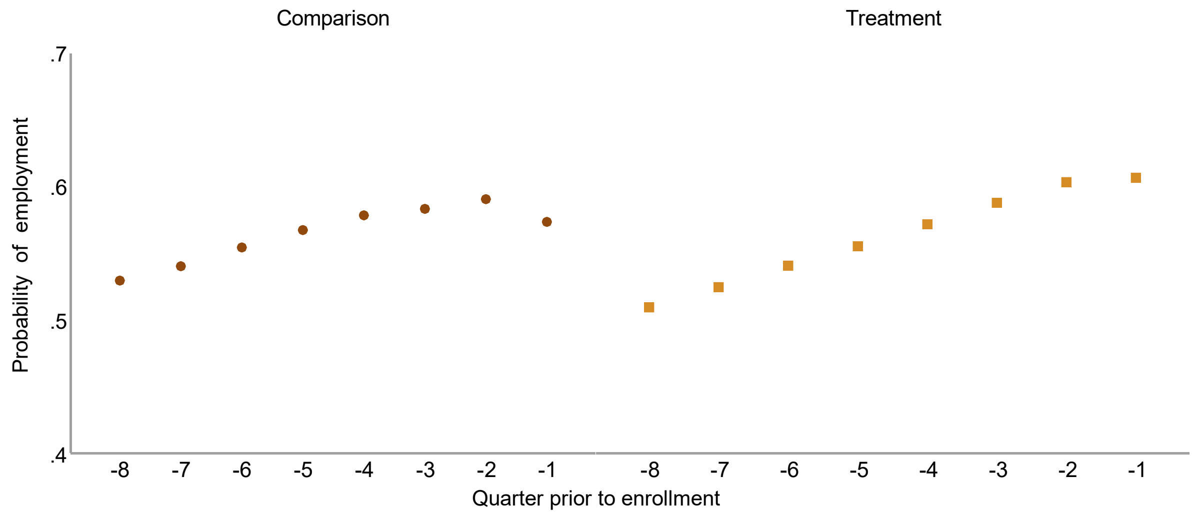
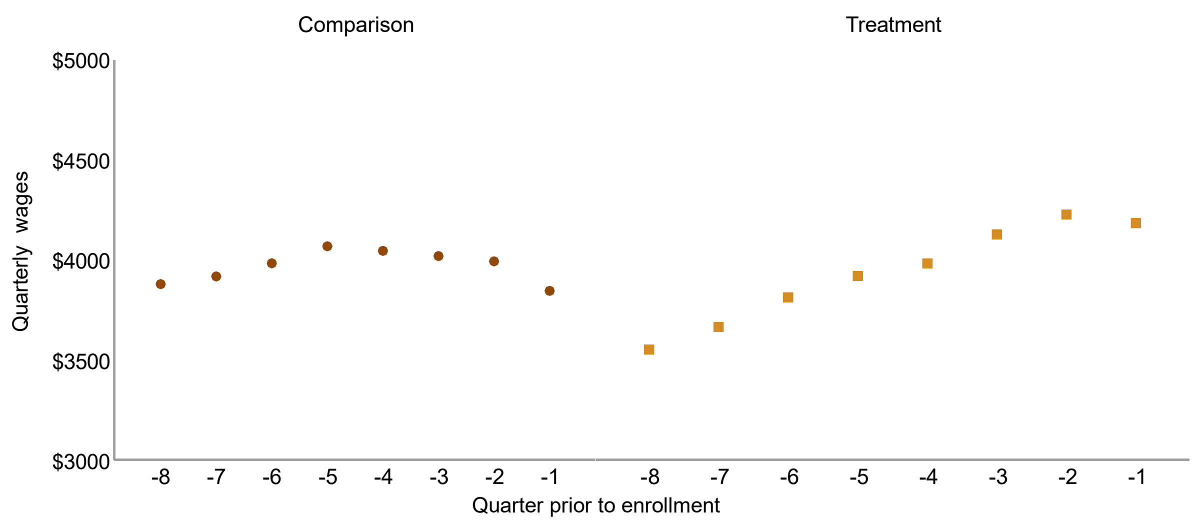


Figure C.3. Pre-Program Wages Trend, Unweighted, RQ 2



RESEARCH QUESTION 3

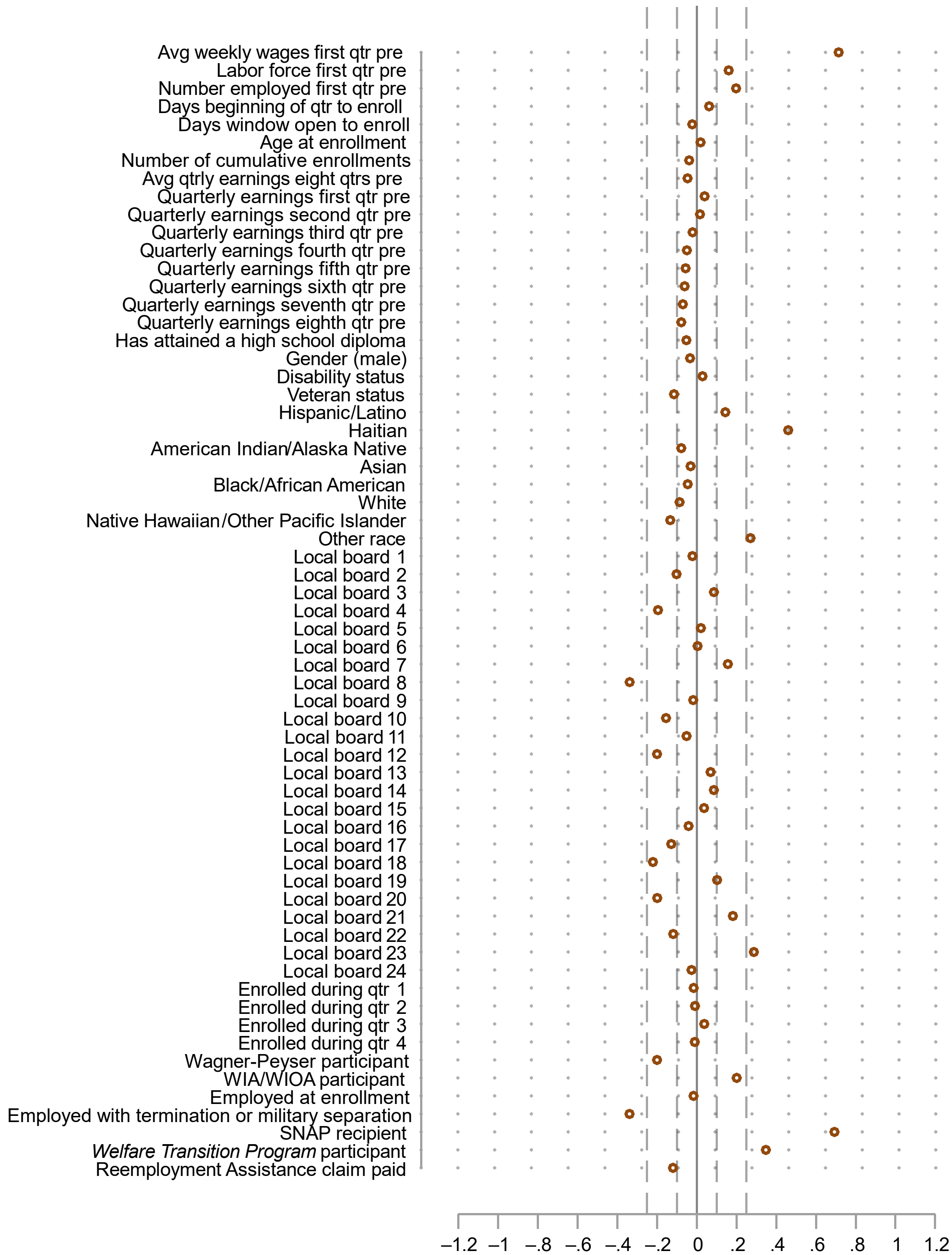
Table C.2. Unweighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3

Variable	Treatment Mean (*n* = 457,449)	Comparison Mean (*n* = 756,820)	Standardized Mean Difference
Continuous Variables			
Average weekly wages 1st quarter prior to enrollment	$886.41	$809.04	0.71
Number in labor force 1st quarter prior to enrollment	733,071	659,274	0.16
Number employed 1st quarter prior to enrollment	697,313	611,980	0.20
Days from beginning of quarter to enrollment date	45.46	43.82	0.06
Days from beginning of study window to enrollment date	480.21	487.26	–0.02
Age at enrollment	39.34	39.08	0.02
Number of cumulative enrollments in workforce services	1.98	2.03	–0.04
Average quarterly earnings for eight quarters prior to enrollment	$1,787.83	$1,989.92	–0.05
Quarterly earnings prior to enrollment:			
1st quarter prior	$1,138.58	$960.69	0.04
2nd quarter prior	$1,668.73	$1,590.16	0.02
3rd quarter prior	$1,834.22	$1,950.44	–0.02
4th quarter prior	$1,845.08	$2,120.10	–0.05
5th quarter prior	$1,967.19	$2,284.11	–0.06
6th quarter prior	$1,989.65	$2,332.47	–0.06
7th quarter prior	$1,969.50	$2,354.42	–0.07
8th quarter prior	$1,889.73	$2,326.98	–0.08
Dichotomous Variables			
Attained a high school diploma	89.1%	89.9%	–0.05
Gender (male)	45.8%	47.2%	–0.03
Disability status	7.1%	6.8%	0.03
Veteran status	6.6%	7.8%	–0.12
Race/ethnicity:			
Hispanic/Latino	28.5%	23.9%	0.14
Haitian	3.0%	1.4%	0.46
American Indian/Alaska Native	1.4%	1.6%	–0.08
Asian	1.4%	1.5%	–0.03
Black/African American	29.7%	31.4%	–0.05
White	50.4%	53.9%	–0.09
Native Hawaiian/Other Pacific Islander	0.5%	0.6%	–0.13
Other race	0.1%	0.0%	0.27

Table C.2. Unweighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3

Variable	Treatment Mean (*n* = 457,449)	Comparison Mean (*n* = 756,820)	Standardized Mean Difference
Dichotomous Variables (continued)			
Local board:			
1	3.5%	3.6%	–0.02
2	1.4%	1.6%	–0.10
3	1.2%	1.1%	0.09
4	1.7%	2.4%	–0.20
5	2.3%	2.2%	0.02
6	1.5%	1.5%	0.00
7	1.1%	0.9%	0.16
8	3.5%	6.0%	–0.34
9	1.9%	1.9%	–0.02
10	2.4%	3.0%	–0.16
11	3.8%	4.1%	–0.05
12	7.8%	10.5%	–0.20
13	4.4%	4.0%	0.07
14	4.8%	4.2%	0.09
15	8.4%	8.0%	0.04
16	2.8%	3.0%	–0.04
17	3.1%	3.8%	–0.13
18	2.6%	3.7%	–0.22
19	1.0%	0.8%	0.10
20	2.2%	3.1%	–0.20
21	6.9%	5.2%	0.18
22	4.8%	5.8%	–0.12
23	23.4%	16.0%	0.29
24	3.4%	3.6%	–0.03
Enrolled during quarter 1	29.5%	30.0%	–0.02
Enrolled during quarter 2	22.2%	22.4%	–0.01
Enrolled during quarter 3	25.0%	23.8%	0.04
Enrolled during quarter 4	23.4%	23.7%	–0.01
Wagner-Peyser participant	94.5%	96.0%	–0.20
WIA/WIOA participant	5.5%	4.0%	0.20
Employed at enrollment	8.2%	8.4%	–0.02
Employed with termination notice or military separation at enrollment	0.4%	0.6%	–0.34
SNAP recipient	9.9%	3.4%	0.69
Welfare Transition Program participant	7.0%	4.1%	0.35
Reemployment Assistance claim paid	9.2%	11.0%	–0.12

Figure C.4. Unweighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3



APPENDIX D. DATA AND VARIABLES

OVERVIEW

No original data were collected for the outcomes study. Individual-level outcome data, covariate data, and contextual/local board-level economic data used for propensity score weighting and analytical modeling were collected from the Florida Department of Economic Opportunity (DEO) and the *U.S. Bureau of Labor Statistics* (BLS).

The outcomes study required receipt of pre- and post-exposure individual-level *Unemployment Insurance* (UI) wage data (i.e., quarterly wages) for individuals who enrolled in services at one of the 24 local boards during the treatment or comparison enrollment period. Data sharing permissions were covered by the original evaluation contract; all necessary precautions to ensure confidentiality and compliance with requirements from the State of Florida regarding data security practices were followed. DEO securely transferred data for the outcomes study on a monthly basis from November 2017 through May 2019. Each month, four individual-level data files were transferred; these files contained Workforce Investment Act (WIA)/Workforce Innovation and Opportunity Act (WIOA) enrollments, Wagner-Peyser (WP) enrollments, demographic data, and certified UI wage data. Data received were sufficient to cover the treatment and comparison enrollment periods, as well as the pre- and post-program data required for conducting the outcomes study. Final analytic sample data were received in May 2019.

For each individual who engaged with a local board during the comparison or treatment enrollment period, we retained eight quarters of pre-exposure (i.e., baseline) wage data and up to eight quarters of data following their enrollment date.[115] Our benchmark analysis of outcomes is based on the employment outcomes at the fourth full-quarter post-enrollment; despite the lag time in receiving fully certified wage data, data for all participants were complete at this time point.[116] Although the specific pre-program quarterly observations varied for each individual, depending on when they engaged with local board services relative to the enrollment period, the range of data received fell between the eight quarters prior to January 1, 2012 (start date of the comparison group enrollment period) through four quarters after June 30, 2017 (end date of the treatment group enrollment period).[117] Data collection procedures were identical for all treatment and comparison group members.

DATA PROCEDURES

As data were obtained in multiple files, a unique identifier was used to create a composite data set. This ID was created by DEO and all data deemed personally identifiable by DEO were removed prior to data transmission. On a monthly basis, staff at DEO used a secure link to upload the data files to The Policy & Research Group's (PRG's) Citrix ShareFile site, and files were then stored on a password-protected, limited-access server that requires two-factor authentication for access.

The participant demographic data were submitted in an individual-level, wide format, with one record per individual, identified with the unique ID. The WIOA, WP, and wage data sets were submitted in an

[115] In the approved *Evaluation Design Report* (EDR), we outlined a plan using five quarters of pre-intervention wage/employment data prior to enrollment for each participant. Because data were available, we included eight quarters of pre-intervention data.

[116] For participants who enrolled earlier in the comparison and treatment windows, we conduct a sensitivity study to analyze outcomes at the eighth full-quarter post-enrollment. Eighth quarter wage data were not available for individuals who enrolled in the latter two quarters of each enrollment period.

[117] This refers to the range for Research Questions 1 and 2. The enrollment period for Research Question 3 was three quarters longer than the enrollment period for Research Questions 1 and 2. For Research Question 3, the range of data received fell between the eight quarters prior to January 1, 2012 through four quarters after March 31, 2018.

individual-level, person-period format, where an individual had multiple observations in the data set based on the number of enrollments (WIOA or WP) or quarters of wage records, again identified using the unique ID.[118]

In order to create an analysis-ready data set, the WIOA, WP, and wage data sets were reshaped into a wide format and merged with the demographic data set using the unique ID. Data were again reshaped to create a final data set that contained, per observation, data pertaining to an individual's enrollment into local board (i.e., WIOA or WP) services, along with all of the individual's demographic and wage data. Next, the unique ID numbers are transposed so that each unique enrollment into local board services is analyzed as a unique individual, regardless of whether it is actually the same individual. We then drop cases that have enrollment dates outside of our comparison or treatment windows. And finally, we keep only wage data that are relevant to the enrollment date for each observation (i.e., eight quarters pre- and post-enrollment).

DATA SOURCES

FLORIDA DEPARTMENT OF ECONOMIC OPPORTUNITY

Beginning in February 2017, DEO acted as the data manager for the PFM project. DEO compiled and verified data on a monthly basis for CareerSource Florida (CSF) to use to calculate the PFM metrics and assess Local Workforce Development Board (LWDB) performance. In the fall of 2017, PRG requested to be included in data sharing; we received our first data submission in November 2017.

The data received from DEO consist of four individual-level text data files. One file contains basic demographic information, such as gender and race – we refer to this as the participant data set. Two files contain information specific to an individual's enrollment(s) within WIOA and/or WP, such as entry and exit date, as well as demographic data collected at entry into those programs. Finally, a certified wage data file includes the wages associated with each quarter and year. All four data sets contain unique identifiers that are used to link individuals after receipt.[119]

With the exception of two variables, the source of the participant data (participant data set, WIOA data set, WP data set) is the *Employ Florida* system, which sends an updated extract of the data to DEO on a nightly basis. The variables that describe an individual's *Supplemental Nutrition Assistance Program* and *Welfare Transition Program* participation are sourced from the *One Stop Service Tracking* system and are incorporated into the data by DEO prior to submission. The *U.S. Department of Revenue* provides the certified UI wage data.

U.S. BUREAU OF LABOR STATISTICS

In addition to data submitted by DEO, PRG accessed publicly available data from the BLS. Specifically, we used publicly available aggregate figures provided by the *Local Area Unemployment Statistics* (LAUS) program and the *Quarterly Census of Employment and Wages* (QCEW). Data made available by BLS include the following contextual economic indicators: unemployment rate, total labor force, total employed, and total unemployed (all from the LAUS), and average weekly wages (from the QCEW). We downloaded data for each economic quarter at the Florida county level between 2010 and 2018; these data were later used to calculate the workforce region-level variables used in the outcomes study.[120]

[118] Data pertaining to WIA program participants were also included in this study. The transition from WIA to WIOA took place on July 1, 2016, during our treatment group enrollment period. For simplicity, we refer to those enrolled in WIA or WIOA as WIOA participants.
[119] This ID was created by DEO and all data deemed personally identifiable by DEO were removed prior to data transmission.
[120] The county-to-workforce board key was obtained here on December 10, 2018: https://careersourceflorida.com/your-local-team/

COVARIATES

We provide an overview of the individual-level covariate data that were considered for inclusion into the propensity score estimating models and the analytic models, including a description and the source of each variable. Overall, covariate data were complete for all participants; some background variables, as noted in the source line in Table D.1, are available for WIOA participants only.[121]

Table D.1. Covariate Variables

Variable Name	Description of Variable
Age at entry	Age at entry is calculated using the individual's date of birth and the entry date. Source: Participant data set
Gender	Gender is reported as either male or female. Source: Participant data set
Hispanic/Latino	Dummy variable reporting ethnicity as Hispanic/Latino (1) or not (0). Source: Participant data set
Native American	Dummy variable reporting race as Native American (1) or not (0). Source: Participant data set
Asian	Dummy variable reporting race as Asian (1) or not (0). Source: Participant data set
Black	Dummy variable reporting race as Black (1) or not (0). Source: Participant data set
Pacific Islander	Dummy variable reporting race as Pacific Islander (1) or not (0). Source: Participant data set
White	Dummy variable reporting race as White (1) or not (0). Source: Participant data set
Other race	Dummy variable reporting race as another race (1) or not (0). Source: Participant data set
Local board ID	Numeric variable (1–24) identifying the LWDB where the enrollment took place. Source: WIOA and WP data sets
Education level	Dummy variable reporting the individual's highest level of education at entry as high school diploma or higher (1) or less than high school diploma (0). Source: WIOA and WP data sets
Veteran status	Dummy variable indicating if the individual is a Veteran at entry (1) or not (0). Source: WIOA and WP data sets
Disability status	Dummy variable indicating if the individual is disabled at entry (1) or not (0). Source: WIOA and WP data sets
Rehabilitation Assistance claim paid	Dummy variable indicating if the individual had a Rehabilitation Assistance claim paid during program participation (1) or not (0). Source: WIOA and WP data sets
Homeless	Dummy variable indicating if the individual is homeless at entry (1) or not (0). Source: WIOA data set
Offender	Dummy variable indicating if the individual is an offender at entry (1) or not (0). Source: WIOA data set
Low income	Dummy variable indicating if the individual is low income at entry (1) or not (0). Source: WIOA data set
Limited English proficiency	Dummy variable indicating if the individual has limited English proficiency at entry (1) or not (0). Source: WIOA data set
Single parent	Dummy variable indicating if the individual is a single parent at entry (1) or not (0). Source: WIOA data set
Average quarterly wages prior to entry	Certified quarterly wages earned for eight quarters prior to entry in local board services during the comparison or treatment enrollment periods. Source: Wage data set
Employment status prior to entry	Employed (1) or not (0) for eight quarters prior to entry in local board services during the comparison or treatment enrollment periods. Employment is defined as having earned $100 or more during that quarter. Source: Wage data set

[121] The data set included a zip code variable; however, a high number of missing or incorrect zip codes prevented its use; in its place, we use the arguably more appropriate local board variable, which tells us where the individual enrolled in services.

FIXED CONTEXTUAL VARIABLES

In Table D.2, we present control variables that provide contextual information about the timing and nature of the individual's entry into the study and that do not change over time.

Table D.2. Fixed Contextual Variables

Variable Name	Description of Variable
Quarter of entry	A series of four dummy variables, constructed using the entry date variable, which indicates that an individual enrolled during any quarter (of any year) (1) or not (0). Q1 = July–September; Q2 = October–December; Q3 = January–March; Q4 = April–June. Source: WIOA and WP data sets
Month of entry	A continuous count variable (1–72), constructed using the entry date variable, which indicates the overall month of entry into the study window, where 1 = the first month of the study window and 72 = the last month of the study window.[122] Source: WIOA and WP data sets
Number of days from beginning of enrollment window to entry date	A continuous variable (0–730 days), constructed using the entry date variable, which indicates the number of days from the beginning of the enrollment window (for comparison – July 1, 2012; for treatment – July 1, 2015) to the entry date. Source: WIOA and WP data sets
Number of days from beginning of enrollment quarter to entry date	A continuous variable (0–91 days), constructed using the entry date variable, which indicates the number of days from the beginning of the quarter of enrollment to the entry date. Source: WIOA and WP data sets
Cumulative number of enrollments in local board services	For each observation, the cumulative number of enrollments is calculated by counting the number of enrollments (either WIOA or WP) occurring in the seven years prior to and including the current enrollment. This gives an indicator of the level of engagement that an individual has had with the local board(s) over time. Source: WIOA and WP data sets
Funding stream	A series of two dummy variables indicating whether an enrollment is a WIOA enrollment (1) or not (0) and whether an enrollment is a WP enrollment (1) or not (0). These variables are mutually exclusive. Source: WIOA and WP data sets
Receipt of WIOA training	A dummy variable indicating whether an individual received WIOA training services during the enrollment (1) or not (0). Source: WIOA data set

TIME-VARIANT ECONOMIC AND CONTEXTUAL VARIABLES

Time-variant economic and contextual variables capture second-order processes existing outside of the control of the study design that may have influence on outcomes for participants who were engaging with local boards at different times and in different locations. Including these variables can help diminish any potential bias stemming from variable economic conditions across time and geographic location. All time-variant and contextual variables that were in included in the analysis are detailed in Table D.3.

[122] Because there are no enrollments occurring during the third year of the comparison and treatment study periods for Research Questions 1 and 2, this variable will only take on a value of 1–24 or 37–60.

Table D.3. Time-Variant Economic and Contextual Variables

Variable Name	Description of Variable
Local board-level unemployment rate	A continuous variable describing the unemployment rate for the local area where the enrollment occurred. Data are downloaded at the county level. In each local area, the unemployment rate is calculated by dividing the total number of individuals that are unemployed in by the total number of individuals. Source: BLS (LAUS)
Local board-level number in the labor force	A continuous variable describing the total number in the labor force for the local area where the enrollment occurred. Data are downloaded at the county level. The local number in the labor force is calculated by summing the total number of individuals in the labor force. Source: BLS (LAUS)
Local board-level number employed	A continuous variable describing the total number of employed individuals for the local area where the enrollment occurred. Data are downloaded at the county level. The local number employed is calculated by summing the total number of employed individuals. Source: BLS (LAUS)
Local board-level number unemployed	A continuous variable describing the total number of unemployed individuals for the local area where the enrollment occurred. Data are downloaded at the county level. The local number unemployed is calculated by summing the total number of unemployed individuals. Source: BLS (LAUS)
Local board-level average weekly wages	A continuous variable describing the average weekly wages for the local area where the enrollment occurred. Data are downloaded at the county level. The local figure is calculated by first calculating the weight that will be applied to each county – this is based on the percentage of jobs attributed to each county within the area. The local average is then calculated using the following formula: (weight_county_1 * avg_weekly_wage_1) + (weight_county_n * avg_weekly_wage_n). Source: BLS (QCEW)

OUTCOME VARIABLES

Employment outcomes are assessed with three measures: employment status (i.e., whether one was employed during a given quarter), wages (i.e., the total wages earned in a quarter), and time to employment. We outline how these measures are constructed in Table D.4.

Table D.4. Outcome Variables

Variable Name	Description of Variable
Employment status (used for Research Questions 1 and 3)	An individual is considered to be employed in any given quarter if they have earned $100 or more in the quarter being measured. Dummy variables are created for each quarter where 1 means that an individual was employed (according to the above definition) and 0 means the individual was not employed. Source: Wage data set
Average quarterly wages (used for Research Question 2)	This is a continuous variable measuring the amount of certified, employer-reported wages reported in the UI wage data set. All earnings variables are adjusted for inflation using the Consumer Price Index for all Urban Consumers (CPI-U) to the beginning of the treatment group outcome period – which is July 2017. Source: Wage data set

MISSING DATA

Missing outcome data are not imputed. At the guidance of DEO, given that the wage data are certified (i.e., final), missing wage data are to be interpreted as a lack of wages. Therefore, missing wage data were logically edited to reflect zero wages for the quarter(s) in which the data were missing. Covariate data were, on the whole, very complete. A minimal number of cases were dropped due to invalid or out-of-range ages.[123] Aside from those individuals, anyone whose data were submitted to us who had an enrollment in either the treatment or comparison period was included in the analysis.

[123] For Research Questions 1 and 2, 1 individual was dropped due to a missing date of birth variable, 20 individuals were dropped because they were under 14 years of age (the eligibility criteria for WIOA youth services), and 7 individuals were dropped because of otherwise invalid birthdates (e.g., birthdate was a "default" missing birthdate, such as 10/10/1910). An additional 3 cases were dropped from the sample for Research Question 3; these individuals were under 14 years of age at enrollment.

APPENDIX E. STUDY PERIOD

The Performance Funding Model (PFM) was officially launched as a statewide policy in July 2015. As our objective is to estimate the average program effects of this statewide policy switch on individuals exposed to the PFM through engagement in their local workforce boards, our treatment study period begins on July 1, 2015 and runs through June 30, 2018. As this is a statewide intervention and randomization into the treatment condition is infeasible, we use a retrospective comparison group covering July 1, 2012 to June 30, 2015.

Evaluating the program from the actual start date has the advantage of improving the proximity of treatment and comparison groups in time. This means that the economic and political contexts, which vary over time, will tend to be more similar. The principal limitation of the study design employed here is the possibility of extraneous historical effects. Keeping the two windows proximate will allow for maximum overlap in the contextual conditions that are experienced by the treatment and comparison groups and should minimize historical and selection confounds.

For Research Questions 1 and 2, our benchmark approach is to observe outcomes for all participants in the fourth quarter post-enrollment. Participants who enroll to receive local board services within the enrollment period will be observed four full quarters following the enrollment quarter. The study period is 72 months (36 months or 12 quarters for the treatment and comparison windows) in duration, but the enrollment period within each of the 12-quarter windows is 8 quarters. The enrollment and observation schedule is illustrated in Table E.1. Schedules for the treatment and comparison windows are identical, except for the year of their start and end dates; Q1 for the comparison group corresponds to July through September 2012 and Q1 for the treatment group corresponds to July through September 2015.

Table E.1. Enrollment and Post-Program Periods for RQ 1 & 2

	Quarter Within Study Window											
Enrollment Quarter	1	2	3	4	5	6	7	8	9	10	11	12
1	Enroll				Post							
2		Enroll				Post						
3			Enroll				Post					
4				Enroll				Post				
5					Enroll				Post			
6						Enroll				Post		
7							Enroll				Post	
8								Enroll				Post

The outcome quarter at which we measure employment outcomes is not congruent with the timing that was proposed in the original *Evaluation Design Report* (EDR).[124] We considered multiple factors in the

[124] In the memo detailing revisions to the approved EDR (submitted to the National Evaluation Coordinator on September 30, 2015), we proposed that the treatment and comparison enrollment and exit periods were 27 months in duration (9 quarters), with the observation of employment outcomes occurring 3 quarters post-exit from services. This meant that participants had to enroll *and exit* within the first 9 quarters in their respective windows to be eligible for inclusion into the study. Although outcomes would be observed for any given client 3 quarters post-exit from services, the treatment and comparison windows consisted of a maximum of 9 quarters in which the participants could potentially enroll and exit, plus 3 quarters to observe outcomes for clients who exited services in the 9th quarter.

process of simplifying the timeline. Ultimately, the decision to change the outcome observation time point to the fourth quarter post-enrollment date was informed by consistent findings in research on active labor market programs which suggest that job training programs tend to have small or negative impacts on employment outcomes for periods of less than one year.[125] Our analytic approach is to compare differenced pre-post change for both groups, which should capture improvement as either reduced declines or increased gains in quarterly wages. Nevertheless, there is the possibility that even the differenced outcome estimates are attenuated in the first year (referred to in the literature as the "lock-in" period), so we have decided to move the follow-up observation point as far back as practicable for the evaluation. Combined with the shortened enrollment period, this allows us to maximize the amount of post-enrollment data that is used to measure outcomes for our participants.

For Research Question 3, clients could enroll at any time during the 36-month treatment or comparison period, except for the last quarter of each period. The event (employment) was observed at any of the 11 quarters in the study period after the client's enrollment, but the number of quarters being observed depended on the quarter of enrollment. The benchmark study period for Research Question 3 does not differ from that proposed in the approved EDR. Table E.2 provides a visual description of observation quarters for the treatment and comparison groups by enrollment quarter; Q1 for the comparison group corresponds to July through September 2012 and Q1 for the treatment group corresponds to July through September 2015. A grey cell indicates a quarter of observation for those enrolled during the identified enrollment quarter.

Table E.2. *Enrollment and Potential Event Periods for Research Question 3*

Enrollment Quarter	Observation Quarter										
	1	2	3	4	5	6	7	8	9	10	11
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											

[125] Heinrich, C., et al. (November 2009). *New Estimates of public employment and training program net impacts: A nonexperimental evaluation of the Workforce Investment Act Program*, IZA Discussion Paper No. 4569. Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *Economic Journal*, *120*(548), F452–F477. Andersson, F., et al. (September 2013). *Does federally-funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms*, IZA Discussion Paper No. 7621.

APPENDIX F. PROPENSITY SCORE ESTIMATION

We predict the propensity score with a *Generalized Boosted Model* (GBM). Although we specified the logistic analysis in the *Evaluation Design Report* (EDR), we ultimately selected the GBM approach because it balanced samples better (on observed covariates) than did the logistic model. The GBM method also produced estimates faster and reduced the number of subjective decisions because it avoided the necessity to specify functional forms of covariates in the model. The GBM is an automated and data-adaptive algorithm that fits several models using a regression tree and then averages the predictions produced by each model. We used the Boost command in Stata and included the following variables in the estimating equation:[126]

- Education level
- Gender
- Mean age at enrollment
- Disability status
- Veteran status
- Hispanic/Latino
- Haitian
- American Indian/Alaska Native
- Asian
- Black/African American
- White
- Native Hawaiian/Other Pacific Islander
- Other race
- Cumulative number of enrollments
- Employed 1st quarter pre
- Employed 2nd quarter pre
- Employed 3rd quarter pre
- Employed 4th quarter pre
- Employed 5th quarter pre
- Employed 6th quarter pre
- Employed 7th quarter pre
- Employed 8th quarter pre
- Quarterly wages 1st quarter pre
- Quarterly wages 2nd quarter pre
- Quarterly wages 3rd quarter pre
- Quarterly wages 4th quarter pre
- Quarterly wages 5th quarter pre
- Quarterly wages 6th quarter pre
- Quarterly wages 7th quarter pre
- Quarterly wages 8th quarter pre
- Local board 1
- Local board 2
- Local board 3
- Local board 4
- Local board 5
- Local board 6
- Local board 7
- Local board 8
- Local board 9
- Local board 10
- Local board 11
- Local board 12
- Local board 13
- Local board 14
- Local board 15
- Local board 16
- Local board 17
- Local board 18
- Local board 19
- Local board 20
- Local board 21
- Local board 22
- Local board 23
- Local board 24
- Enrolled during quarter 1
- Enrolled during quarter 2
- Enrolled during quarter 3
- Enrolled during quarter 4
- Days from beginning of quarter to enrollment date
- Days from beginning of study window to enrollment date
- SNAP participant
- *Welfare Transition Program* participant
- Reemployment Assistance claim paid
- Employed at enrollment
- Employed with termination notice or military separation at enrollment
- Wagner-Peyser participant
- WIA/WIOA participant

[126] Boost implements the MART boosting algorithm described in Hastie T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York, NY: Springer-Verlag.

We do not include regional-level labor context variables in the propensity score model. When included with the other variables, they resulted in a perfect prediction of the propensity score and no overlap in scores for both groups. Instead of accounting for economic conditions in the selection into groups (i.e., balancing on economic conditions), we employ them explicitly as part of the conditioning variables in the prediction of the parameter of interest (i.e., control for them in the analytic model), in addition to the CSITS modeling of individual-level trends in prior earnings. We are balancing on the characteristics of participants in the propensity score model and then including time-variant controls for the differing regional and temporal economic conditions and individual-level earnings trends in the analytic model.

Our benchmark approach is to use "trimmed" propensity scores.[127] We do not drop any observations that fall outside the region of common support, but re-code the propensity scores that transcend this region at the threshold value. In more specific terms, we first determine the maximum propensity score for the comparison group and the minimum score for the treatment group. For any participants in the treatment group whose propensity score is more than this (comparison) maximum value – we re-code their propensity score to the threshold value. Similarly, any individuals in the comparison group whose propensity score is less than the (treatment) minimum are re-coded to the minimum threshold value.

We also tested alternative strategies – including no trimming and stabilizing, which is another method to mitigate the effect of extreme propensity scores.[128] In all preliminary testing, trimming and alternative strategies produced substantively identical results.

We do not explicitly match comparison to treatment cases. Rather, we weight each case according to its propensity score so that the regression analysis is conducted on the full sample of participants enrolled in the study, but who have been nonparametrically adjusted so that they are more similar: cases are upweighted if they are more alike (according to observed covariates) and down weighted if they are less alike. We use inverse proportional treatment weights to balance the treatment and comparison groups and estimate the average treatment effect (ATE). The formula for the weighting procedure is:

$$\mathrm{w(D,x)} = \frac{D}{e^{\wedge}(x)} + \frac{1-D}{1-e^{\wedge}(x)}$$

where w equals the ATE weight, conditional on treatment status D and conditioning set x, and $e^{\wedge(x)}$ equals the estimated propensity score.

---

[127] Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*, 3661–3679. Imbens, G., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York, NY: Cambridge University Press.

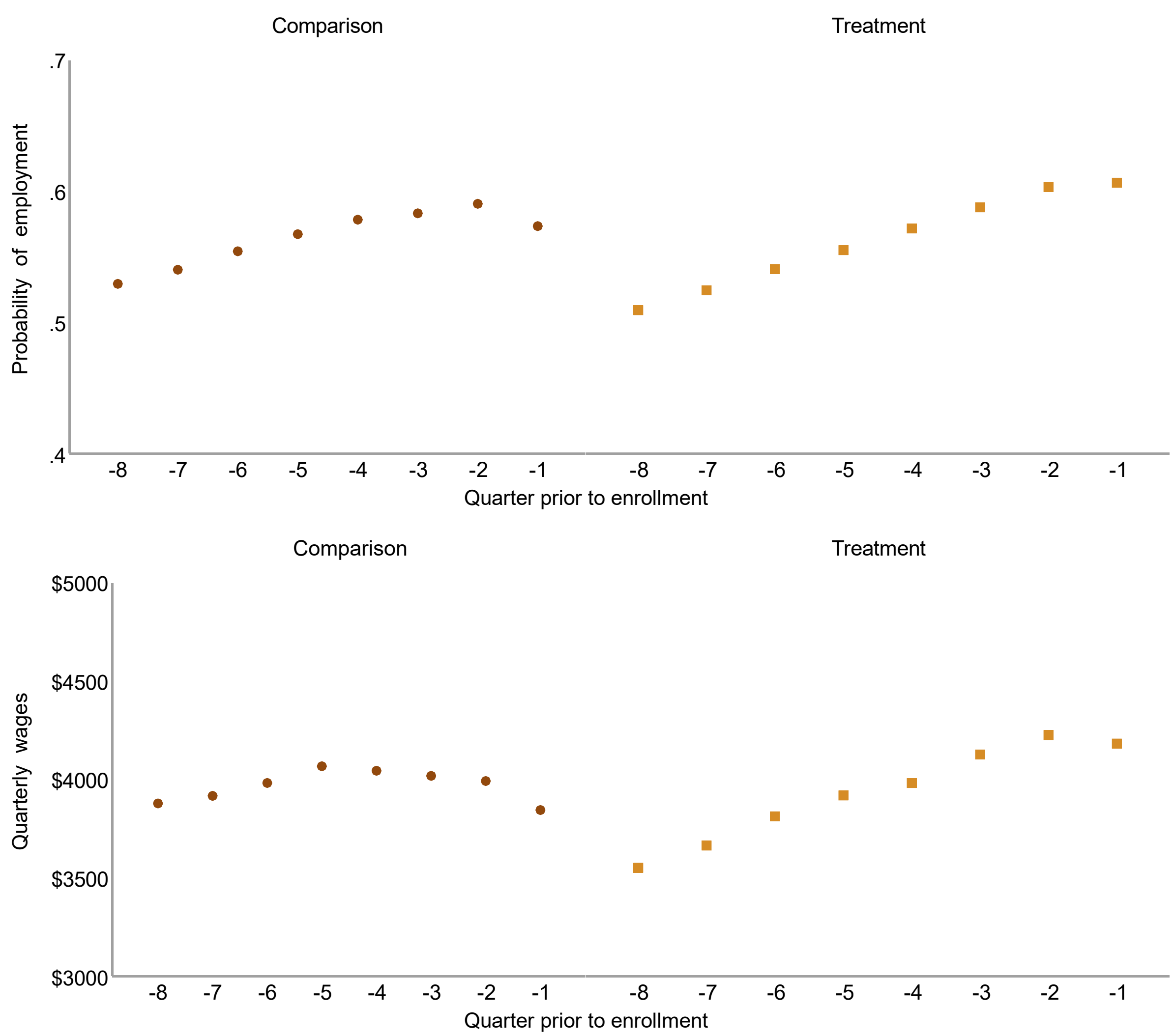
[128] Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*, 550–560. Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed., p. 245). Thousand Oaks, CA: Sage.

APPENDIX G. SELECTING THE APPROPRIATE DIFFERENCING MODEL

In the *Evaluation Design Report* (EDR), we wrote that we would analyze pre-program outcomes to determine if the comparative short interrupted time series (CSITS) or the conventional difference-in-difference (DID) design would be more appropriate. This determination, we said, would be based on whether there was evidence that pre-program outcomes for the PFM and comparison group demonstrate differential trends over time.

A preliminary analysis of quarterly wage and employment data indicate that a CSITS model is a more appropriate analytic design than DID. The internal validity of the DID estimating strategy rests on the assumption of parallel trends, which means that the relative rate of improvement or decline in the outcome must be the same for both groups in the absence of the intervention. Pre-intervention plots of mean employment status (top panel) and wages (bottom panel) for the unweighted, full analytic sample in Figure G.1 illustrate that this assumption is not tenable.

Figure G.1. Pre-Program Employment and Wage Trends, Unweighted, RQ 1[129]



[129] The *y*-axes for Figure G.1 are shortened to see the unweighted pre-program trends clearly.

Figure G.1 illustrates mean quarterly wages and employment status across eight pre-intervention quarters, ordered from the most distal to the final preenrollment quarter. These trends can be understood as the mean relative rate of earnings and employment status change for client cases in the eight quarters prior to enrollment for the population of client cases that received Local Workforce Development Board services under the Wagner-Peyser and Workforce Innovation and Opportunity Act funding streams for both the treatment and comparison groups. The graphic illustrates reasonably well-defined linear trends for treatment and comparison groups.[130]

Figure G.1 also illustrates that the trends for the comparison and treatment groups differ. Although it is more evident for wages than employment status, one can observe in both that the "slope" of the implied linear trajectory for the treatment group is greater than that of the comparison group. Quarter-by-quarter, across the two years prior to enrolling in the study, members of the treatment group on average saw greater rates of improvement to their labor market outcomes (i.e., average quarterly wages and employment status). This is convincing evidence that the parallel trends assumption that is required to produce an unbiased estimate of the effect of the PFM on quarterly wages and employment has been violated. If we were to employ a DID, we would undoubtedly bias any estimate of the effect of the PFM by failing to account for the difference in trends.[131]

In this case, the more appropriate design is the CSITS, which allows us to model and control for those differences in trends in our statistical model. Additionally, from an identification perspective, because these trends are manifest in our data, the CSITS model allows us to quantify the difference in labor market conditions (see the Analytic Methods section) that the two groups are exposed to in their respective study windows. This means we have data to empirically adjust for the so-called time confound that results from using a retrospective comparison window. This sort of control is made even more credible and robust after we apply propensity score weights to maximize the observed equivalence of both groups on individual-level background characteristics (see Appendix M for weighted baseline equivalence statistics).[132]

[130] The mean earnings and employment of study participants decline in varying degrees from a linear pattern in the quarter prior to enrollment in the study. The literature refers to this pattern as the Ashenfelter dip.

[131] By way of example, if we were to use a DID analytic strategy with these data – which would take the mean of eight quarters of wages or employment status as the preenrollment or baseline value for both conditions – we would bias the treatment effect upward because any resulting estimate would confound any growth that occurs with the effect of treatment.

[132] For any remaining variation in labor market conditions that influences the outcomes, we include time-variant controls in the estimating model. However, as the results illustrate, the CSITS model with and without covariates generates substantively identical results, suggesting that the CSITS model is effective at soaking up the extraneous influence of labor market differences on wages and employment.

APPENDIX H. MODIFICATIONS

GENERALIZED BOOSTED MODELING

We initially proposed that we would use logit regression to estimate the propensity score; however, we had difficulty generating weights that improved the overall balance on observed pre-program characteristics. This may have been because the sample was well balanced already in participant characteristics, and it was certainly complicated by the size of the data set. We tried alternative approaches, including the generalized boosted modeling (GBM) approach recommended by Guo and Fraser (2015). The GBM proved more effective at producing propensity scores that balanced the two groups (treatment and comparison) in terms of pre-program characteristics. For details on this procedure, see Appendix F.

SINGLE-LEVEL ANALYTIC MODELS FOR RESEARCH QUESTIONS 1 AND 2

The sample we analyze to respond to Research Questions 1 and 2 is extremely large. The analytic sample for both questions is composed of 2,323,339 individuals observed at 9 time points, resulting in an analysis of 20,910,051 unique observations. It was computationally infeasible to run multilevel models with weights on the full sample. We attempted to run a few preliminary models (mixed and menbreg) with an inverse probability weight in Stata but found that the model never converged to solution. We considered several alternative modeling strategies – subsampling with multilevel models, repeated subsampling with multilevel models and averaging parameter estimates (akin to bootstrap sampling), and full-sample single-level models.

First, we tried selecting a random sample of the full analytic sample with a large (but not too large) sample, such that statistical rules and conventions remain appropriate for statistical inference (e.g., statistical significance is inferred if $p < .05$ using a two-tailed test). We randomly selected 5,000 observations from the full sample and ran multilevel versions of the benchmark analyses, conducting tests of significance on the parameters of interest. Across numerous random samples, we found that inference was, for the most part, substantively consistent (findings were consistently insignificant) but the point estimates themselves were highly variable.

We then tried selecting much larger random samples of the full analytic sample, such that statistical rules and conventions were less appropriate for inference, but because the sample was so large ($n > 10,000$), we expected the variance of those estimates to be narrower and more consistent. We were, as it turned out, misguided. Because wages are not normal in distribution, estimates of mean differences continued to be highly variable and standard deviations remained consistently large regardless of sample size. At the same time, we found that the standard error of the estimate became increasingly narrow as sample size increased. This meant that that statistical significance largely became a function of sample size (as one would expect), but the point estimate was highly variable because we were sampling from a distribution with large standard deviation. With extremely narrow confidence intervals, sampling variation and variability overpowered and dominated the observed effects. Results were unstable and highly variable in material ways. In a series of random samples, we obtained significant negative and significant positive estimates for the coefficient of interest.

Next, we considered conducting multiple subsamples in a way that is similar to a bootstrap sampling method. In this method, we set up an iterative loop of the process we have just outlined, where we randomly sampled a large sample ($n > 10,000$), constructed a propensity score, and then estimated the program effect with the benchmark analytical model multiple times. If we then averaged the estimates produced by each iteration, we found that the coefficients produced by the multilevel models converged toward estimates that had been produced by single-level models using the benchmark methods on the full sample.

We therefore decided to use the single-level model as the estimator of the effect of the PFM on wages and employment because it represented the simplest to explain, most transparent, and parsimonious method of estimating the effect of the program. Further, although standard errors are likely erroneous, a single, single-level model should provide a more accurate estimate of the effect of the program, because it is an estimate derived from a sample that represents the full population of participants rather than an average of sample estimates.

Multilevel and single-level models should provide comparable estimates for the fixed-effects coefficients when the models are identically specified (other than the multilevel structure). This means that the estimates produced by our benchmark single-level model and a multilevel model (which would not converge) should be very similar, if not identical. The limitation to the single-level model is that it can produce erroneous standard errors and confidence intervals because it does not account for clustering. Depending on the degree of variation explained by the group or cluster (in this case, the individual observed over multiple occasions), the single-level model will produce standard errors and confidence intervals that are too small. This seems to be an acceptable limitation because we are primarily interested in producing a reliable and accurate estimate of program effect, and because null-hypothesis testing is of secondary concern in a sample this large.

ESTIMATING TREATMENT EFFECT WITH COUNT MODEL FOR RESEARCH QUESTION 2

In the *Evaluation Design Report* (EDR), we specified using a linear model to estimate the effect of the PFM on wages. Preliminary analysis of the wage data and distributions of the errors produced by the linear models convinced us to alter our benchmark approach and use a count model to estimate effects. We selected a zero-inflated negative binomial (ZINB) model over other candidate count models based on a visual analysis of the distribution of the data, goodness-of-fit statistics, and predicted versus observed statistics.

The wage data are evidently not normal in distribution. Wage data take on integer values (only) and are nonnegative. A histogram of the data demonstrates that they are right-skewed in distribution, with a modal value of zero, roughly $3,900 away from the mean (see Figure J.1 in Appendix J). Although linear regression does not require a normally distributed outcome to produce valid estimates in large samples such as this, it does assume that residuals are independent and identically distributed.[133] A quantile plot of the errors demonstrated that this was not the case. Residuals are not normal in distribution and grow systematically for higher values of wages.[134]

We believe the most conservative approach is to use a statistical count model as our benchmark analytical model. We then used descriptive data and model-fit statistics (Akaike information criterion [AIC] and Bayesian information criterion [BIC], log likelihood) statistics to determine which count model fits the data

[133] It is widely believed that a linear regression model (like OLS) requires a normally distributed outcome to produce valid estimates. This is not accurate in most cases. Especially in large samples, OLS can produce valid estimates for any distribution. Nevertheless, applied researchers are often encouraged to estimate parameters for an underlying probability distribution that best represents these data.

[134] In cases of extremely large samples (such as this), some statisticians rely on the Central Limit and the Gauss–Markov Theorems to argue that even with non-normal errors OLS should perform better (i.e., produce the best linear unbiased estimator). Others contend this argument is less convincing in situations where you have long tailed errors with correlation and nonconstant variance (such as this).

best. These statistics are reported in Appendix J. The outcome data are overdispersed, indicating that a negative binomial model might be more appropriate than a Poisson. Additionally, there are an excessive number of zeros, which suggests that a ZINB model is more appropriate. Finally, model fit statistics, produced in Appendix J, indicate that the ZINB model is the best fitting model. We test our benchmark approach by including a sensitivity study that employs ordinary least squares (OLS) regression with an otherwise identical model. Results for this study are reproduced in Appendix O.

APPENDIX I. SPECIFICATION OF PREFERRED LOGISTIC CSITS MODEL

For Research Question 1, we use the comparative short interrupted time series (CSITS) design to estimate the effect of the PFM on employment. As outlined in the design plan, we fit the model with a logistic regression model because the outcome is dichotomous (employed = 1, 0 otherwise).

The probability of employment (π) is estimated by way of a logistic model with a logit link function. We specify the regression equation:

Logit (π_{ij}) = $\alpha + \beta_1 TX_i + \beta_2 Quarter_j + \beta_3 TX_i * Quarter_j + \beta_4 Post_j + \beta_5 TX_i * Post_j + \beta_n X_{ni}$

where π is the probability of employment (in log odds) for individual *i* at time *j*; *TX* is a treatment group indicator (1 for PFM and 0 otherwise); *Quarter* is a variable for the quarter of observation, which counts from –7 (eighth preenrollment quarter) to 5 (fourth full-quarter post-enrollment); *Post* is a 0/1 indicator of the benchmark outcome observation quarter (1 = fourth full-quarter post-enrollment, 0 otherwise); TX_i * *Quarter* is an interaction term of the treatment and quarter variables; TX_i * *Post* is an interaction term of the treatment and *Post* indicator; X_{ni} is a vector of time-variant economic factors, local board indicators, policy indicators, and individual-level covariates (all mean-centered at zero); and:

α is the intercept or estimated probability of employment when all other variables are held at zero. In this CSITS logit model specification, this equals the log odds of the predicted mean wages for a comparison group member at the last preenrollment quarter;

β_1 is the coefficient for the treatment group indicator. In the CSITS logit model this equals the additive effect in the probability of employment (expressed as log odds) for being in the treatment group (in addition to α) in the last preenrollment quarter;

β_2 is the coefficient for the Quarter counter variable; this is the model estimate of the preenrollment probability of employment trend (expressed in log odds) for the comparison group;

β_3 is the coefficient for the TX_i * Quarter interaction term; this is the model estimate of the additive effect on the preenrollment wage trend for the treatment group. The treatment group's preenrollment wage trend (in log odds) is $\beta_2 + \beta_3$;

β_4 is the coefficient for the Post variable; this is the estimate of the deviation (in log odds) from the preenrollment wage trend for the comparison group four full quarters after enrolling in Local Workforce Development Board (LWDB) services; and:

β_5 is the coefficient for the TX_i * Post interaction term; this is the difference-in-difference estimate for the CSITS model and the estimate of interest for hypothesis testing purposes. It represents the difference in deviation from the preenrollment probability of employment trend (expressed as log odds) between the treatment and comparison groups four full quarters after enrolling in LWDB services. The estimated difference in probability of employment (in log odds) for the treatment group four quarters post-enrollment in LWDB services is $\beta_4 + \beta_5$.

Model estimates are transposed into real quarterly wages in our analysis by way of an algorithm that incorporates both the logit and binomial processes, and a linear regression equation that includes estimated coefficients multiplied by variables fixed at specific times for both groups (see Figure 4). All other non-CSITS indicators (control variables in the count model and predictors in the logit model) are held at their mean values (i.e., zero).

Table I.1. Covariate and Control Variables Included in Candidate Models and Goodness-of-Fit Statistics, RQ 1

Covariates	Model 1	Model 2	Model 3	Model 4	Model 5 (Preferred Model)
CSITS variables					
Tx group indicator	✓	✓	✓	✓	✓
Quarter counter	✓	✓	✓	✓	✓
Tx * qtr interaction	✓	✓	✓	✓	✓
Post variable	✓	✓	✓	✓	✓
Tx * post interaction	✓	✓	✓	✓	✓
Economic control variables					
Average weekly wages		✓	✓	✓	✓
Number in labor force		✓	✓	✓	✓
Number employed		✓	✓	✓	✓
Individual-level covariates					
High school graduate			✓	✓	✓
Gender			✓	✓	✓
Age at enrollment			✓	✓	✓
Disability status			✓	✓	✓
Veteran status			✓	✓	✓
Race/ethnicity			✓	✓	✓
Cumulative enrollments			✓	✓	✓
Program type (WP/WIOA)			✓	✓	✓
Employed at enrollment			✓	✓	✓
Military separation			✓	✓	✓
SNAP recipient			✓	✓	✓
Welfare Transition			✓	✓	✓
Reemployment Assistance			✓	✓	✓
Timing of enrollment[135]			✓	✓	✓
Local Board Variables					
Local boards 1–24				✓	✓
Policy Variable					
WIOA implementation[136]					✓
Goodness of Fit					
AIC	55,407,988	55,372,408	49,876,700	49,609,556	49,608,092
BIC	55,408,076	55,372,540	49,877,236	49,610,448	49,608,996
Likelihood-ratio test					
Model 1 nested in Model 2		35,586***			
Model 2 nested in Model 3			5,495,760***		
Model 3 nested in Model 4				267,194***	
Model 4 nested in Model 5					1,465***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

[135] This includes the quarter of enrollment, the number of days from the beginning of the quarter to the enrollment date, and the number of days from the beginning of the study window to enrollment.
[136] To control for the WIOA policy implementation, we use the date that the legislation was mandated: July 1, 2016.

APPENDIX J. SPECIFICATION OF PREFERRED ZINB CSITS MODEL

For Research Question 2, the zero-inflated negative binomial model (ZINB) is a mixture model that estimates the complete distribution of counts (wages) in two separate components – a zero component that models the probability of false or inflated zeros, and a component that accounts for the "true" zeros and nonzero counts. The implication of the ZINB model is that the excessive zeros, true zeros, and nonzero counts are generated by separate processes. For our purposes, we were not motivated to select the ZINB model for any *a priori* theoretical or hypothetical reasons, but rather because, as shown in Figure J.1, the distribution had more zeros than were predicted by the other count models and the ZINB model was selected as the most explanatory model.

The separate processes – one for the inflated zeros and the other for the true zeros and nonzero counts – are reflected in the model specification. The false zeros are estimated by way of a logistic model with a logit link function. The second part of the model, which estimates the probability of true zeros and nonzero counts is estimated with a regular negative binomial model with a log-link function. We specify the two components here. First the specification of the false zero regression equation:

$$\text{Logit}\ (\pi) = \alpha + \beta_1 Employed \quad + \beta_2 Wages_2$$

where π is the probability of a false zero, α is the intercept, and β1 and β2 are slope coefficients for the inflated zero model. *Employed* is an indicator of employment (1 = yes, 0 otherwise) at baseline or the last preenrollment quarter, and *Wages* is the average wages over the full eight quarters preenrollment. We have no substantive interest in the first part of the model (the inflated portion) but include it to empirically account for the zero inflation in the distribution.

Then the count component of the model:

$$\text{Log}(\mu_{ij}) = \alpha + \beta_1 TX_i + \beta_2 Quarter_j + \beta_3 TX_i * Quarter_j + \beta_4 Post_j + \beta_5 TX_i * Post_j + \beta_n X_n$$

where μ is the expected value, or regression-adjusted mean value of the outcome variable (wages) for individual *i* at time *j*; *TX* is a treatment group indicator (1 for PFM and 0 otherwise); Quarter is a variable for the quarter of observation, which counts from –7 (eighth preenrollment quarter) to 5 (fourth full-quarter post-enrollment); *Post* is a 0/1 indicator of the benchmark outcome observation quarter (1 = fourth full-quarter post-enrollment, 0 otherwise); TX_i * *Quarter* is an interaction term of the treatment and quarter variables; TX_i * *Post* is an interaction term of the treatment and *Post* indicator; X_n is a vector of time-variant economic factors, regional indicators, policy indicators, and individual-level covariates (all mean-centered at zero); and:

α is the intercept or estimated wages when all other variables are held at zero. In this CSITS ZINB model specification, this equals the log of the predicted mean wages for a comparison group member at the last preenrollment quarter;

β_1 is the coefficient for the treatment group indicator. In the CSITS ZINB model this equals the additive effect (in log wages) of being in the treatment group (in addition to α) in the last preenrollment quarter;

β_2 is the coefficient for the Quarter counter variable; this is the model estimate of the preenrollment wage trend (in log wages) for the comparison group;

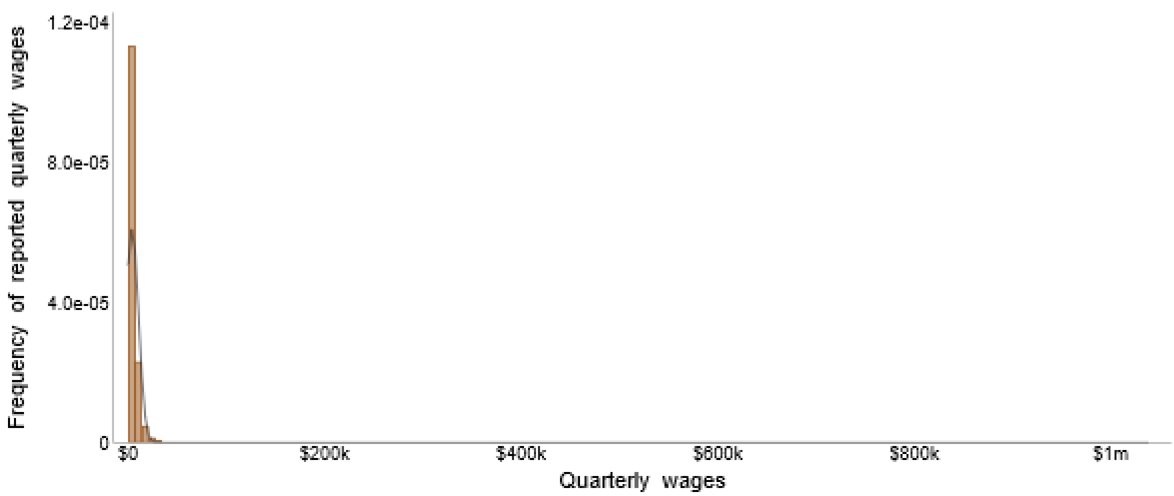
β_3 is the coefficient for the *TX*$_i$ * *Quarter* interaction term; this is the model estimate of the additive effect on the preenrollment wage trend for the treatment group. The treatment group's preenrollment wage trend (in log quarterly wages) is $\beta_2 + \beta_3$;

β_4 is the coefficient for the *Post* variable; this is the estimate of the deviation (in log wages) from the preenrollment wage trend for the comparison group four full quarters after enrolling in Local Workforce Development Board (LWDB) services; and:

β_5 is the coefficient for the *TX*$_i$ * *Post*$_j$ interaction term; this is the difference-in-difference estimate for the CSITS model and the estimate of interest for hypothesis testing purposes. It represents the difference in deviation from the preenrollment wage trend between the treatment and comparison groups four full quarters after enrolling in LWDB services. The estimated difference in (log) wages for the treatment group four quarters post-enrollment in LWDB services is $\beta_4 + \beta_5$.

Model estimates are transposed into real quarterly wages in our analysis by way of an algorithm that incorporates both the logit and binomial processes and a linear regression equation that includes estimated coefficients multiplied by variables fixed at specific times for both groups (see Figure 6 in the main report). All other non-CSITS indicators (control variables in the count model and predictors in the logit model) are held at their mean values (i.e., zero).

Figure J.1. Frequency and Distribution of Pre- and Post-Program Wages, Full Sample[137]



[137] Although not evident in the graphic, the range of quarterly wages is from $0 to $1,037,626, with a modal value of $0.

Table J.1. Covariate and Control Variables Included in Candidate Models and Goodness-of-Fit Statistics, RQ 2

Covariates	Model 1	Model 2	Model 3	Model 4	Model 5 (Preferred Model)
CSITS variables					
Tx group indicator	✓	✓	✓	✓	✓
Quarter counter	✓	✓	✓	✓	✓
Tx * qtr interaction	✓	✓	✓	✓	✓
Post variable	✓	✓	✓	✓	✓
Tx * post interaction	✓	✓	✓	✓	✓
Economic control variables					
Average weekly wages		✓	✓	✓	✓
Number in labor force		✓	✓	✓	✓
Number employed		✓	✓	✓	✓
Individual-level covariates					
High school graduate			✓	✓	✓
Gender			✓	✓	✓
Age at enrollment			✓	✓	✓
Disability status			✓	✓	✓
Veteran status			✓	✓	✓
Race/ethnicity			✓	✓	✓
Cumulative enrollments			✓	✓	✓
Program type (WP/WIOA)			✓	✓	✓
Employed at enrollment			✓	✓	✓
Military separation			✓	✓	✓
SNAP recipient			✓	✓	✓
Welfare Transition			✓	✓	✓
Reemployment Assistance			✓	✓	✓
Timing of enrollment[138]			✓	✓	✓
Local Board Variables					
Local boards 1–24				✓	✓
Policy Variable					
WIOA implementation[139]					✓
Goodness of Fit					
AIC	4.970E+08	4.968E+08	4.927E+08	4.926E+08	4.926E+08
BIC	4.970E+08	4.968E+08	4.927E+08	4.926E+08	4.926E+08
Likelihood-ratio test					
Model 1 nested in Model 2		161,762***			
Model 2 nested in Model 3			4,086,429***		
Model 3 nested in Model 4				79,574***	
Model 4 nested in Model 5					220***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

[138] This includes the quarter of enrollment, the number of days from the beginning of the quarter to the enrollment date, and the number of days from the beginning of the study window to enrollment.
[139] To control for the WIOA policy implementation, we use the date that the legislation was mandated: July 1, 2016.

Table J.2. Test and Fit Statistics

Model Comparison			Preferred Model	Evidence
Poisson regression	BIC = 1.490e+11			
	AIC = 1.490e+11			
vs. negative binomial regression	BIC = 2.836e+08	dif = 1.487e+11	NBRM over PRM	Very strong
	AIC = 2.836e+08	dif = 1.487e+11	NBRM over PRM	
	LRX2 = 1.49e+11	prob = 0.000	NBRM over PRM	*P* = 0.000
Poisson regression	BIC = 1.490e+11			
	AIC = 1.490e+11			
vs. zero-inflated negative binomial	BIC = 2.636e+08	dif = 1.488e+11	ZINB over PRM	Very strong
	AIC = 2.636e+08	dif = 1.488e+11	ZINB over PRM	
Negative binomial regression	BIC = 2.836e+08			
	AIC = 2.836e+08			
vs. Zero-Inflated Negative Binomial	BIC = 2.636e+08	dif = 2.007e+07	ZINB over NBRM	Very strong
	AIC = 2.636e+08	dif = 2.007e+07	ZINB over NBRM	
	Vuong = 923.038	Prob = 0.000	ZINB over NBRM	*P* = 0.000

Table J.3. Comparison of Mean Observed and Predicted Count

Model	Maximum Difference	At Value	Mean Difference
Poisson regression	0.426	0	0.043
Negative binomial regression	0.094	0	0.021
Zero-inflated negative binomial	0.000	9	0.000

Table J.4. Predicted and Actual Probabilities by Model Type

Model		Count	Actual Probability	Predicted Probability	Difference	Pearson
Poisson regression		0	0.426	0	0.426	.
		1	0	0	0	.
		2	0	0	0	.
		3	0	0	0	.
		4	0	0	0	.
		5	0	0	0	.
		6	0	0	0	.
		7	0	0	0	.
		8	0	0	0	.
		9	0	0	0	.
	Sum		**0.427**	**0**	**0.427**	**0**
Negative binomial regression		0	0.426	0.332	0.094	5.60e+05
		1	0	0.035	0.035	7.30e+05
		2	0	0.019	0.019	4.00e+05
		3	0	0.013	0.013	2.80e+05
		4	0	0.010	0.010	2.20e+05
		5	0	0.009	0.009	1.80e+05
		6	0	0.007	0.007	1.50e+05
		7	0	0.006	0.006	1.30e+05
		8	0	0.006	0.006	1.20e+05
		9	0	0.005	0.005	1.10e+05
	Sum		**0.427**	**0.443**	**0.205**	**2.90e+06**
Zero-inflated negative binomial		0	0.426	0.426	0	0.000
		1	0	0	0	762.796
		2	0	0	0	632.503
		3	0	0	0	374.000
		4	0	0	0	535.269
		5	0	0	0	214.152
		6	0	0	0	72.499
		7	0	0	0	309.194
		8	0	0	0	455.498
		9	0	0	0	131.085
	Sum		**0.427**	**0.426**	**0**	**3,486.996**

APPENDIX K. SPECIFICATION OF PREFERRED LOGISTIC DISCRETE-TIME HAZARD MODEL

For Research Question 3, the discrete-time hazard model is as follows:

$$Logit\ (t_{ij}) = [\alpha_j D_j] + \beta_1 VC_i + \beta_p X_{pi}$$

where t_{ij} = the discrete-time hazard for individual *i* at time *j*. In the estimating model, the dependent variable is the indicator (0 = no; 1 = yes) of event occurrence (employment) for individual *i* at time *j*;

D_{ij} = a series of *j* dummy variables that indicate each discrete time period in the study in which the event may happen.[140] In this analysis, the time period is the fiscal quarter. Because the proposed comparison and treatment samples will be followed for the same length of time (11 quarters), we will include 11 dummy variables (e.g., Quarter 1, Quarter 2, Quarter 3). Note that since the quarter of enrollment is the time period start for individual *i*, Quarter 1 refers to the first quarter for that individual and not the first quarter in the study window;

VC_i = a dummy variable that indicates whether an individual *i* is a member of the treatment group (1) or is a member of the comparison group (0);

α_j = the estimate of the conditional logit hazard for individuals in the "baseline" or comparison group at time period *j*. This represents the "risk" of becoming employed for the comparison group at time *j* (expressed in log odds);

β_1= the substantive estimate of interest. This represents the difference in the logit hazard between the comparison (quantified by α) and the treatment groups. This is the incremental shift in the conditional probability of employment (expressed as log odds) for individuals in the treatment group, statistically controlling for the effects of covariates included in the model;

X_{pi} = p vector of X covariates for individual *I*; and:

β_p = the effect of the covariate on the logit hazard for a one-unit change in covariate.

CENSORING

Appropriately dealing with events that are not observed within the study period, known as censoring, is one of challenges in the analysis of data such as these. Censoring arises when the event captured by the dependent variable – in this case becoming employed – fails to occur in the time period under investigation. More conventional analytic models force the issue of how and whether to include these data in the analysis. The options, in broad strokes, are to treat these individuals as missing or to include some systematic decision rules for coding of these data for inclusion (e.g., if the event is not observed, it is coded as 0 meaning an individual *never* became employed). This approach may result in some form of apparent or non-apparent bias. Time hazard modeling, however, permits us to retain the censored data for analysis of risk at each point for which there are data. That is, whereas the censored cases may not be able to inform our estimates of risk beyond the range of time available, discrete-time hazard modeling

[140] Singer and Willett advise that "time should be recorded in the smallest possible units relevant to the process under study." Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

does not require that the case be dropped and does not force re-coding of these data; it permits the use of the data for estimates for each of the periods that are available.

Table K.1. Measures of Fit for Models With Different Representations for Main Effect of Time (n = 4,399,296)

Representation of Time	*n* Parameters	ll (Null)	ll (Model)	AIC	BIC
Constant	1	−1,853,615	−1,853,615	3,707,231	3,707,244
Linear	2	−1,853,615	−1,742,801	3,485,605	3,485,632
Quadratic	3	−1,853,615	−1,736,088	3,472,182	3,472,222
Cubic	4	−1,853,615	−1,735,180	3,470,368	3,470,422
4^{th} order	5	−1,853,615	−1,735,104	3,470,217	3,470,283
5^{th} order	6	−1,853,615	−1,735,103	3,470,218	3,470,298
General	11	−1,853,615	−1,735,088	3,470,198	3,470,344

Table K.2. Covariate and Control Variables Included in Candidate Models and Goodness-of-Fit Statistics, RQ 3

Covariates	Model 1	Model 2	Model 3	Model 4	Model 5 (Preferred Model)
Discrete-time hazard variables					
Tx group indicator	✓	✓	✓	✓	✓
Time dummy variables	✓	✓	✓	✓	✓
Economic control variables					
Average weekly wages		✓	✓	✓	✓
Number in labor force		✓	✓	✓	✓
Number employed		✓	✓	✓	✓
Individual-level covariates					
High school graduate			✓	✓	✓
Gender			✓	✓	✓
Age at enrollment			✓	✓	✓
Disability status			✓	✓	✓
Veteran status			✓	✓	✓
Race/ethnicity			✓	✓	✓
Cumulative enrollments			✓	✓	✓
Program type (WP/WIOA)			✓	✓	✓
Employed at enrollment			✓	✓	✓
Military separation			✓	✓	✓
SNAP recipient			✓	✓	✓
Welfare Transition			✓	✓	✓
Reemployment Assistance			✓	✓	✓
Timing of enrollment[141]			✓	✓	✓
Local Board Variables					
Local boards 1–24				✓	✓
Policy Variable					
WIOA implementation[142]					✓
Goodness of Fit					
AIC	3,470,196	3,469,280	3,405,914	3,399,399	3,399,072
BIC	3,470,356	3,469,480	3,406,446	3,400,237	3,399,923
Likelihood-ratio test					
Model 1 nested in Model 2		922***			
Model 2 nested in Model 3			63,417***		
Model 3 nested in Model 4				6,561***	
Model 4 nested in Model 5					329***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

[141] This includes the quarter of enrollment, the number of days from the beginning of the quarter to the enrollment date, and the number of days from the beginning of the study window to enrollment.
[142] To control for the WIOA policy implementation, we use the date that the legislation was mandated: July 1, 2016.

APPENDIX L. OVERVIEW OF SENSITIVITY STUDIES

BENCHMARK MODEL, NO COVARIATES

We have presented theoretical and empirical justification for the specification of the preferred model. Nevertheless, to test the robustness of this decision we also run regressions, without any independent variables in the estimating model. The logit model retains the comparative short interrupted time series (CSITS) dummy variables and interaction terms and the zero-inflated negative binomial (ZINB) model retains the CSITS components and the inflate terms. Results for these tests are discussed in the narrative and presented in Appendices N and O.

OLS REGRESSION, PREFERRED MODEL

To empirically test whether results were sensitive to our selection of benchmark statistical models (logit and ZINB) we fit the preferred models for both Research Questions 1 and 2 using OLS. The estimating equations were identical – except for the fact that in the regression for Research Question 2, the inflate part of the model is excluded. Results for these tests are discussed in the narrative and presented in Appendices N and O.

For Research Question 1, the linear probability model (LPM) is an alternative and simpler to interpret model than a logistic regression. Although a logit remains the preferred model, the LPM should provide similar estimates when the modeled probability is between 0.2 and 0.8.[143] The LPM is also a viable alternative to the logit when the regressors are categorical – because one is not really modeling a continuous probability function, but rather discrete probabilities associated with different discrete values of *x*. This is especially true if the model includes interactions between the discrete terms.[144]

For Research Question 2, the outcome and model errors were non-normal in distribution. As described in the report, with a sample this size there is little consensus on which the estimating model is more appropriate – a count model whose probability distribution best represents the data, or ordinary least squares (OLS), which should provide the best linear unbiased estimate in (extremely) large analytic samples.

BENCHMARK MODEL, NO WEIGHTING

Propensity scores have been widely used in applied research; however, the literature has long warned that propensity scores can be misused or misapplied and that propensity score methods may exacerbate selection issues if the required propensity score assumptions have not been met.[145] It was our expectation that because our sample was well balanced to begin with and the overlap on conditioning variables and resulting propensity scores was good, the application of the inverse probability weight (IPW) in our preferred estimating models would not unwittingly insinuate bias into our estimates. However, because the potential selection issues would only be manifest in the unobservable, this is all speculation. To empirically test this expectation and determine what the estimate would be without weighting (and whether that would impact substantive findings), we fit the preferred models without IPW (weights). We expected results would be different, but not dramatically so, because the samples were well balanced to

[143] Long, J. S. (1997). *Regression models for categorical and limited dependent variables* (1st ed.). Thousand Oaks, CA: Sage.
[144] Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion* (1st ed.). Princeton, NJ: Princeton University Press.
[145] Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*(5), 1017–1098. Smith, J., & Todd, P. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1–2), 305–353.

begin with. Results for these tests are discussed in the Effect of PFM on Employment and Wages section of the main report and are presented in Appendices N and O.

CONTEMPORANEOUS COMPARISON GROUP

To empirically test whether our benchmark analyses are identifying the effect of the Performance Funding Model (PFM) on wages and employment without apparent bias attributable to the off-year comparison, we conduct a sensitivity analysis that contrasts wage and employment outcomes for the original (full) PFM group with a contemporaneous but nonequivalent comparison group that did not benefit from the PFM. This new contrast removes the “time” confound that is characteristic of the benchmark natural experiment, but it does so at the cost of changing the nature of counterfactual. In this study, we modify our selection procedures for the comparison group and observe outcomes for that group during the treatment period itself. We explain these procedures below.

To be eligible for selection into the comparison sample, individuals must:

1. Be selected into the benchmark comparison group
2. Have complete wage and employment data throughout the treatment period
3. Not have an enrollment in LWDB services at any time during the treatment group enrollment period

Because the members of this group are not enrolling in Local Workforce Development Board (LWDB) services during the treatment window, we constructed an artificial “enrollment” criterion that retains much of the exogenous assignment (i.e., the independence of treatment assignment and outcomes). The selection procedure also isolates the comparison group from any PFM exposure. Individuals who met the above criteria were considered enrolled in the sensitivity study as comparison group members exactly three years after the date that they originally enrolled in the benchmark comparison period. In other words, we are creating an artificial enrollment for these participants in the treatment window, three years after their enrollment in the benchmark comparison window. The delay of three years ensures that both enrollments occurred during the same time of year and in the same relative time point within both windows.

Outcome data are collected for these subjects in the same manner as the benchmark study, but this time we collect those data relative to the artificial enrollment date. We collect eight quarters of wage and employment data prior to this enrollment date and observe post-enrollment outcomes four full quarters after it.

Identification is aided by the fact that both groups have elected to use LWDB services in Florida. However, since at the time of their artificial enrollment, the sensitivity study comparison group has not experienced the same shock (e.g., unemployment or underemployment) that motivated their actual enrollment in LWDB services in the benchmark comparison period, we recognize that the two groups may have meaningful differences in motivation and unobserved selection effects may be lurking. Identification in this study, therefore, leans more heavily on the conditioning set in the propensity score weighting. Baseline equivalence statistics are acceptable and presented in Appendices C (unweighted) and M (weighted).

Additionally, it is important to remember that the contrast (and counterfactual) are different from the benchmark study. In the benchmark study, we are contrasting receipt of LWDB services in Florida with the PFM turned on (treatment) with the receipt of LWDB services in Florida with the PFM turned off (at an earlier time). In this case, we are contrasting the receipt of LWDB services in Florida with the PFM turned on (treatment) with not receiving LWDB services in Florida during the same period (but with a group that once received LWDB services in the past). The counterfactual thus becomes the deviation from the trend established by a group of individuals who once received LWDB services (but do not currently), rather than those who have actually received LWDB services (but at a different time).

We employ this procedure as a sensitivity study for both Research Questions 1 and 2. All analytical methods are identical to those used in the preferred benchmark models. Model estimates are presented in Appendices N and O.

APPENDIX M. BASELINE EQUIVALENCE FOR BENCHMARK ANALYSES (WEIGHTED SAMPLES)

RESEARCH QUESTIONS 1 AND 2

Table M.1. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2

Variable	Treatment Mean (*n* = 994,578)	Comparison Mean (*n* = 1,328,761)	Standardized Mean Difference
Continuous Variables			
Average weekly wages 1st quarter prior to enrollment	$876.81	$804.22	0.58
Number in labor force 1st quarter prior to enrollment	707,268	644,938	0.05
Number employed 1st quarter prior to enrollment	671,659	596,407	0.09
Days from beginning of quarter to enrollment date	44.7	43.4	0.01
Days from beginning of study window to enrollment date	360.1	350.4	–0.01
Age at enrollment	39.3	39.3	0.03
Number of cumulative enrollments into workforce services	2.1	2.1	0.01
Average quarterly earnings for eight quarters prior to enrollment	$3,934.59	$3,969.82	0.01
Quarterly earnings prior to enrollment:			
1st quarter prior	$4,184.55	$3,846.24	0.03
2nd quarter prior	$4,227.99	$3,993.99	0.03
3rd quarter prior	$4,129.06	$4,020.21	0.03
4th quarter prior	$3,983.54	$ 4,046.31	0.01
5th quarter prior	$3,920.44	$4,069.22	0.00
6th quarter prior	$3,813.83	$3,984.11	0.00
7th quarter prior	$3,665.19	$3,918.54	–0.01
8th quarter prior	$3,552.12	$3,879.94	–0.02
Dichotomous Variables			
Attained a high school diploma	91.3%	91.1%	0.04
Gender (male)	46.3%	48.1%	–0.03
Disability status	5.1%	4.8%	0.08
Veteran status	5.4%	6.9%	–0.07
Race/ethnicity:			
Hispanic/Latino	27.6%	24.4%	0.04
Haitian	3.0%	1.2%	0.15
American Indian/Alaska Native	1.3%	1.3%	–0.03
Asian	1.4%	1.4%	–0.01
Black/African American	30.2%	29.3%	–0.01
White	52.1%	56.0%	–0.04
Native Hawaiian/Other Pacific Islander	0.5%	0.6%	–0.15
Other race	0.0%	0.0%	0.33

Table M.1. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2

Variable	Treatment Mean (*n* = 994,578)	Comparison Mean (*n* = 1,328,761)	Standardized Mean Difference
Dichotomous Variables (continued)			
Local board:			
1	3.2%	3.1%	0.08
2	1.2%	1.5%	–0.07
3	1.0%	0.9%	0.10
4	1.9%	2.6%	–0.12
5	1.9%	2.1%	–0.04
6	1.2%	1.3%	0.06
7	0.9%	0.7%	0.21
8	3.7%	5.8%	–0.09
9	1.8%	1.9%	0.01
10	2.5%	2.9%	–0.04
11	4.1%	3.9%	0.07
12	8.3%	10.5%	–0.06
13	4.0%	3.9%	0.05
14	6.2%	5.2%	0.07
15	9.3%	9.5%	–0.02
16	3.1%	3.1%	0.02
17	3.1%	3.9%	–0.09
18	3.4%	4.3%	–0.09
19	1.1%	0.9%	0.14
20	2.1%	3.2%	–0.10
21	7.0%	5.0%	0.08
22	5.1%	6.2%	–0.08
23	19.9%	13.9%	0.05
24	4.0%	3.8%	0.07
Enrolled during quarter 1	26.7%	27.8%	0.00
Enrolled during quarter 2	24.0%	22.6%	0.04
Enrolled during quarter 3	24.8%	24.0%	0.00
Enrolled during quarter 4	24.5%	25.6%	–0.05
Wagner-Peyser participant	95.1%	96.2%	–0.07
WIA/WIOA participant	4.9%	3.8%	0.07
Employed at enrollment	16.6%	16.7%	–0.02
Employed with termination notice or military separation at enrollment	0.7%	1.1%	–0.15
SNAP recipient	8.2%	1.2%	0.23
Welfare Transition Program participant	4.5%	1.8%	0.14
Reemployment Assistance claim paid	16.3%	22.8%	–0.05

Figure M.1. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 1 & 2

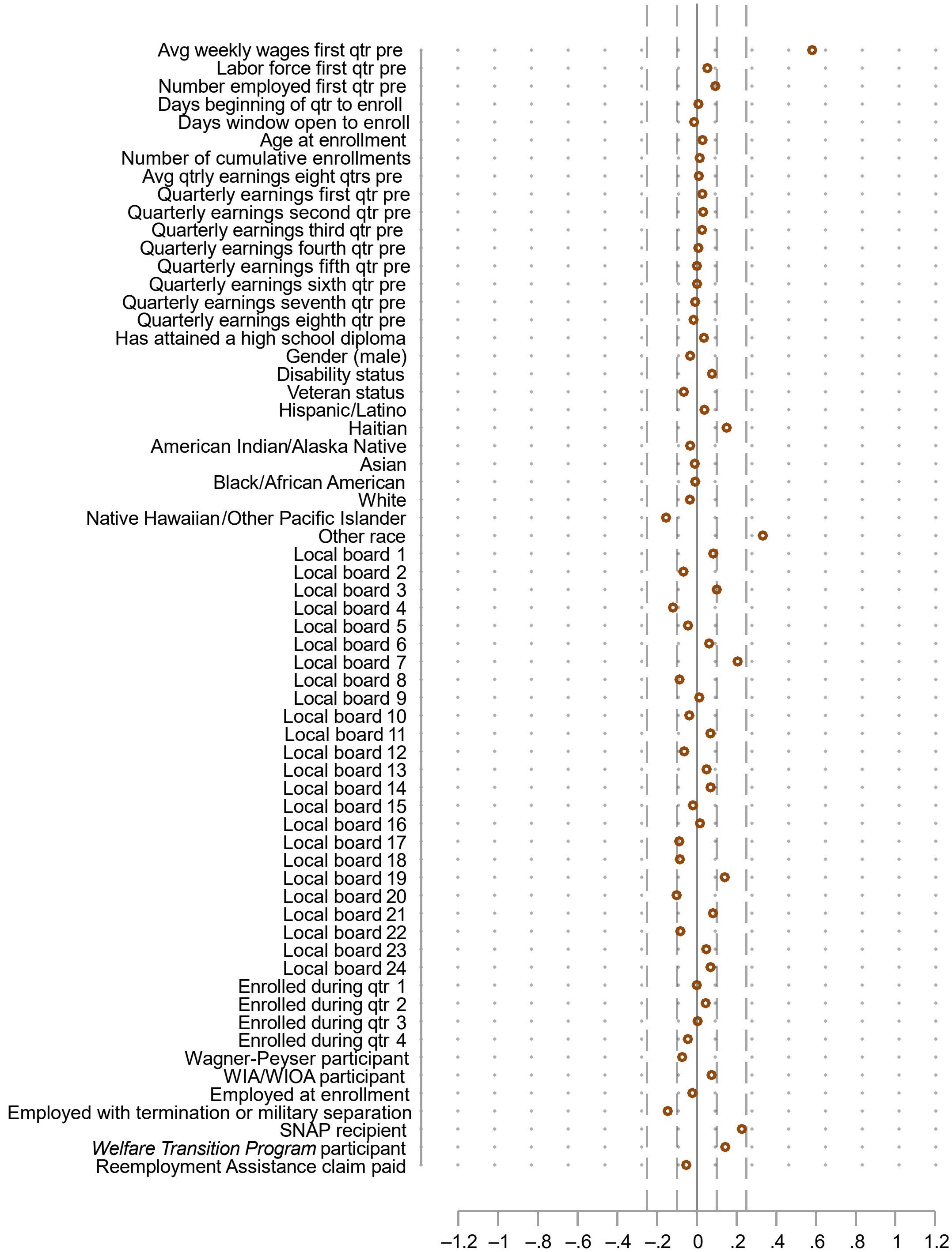


Figure M.2. Pre-Program Employment Trend, Weighted, RQ 1

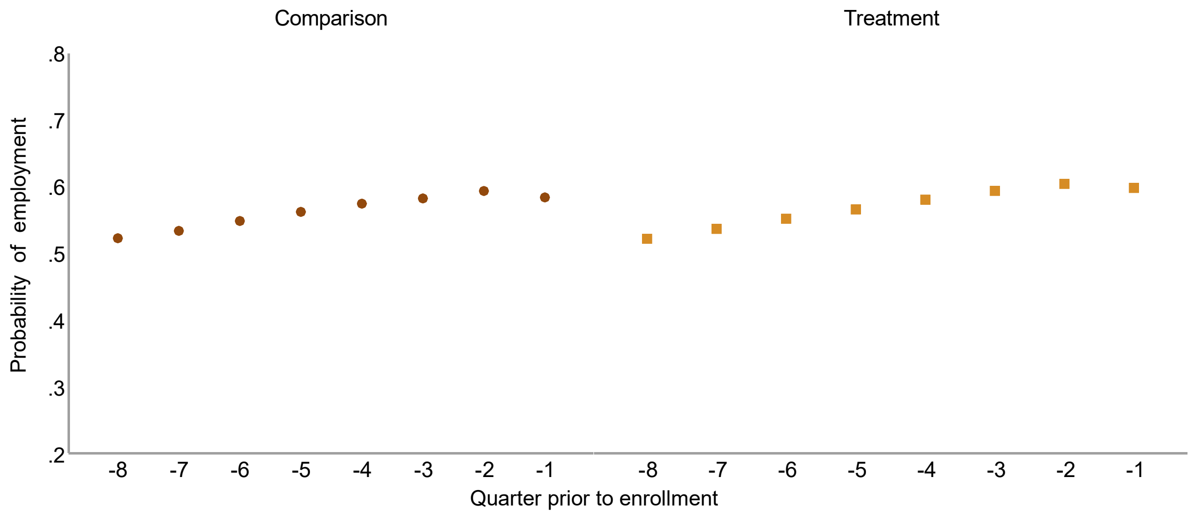


Figure M.3. Pre-Program Wages Trend, Weighted, RQ 2

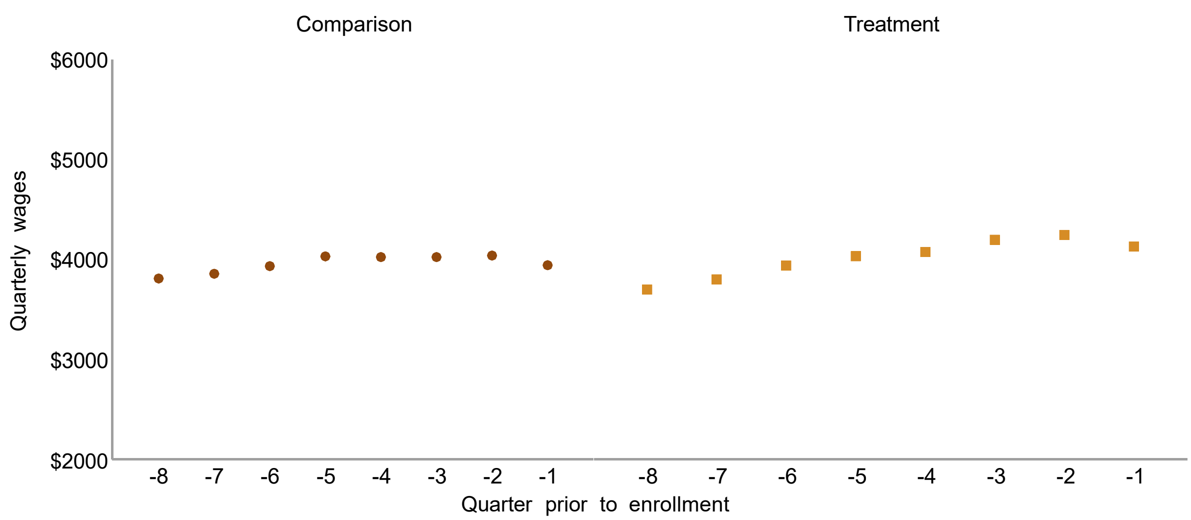
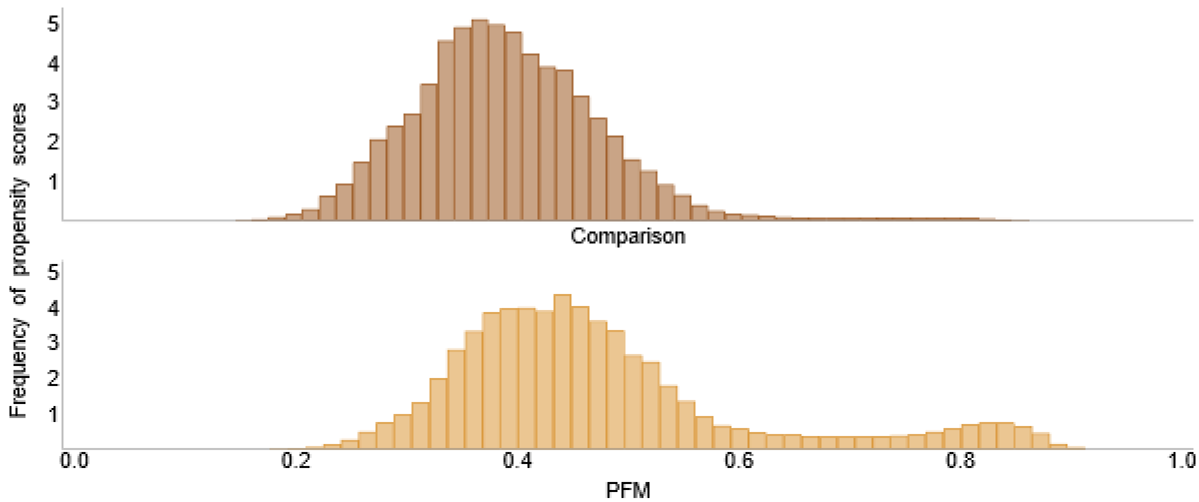


Figure M.4. Frequency and Distribution of Propensity Scores by Group, RQ 1 & 2



RESEARCH QUESTION 3

Table M.2. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3

Variable	Treatment Mean (*n* = 457,449)	Comparison Mean (*n* = 756,820)	Standardized Mean Difference
Continuous Variables			
Average weekly wages 1st quarter prior to enrollment	$886.41	$809.04	0.60
Number in labor force 1st quarter prior to enrollment	733,071	659,274	0.06
Number employed 1st quarter prior to enrollment	697,313	611,980	0.09
Days from beginning of quarter to enrollment date	45.46	43.82	0.02
Days from beginning of study window to enrollment date	480.21	487.26	–0.01
Age at enrollment	39.34	39.08	0.04
Number of cumulative enrollments into workforce services	1.98	2.03	0.00
Average quarterly earnings for eight quarters prior to enrollment	$1,787.83	$1,989.92	0.00
Quarterly earnings prior to enrollment:			
1st quarter prior	$1,138.58	$960.69	0.03
2nd quarter prior	$1,668.73	$1,590.16	0.03
3rd quarter prior	$1,834.22	$1,950.44	0.01
4th quarter prior	$1,845.08	$2,120.10	–0.01
5th quarter prior	$1,967.19	$2,284.11	–0.01
6th quarter prior	$1,989.65	$2,332.47	–0.01
7th quarter prior	$1,969.50	$2,354.42	–0.01
8th quarter prior	$1,889.73	$2,326.98	–0.02
Dichotomous Variables			
Attained a high school diploma	89.1%	89.9%	–0.03
Gender (male)	45.8%	47.2%	–0.02
Disability status	7.1%	6.8%	0.07
Veteran status	6.6%	7.8%	–0.03
Race/ethnicity:			
Hispanic/Latino	28.5%	23.9%	0.05
Haitian	3.0%	1.4%	0.17
American Indian/Alaska Native	1.4%	1.6%	–0.09
Asian	1.4%	1.5%	–0.07
Black/African American	29.7%	31.4%	–0.03
White	50.4%	53.9%	–0.04
Native Hawaiian/Other Pacific Islander	0.5%	0.6%	–0.16
Other race	0.1%	0.0%	0.13

Table M.2. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3

Variable	Treatment Mean (*n* = 457,449)	Comparison Mean (*n* = 756,820)	Standardized Mean Difference
Dichotomous Variables (continued)			
Local board:			
1	3.5%	3.6%	0.06
2	1.4%	1.6%	–0.02
3	1.2%	1.1%	0.18
4	1.7%	2.4%	–0.10
5	2.3%	2.2%	0.06
6	1.5%	1.5%	0.11
7	1.1%	0.9%	0.20
8	3.5%	6.0%	–0.11
9	1.9%	1.9%	0.01
10	2.4%	3.0%	–0.10
11	3.8%	4.1%	0.02
12	7.8%	10.5%	–0.07
13	4.4%	4.0%	0.08
14	4.8%	4.2%	0.09
15	8.4%	8.0%	–0.01
16	2.8%	3.0%	–0.04
17	3.1%	3.8%	–0.07
18	2.6%	3.7%	–0.13
19	1.0%	0.8%	0.17
20	2.2%	3.1%	–0.11
21	6.9%	5.2%	0.10
22	4.8%	5.8%	–0.10
23	23.4%	16.0%	0.05
24	3.4%	3.6%	0.01
Enrolled during quarter 1	29.5%	30.0%	–0.01
Enrolled during quarter 2	22.2%	22.4%	0.00
Enrolled during quarter 3	25.0%	23.8%	0.02
Enrolled during quarter 4	23.4%	23.7%	–0.02
Wagner-Peyser participant	94.5%	96.0%	–0.09
WIA/WIOA participant	5.5%	4.0%	0.09
Employed at enrollment	8.2%	8.4%	0.04
Employed with termination notice or military separation at enrollment	0.4%	0.6%	–0.26
SNAP recipient	9.9%	3.4%	0.16
Welfare Transition Program participant	7.0%	4.1%	0.09
Reemployment Assistance claim paid	9.2%	11.0%	–0.04

Figure M.5. Weighted Baseline Equivalence of Treatment and Comparison Samples, RQ 3

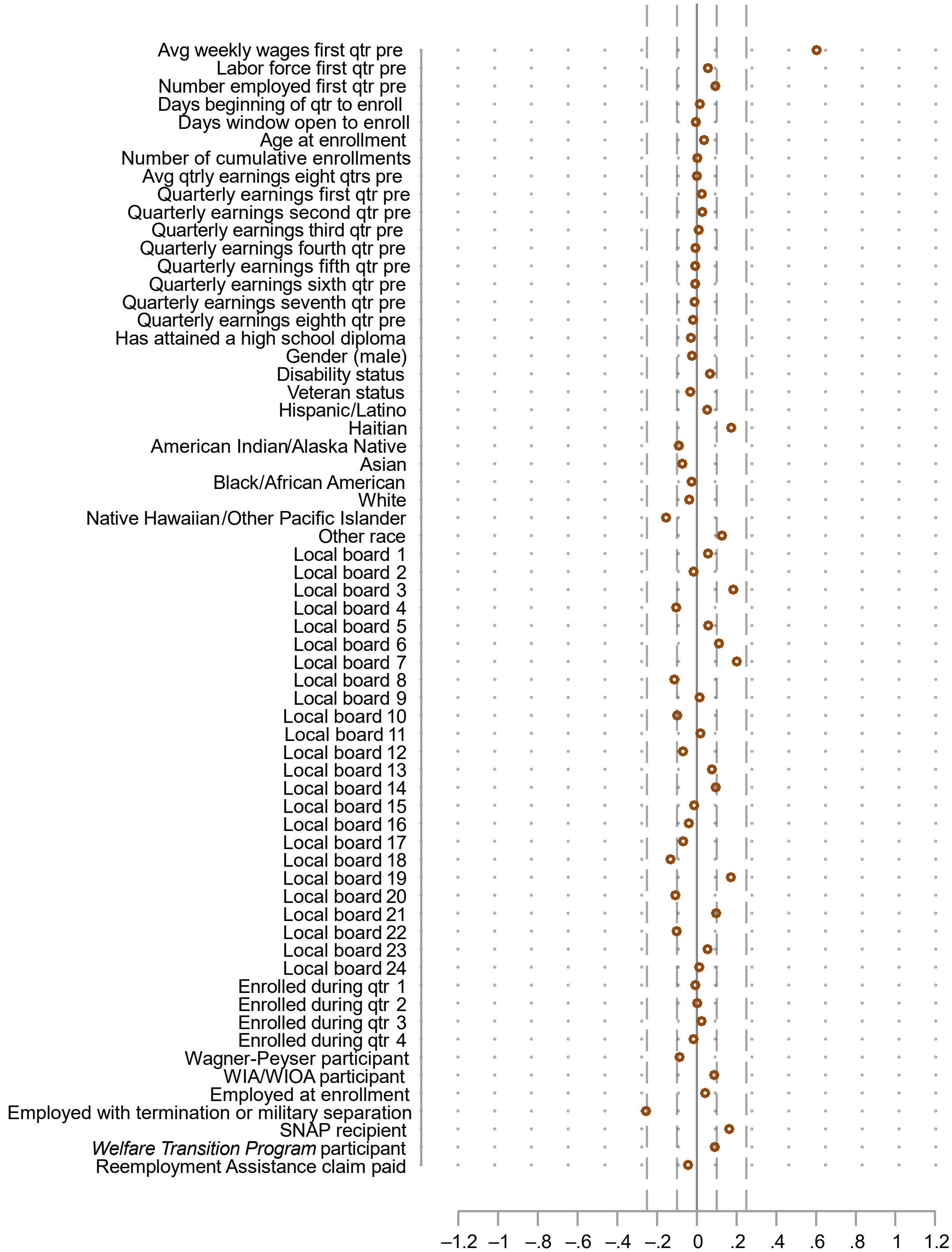
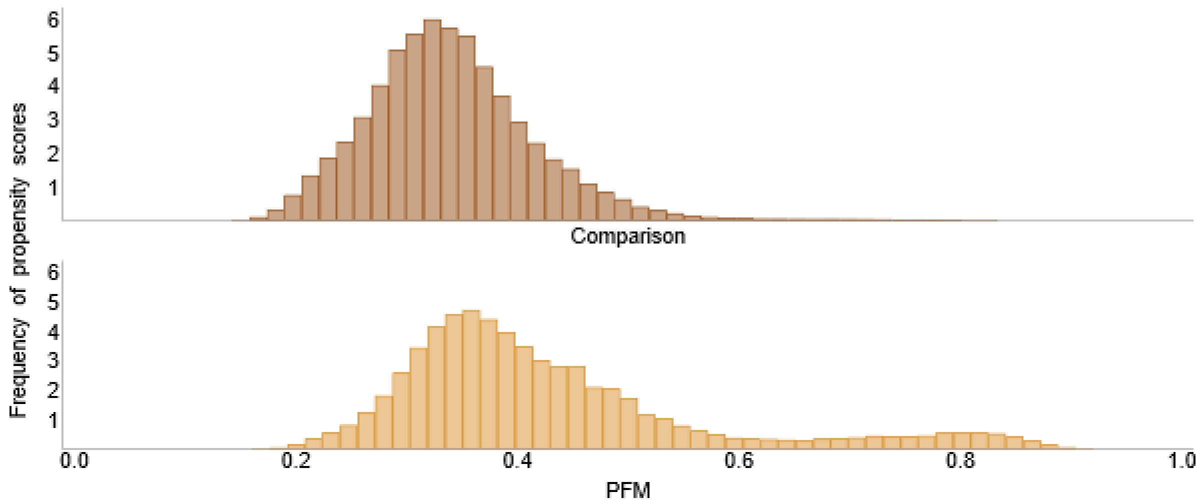


Figure M.6. Frequency and Distribution of Propensity Scores by Group, RQ 3



APPENDIX N. RESULTS: RESEARCH QUESTION 1

BENCHMARK

Table N.1. Preferred Model Results, Benchmark, RQ1

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.16***	0.00	55.71	0.000	0.15	0.16
Quarter counter	0.05***	0.00	159.23	0.000	0.05	0.05
Post variable	–0.33***	0.04	–7.68	0.000	–0.42	–0.25
Tx * qtr interaction	0.01***	0.00	21.21	0.000	0.01	0.01
Tx * post interaction	0.07***	0.01	12.49	0.000	0.06	0.08
Economic control variables						
Average weekly wages	0.00**	0.00	2.89	0.004	0.00	0.00
Number in labor force	0.00***	0.00	7.76	0.000	0.00	0.00
Number employed	0.00***	0.00	–11.22	0.000	0.00	0.00
Individual-level covariates						
High school graduate	0.41***	0.00	236.43	0.000	0.41	0.42
Gender	–0.02***	0.00	–17.65	0.000	–0.02	–0.02
Age at enrollment	0.01***	0.00	146.49	0.000	0.01	0.01
Disability status	–0.54***	0.00	–231.56	0.000	–0.54	–0.53
Veteran status	–0.44***	0.00	–202.54	0.000	–0.45	–0.44
Hispanic/Latino	0.00***	0.00	1.23	0.220	0.00	0.00
Race: Haitian	0.14***	0.00	38.44	0.000	0.13	0.15
Race: Native American	–0.19***	0.00	–44.79	0.000	–0.20	–0.19
Race: Asian	–0.03***	0.00	–6.81	0.000	–0.04	–0.02
Race: Black or African American	0.10***	0.00	62.41	0.000	0.10	0.10
Race: White	0.17***	0.00	121.51	0.000	0.17	0.17
Race: Pacific Islander	0.12***	0.01	17.40	0.000	0.11	0.13
Race: other race	–0.60***	0.03	–21.70	0.000	–0.65	–0.54
Cumulative enrollments	0.15***	0.00	386.76	0.000	0.15	0.15
Employed at enrollment	0.91***	0.00	635.23	0.000	0.91	0.91
Military separation	0.76***	0.01	136.58	0.000	0.75	0.77
SNAP recipient	–0.45***	0.00	–154.92	0.000	–0.46	–0.45
Welfare Transition	–0.48***	0.00	–160.17	0.000	–0.48	–0.47
Reemployment Assistance	1.74***	0.00	1079.29	0.000	1.74	1.74
Days from beginning of quarter to enrollment	0.00***	0.00	–31.40	0.000	0.00	0.00
Days from beginning of study window to enrollment	0.00***	0.00	69.17	0.000	0.00	0.00
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.02***	0.00	11.51	0.000	0.02	0.02
Q2 – October–December	–0.02***	0.00	–12.09	0.000	–0.02	–0.02
Q3 – January–March	–0.04***	0.00	–21.47	0.000	–0.04	–0.03
Program type (reference = WIOA)						
Wagner-Peyser	0.07***	0.00	30.52	0.000	0.07	0.08

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Local board variables (reference = local board 24)						
Local board 1	−0.29***	0.02	−14.15	0.000	−0.33	−0.25
Local board 2	−0.37***	0.03	−14.24	0.000	−0.43	−0.32
Local board 3	−0.44***	0.03	−14.27	0.000	−0.50	−0.38
Local board 4	−0.11***	0.03	-3.87	0.000	−0.16	−0.05
Local board 5	−0.24***	0.02	−10.45	0.000	−0.28	−0.19
Local board 6	−0.48***	0.03	−15.73	0.000	−0.54	−0.42
Local board 7	−0.33***	0.03	−10.73	0.000	−0.39	−0.27
Local board 8	−0.07***	0.01	−6.26	0.000	−0.10	−0.05
Local board 9	−0.18***	0.02	−7.35	0.000	−0.23	−0.13
Local board 10	−0.25***	0.02	−11.51	0.000	−0.29	−0.20
Local board 11	−0.08***	0.02	−4.74	0.000	−0.11	−0.05
Local board 12	0.08*	0.04	2.02	0.043	0.00	0.16
Local board 13	−0.30***	0.02	−16.72	0.000	−0.33	−0.26
Local board 14	0.22***	0.01	33.01	0.000	0.21	0.23
Local board 15	0.05***	0.01	5.62	0.000	0.03	0.06
Local board 16	−0.10***	0.02	−5.84	0.000	−0.13	−0.06
Local board 17	−0.18***	0.02	−10.81	0.000	−0.21	−0.15
Local board 18	0.14***	0.01	10.31	0.000	0.12	0.17
Local board 19	−0.13***	0.03	−4.43	0.000	−0.18	−0.07
Local board 20	−0.19***	0.02	−10.89	0.000	−0.22	−0.16
Local board 21	0.09***	0.01	10.14	0.000	0.07	0.11
Local board 22	0.04	0.02	1.58	0.114	−0.01	0.09
Local board 23	−0.35***	0.05	−7.59	0.000	−0.45	−0.26
Policy variable						
WIOA implementation[146]	−0.06***	0.00	−26.37	0.000	−0.06	−0.06
Constant	0.44	0.00	89.39	0.000	0.43	0.45

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

[146] To control for the WIOA policy implementation, we use the date that the legislation was mandated: July 1, 2016.

Table N.2. Model-Based Predicted Employment Probabilities

Time	Treatment Group	Comparison Group
Quarter prior to enrollment		
8	0.43	0.40
7	0.44	0.42
6	0.45	0.43
5	0.47	0.44
4	0.48	0.45
3	0.50	0.46
2	0.51	0.47
1	0.52	0.49
Fourth quarter after enrollment	0.59	0.54

SENSITIVITY STUDIES

Table N.3. Preferred Model Results, OLS, RQ1

Variable	β	Standard Error	*t*	*p*-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.03***	0.00	57.18	0.000	0.03	0.04
Quarter counter	0.01***	0.00	159.04	0.000	0.01	0.01
Post variable	–0.07***	0.01	–7.27	0.000	–0.09	–0.05
Tx * qtr interaction	0.00***	0.00	20.89	0.000	0.00	0.00
Tx * post interaction	0.01***	0.00	12.30	0.000	0.01	0.02
Economic control variables						
Average weekly wages	0.00*	0.00	2.28	0.023	0.00	0.00
Number in labor force	0.00***	0.00	9.19	0.000	0.00	0.00
Number employed	0.00***	0.00	–13.20	0.000	0.00	0.00
Individual-level covariates						
High school graduate	0.09***	0.00	246.45	0.000	0.09	0.09
Gender	0.00***	0.00	–17.73	0.000	0.00	0.00
Age at enrollment	0.00***	0.00	150.59	0.000	0.00	0.00
Disability status	–0.12***	0.00	–236.98	0.000	–0.12	–0.12
Veteran status	–0.09***	0.00	–201.88	0.000	–0.10	–0.09
Hispanic/Latino	0.00~	0.00	1.84	0.066	0.00	0.00
Race: Haitian	0.03***	0.00	38.84	0.000	0.03	0.03
Race: Native American	–0.04***	0.00	–45.47	0.000	–0.04	–0.04
Race: Asian	–0.01***	0.00	–5.63	0.000	–0.01	0.00
Race: Black or African American	0.02***	0.00	64.00	0.000	0.02	0.02
Race: White	0.04***	0.00	122.06	0.000	0.04	0.04
Race: Pacific Islander	0.03***	0.00	18.29	0.000	0.02	0.03
Race: other race	–0.13***	0.01	–24.36	0.000	–0.14	–0.12
Cumulative enrollments	0.03***	0.00	400.14	0.000	0.03	0.03
Employed at enrollment	0.20***	0.00	688.24	0.000	0.20	0.20
Military separation	0.15***	0.00	148.07	0.000	0.15	0.15

Variable	β	Standard Error	*t*	*p*-value	95% Confidence Interval	
SNAP recipient	–0.10***	0.00	–160.37	0.000	–0.10	–0.10
Welfare Transition	–0.11***	0.00	–165.95	0.000	–0.11	–0.11
Reemployment Assistance	0.34***	0.00	1398.11	0.000	0.34	0.34
Days from beginning of quarter to enrollment	0.00***	0.00	–32.75	0.000	0.00	0.00
Days from beginning of study window to enrollment	0.00***	0.00	71.70	0.000	0.00	0.00
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.00***	0.00	12.44	0.000	0.00	0.01
Q2 – October–December	0.00***	0.00	–11.28	0.000	0.00	0.00
Q3 – January–March	–0.01***	0.00	–20.36	0.000	–0.01	–0.01
Program type (reference = Wagner-Peyser)						
WIOA	–0.01***	0.00	–28.75	0.000	–0.02	–0.01
Local board variables (reference = local board 12)						
Local board 1	–0.08***	0.01	–6.43	0.000	–0.11	–0.06
Local board 2	–0.10***	0.01	–7.05	0.000	–0.12	–0.07
Local board 3	–0.11***	0.01	–7.63	0.000	–0.14	–0.08
Local board 4	–0.04**	0.01	–2.83	0.005	–0.07	–0.01
Local board 5	–0.07***	0.01	–5.20	0.000	–0.09	–0.04
Local board 6	–0.12***	0.01	–8.28	0.000	–0.15	–0.09
Local board 7	–0.09***	0.01	–6.03	0.000	–0.12	–0.06
Local board 8	–0.03***	0.01	–5.28	0.000	–0.04	–0.02
Local board 9	–0.06***	0.01	–4.18	0.000	–0.08	–0.03
Local board 10	–0.07***	0.01	–5.53	0.000	–0.10	–0.05
Local board 11	–0.03**	0.01	–2.96	0.003	–0.06	–0.01
Local board 13	–0.08***	0.01	–6.75	0.000	–0.10	–0.06
Local board 14	0.03**	0.01	2.87	0.004	0.01	0.05
Local board 15	–0.01	0.01	–1.09	0.275	–0.02	0.01
Local board 16	–0.04**	0.01	–3.30	0.001	–0.06	–0.02
Local board 17	–0.06***	0.01	–4.76	0.000	–0.08	–0.03
Local board 18	0.01	0.01	1.22	0.221	–0.01	0.04
Local board 19	–0.05**	0.01	–3.24	0.001	–0.07	–0.02
Local board 20	–0.06***	0.01	–4.87	0.000	–0.08	–0.03
Local board 21	0.00	0.01	0.27	0.784	–0.01	0.02
Local board 22	–0.01**	0.00	–2.68	0.007	–0.01	0.00
Local board 23	–0.09***	0.00	–52.88	0.000	–0.10	–0.09
Local board 24	–0.02*	0.01	–2.19	0.029	–0.03	0.00
Policy variable						
WIOA implementation	–0.01***	0.00	–27.62	0.000	–0.01	–0.01
Constant	0.59***	0.00	555.50	0.000	0.58	0.59

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table N.4. *Preferred Model Results, No Covariates, RQ1*

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.06***	0.00	30.92	0.000	0.05	0.06
Quarter counter	0.04***	0.00	146.66	0.000	0.04	0.04
Post variable	–0.29***	0.00	–95.37	0.000	–0.30	–0.29
Tx * qtr interaction	0.01***	0.00	19.62	0.000	0.01	0.01
Tx * post interaction	0.06***	0.00	13.02	0.000	0.05	0.07
Constant	0.40***	0.00	338.84	0.000	0.39	0.40

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table N.5. *Preferred Model Results, No Weighting, RQ1*

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.37***	0.00	131.41	0.000	0.36	0.37
Quarter counter	0.04***	0.00	127.05	0.000	0.04	0.04
Post variable	–0.71***	0.04	–16.86	0.000	–0.79	–0.63
Tx * qtr interaction	0.03***	0.00	71.27	0.000	0.03	0.03
Tx * post interaction	–0.09***	0.00	–17.13	0.000	–0.10	–0.08
Economic control variables						
Average weekly wages	0.00***	0.00	3.52	0.000	0.00	0.00
Number in labor force	0.00***	0.00	31.65	0.000	0.00	0.00
Number employed	0.00***	0.00	–38.52	0.000	0.00	0.00
Individual-level covariates						
High school graduate	0.40***	0.00	237.12	0.000	0.40	0.41
Gender	–0.02***	0.00	–18.63	0.000	–0.02	–0.02
Age at enrollment	0.01***	0.00	161.82	0.000	0.01	0.01
Disability status	–0.53***	0.00	–237.66	0.000	–0.53	–0.52
Veteran status	–0.44***	0.00	–214.40	0.000	–0.44	–0.44
Hispanic/Latino	0.00**	0.00	–3.28	0.001	–0.01	0.00
Race: Haitian	0.13***	0.00	37.43	0.000	0.12	0.14
Race: Native American	–0.20***	0.00	–47.96	0.000	–0.21	–0.19
Race: Asian	–0.03***	0.00	–8.16	0.000	–0.04	–0.03
Race: Black or African American	0.09***	0.00	58.20	0.000	0.09	0.09
Race: White	0.17***	0.00	122.15	0.000	0.16	0.17
Race: Pacific Islander	0.12***	0.01	18.89	0.000	0.11	0.14
Race: other race	–0.58***	0.03	–20.37	0.000	–0.63	–0.52
Cumulative enrollments	0.14***	0.00	412.33	0.000	0.14	0.14
Employed at enrollment	0.91***	0.00	668.80	0.000	0.91	0.92
Military separation	0.74***	0.01	140.33	0.000	0.73	0.75
SNAP recipient	–0.40***	0.00	–164.26	0.000	–0.40	–0.39
Welfare Transition	–0.48***	0.00	–168.52	0.000	–0.48	–0.47
Reemployment Assistance	1.73***	0.00	1173.47	0.000	1.73	1.74
Days from beginning of quarter to enrollment	0.00***	0.00	–32.34	0.000	0.00	0.00

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Days from beginning of study window to enrollment	0.00***	0.00	79.21	0.000	0.00	0.00
Quarter of enrollment (Reference = Q4 – April–June)						
Q1 – July–September	0.02***	0.00	12.77	0.000	0.02	0.02
Q2 – October–December	–0.03***	0.00	–18.82	0.000	–0.03	–0.03
Q3 – January–March	–0.04***	0.00	–22.91	0.000	–0.04	–0.03
Program type (reference = WIOA)						
Wagner-Peyser	0.06***	0.00	27.17	0.000	0.06	0.07
Local board variables (reference = local board 24)						
Local board 1	–0.03	0.02	–1.35	0.178	–0.07	0.01
Local board 2	–0.05*	0.03	–2.03	0.043	–0.10	0.00
Local board 3	–0.07*	0.03	–2.47	0.014	–0.13	–0.02
Local board 4	0.22***	0.03	8.15	0.000	0.16	0.27
Local board 5	0.04*	0.02	2.00	0.046	0.00	0.09
Local board 6	–0.12***	0.03	–4.06	0.000	–0.18	–0.06
Local board 7	0.03	0.03	1.12	0.262	–0.02	0.09
Local board 8	–0.21***	0.01	–18.99	0.000	–0.23	–0.19
Local board 9	0.12***	0.02	5.11	0.000	0.08	0.17
Local board 10	–0.01	0.02	–0.34	0.735	–0.05	0.03
Local board 11	0.11***	0.02	7.09	0.000	0.08	0.14
Local board 12	–0.37***	0.04	–9.85	0.000	–0.45	–0.30
Local board 13	–0.09***	0.02	–5.47	0.000	–0.13	–0.06
Local board 14	0.28***	0.01	44.13	0.000	0.27	0.29
Local board 15	–0.02**	0.01	–2.86	0.004	–0.04	–0.01
Local board 16	0.08***	0.02	5.00	0.000	0.05	0.11
Local board 17	0.01	0.02	0.74	0.458	–0.02	0.04
Local board 18	0.30***	0.01	22.83	0.000	0.28	0.33
Local board 19	0.22***	0.03	7.73	0.000	0.16	0.27
Local board 20	0.01	0.02	0.83	0.405	–0.02	0.05
Local board 21	0.00	0.01	–0.33	0.740	–0.02	0.01
Local board 22	–0.24***	0.02	–9.97	0.000	–0.28	–0.19
Local board 23	–0.92***	0.05	–20.48	0.000	–1.01	–0.84
Policy variable						
WIOA implementation	–0.07***	0.00	–33.41	0.000	–0.08	–0.07
Constant	0.40***	0.00	82.96	0.000	0.39	0.41

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table N.6. Preferred Model Results, Contemporaneous Comparison Group, RQ1[147]

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.17***	0.00	76.05	0.000	0.16	0.17
Quarter counter	0.01***	0.00	44.01	0.000	0.01	0.02
Post variable	–0.18***	0.00	–51.53	0.000	–0.19	–0.18
Tx * qtr interaction	0.03***	0.00	61.23	0.000	0.03	0.03
Tx * post interaction	0.01*	0.01	1.98	0.048	0.00	0.02
Individual-level covariates						
High school graduate	0.50***	0.00	254.51	0.000	0.50	0.51
Gender	–0.03***	0.00	–30.15	0.000	–0.04	–0.03
Age at enrollment	–0.01***	0.00	–174.59	0.000	–0.01	–0.01
Disability status	–0.62***	0.00	–225.89	0.000	–0.63	–0.62
Veteran status	–0.40***	0.00	–148.25	0.000	–0.40	–0.39
Hispanic/Latino	0.03***	0.00	16.89	0.000	0.02	0.03
Race: Haitian	0.10***	0.00	23.80	0.000	0.10	0.11
Race: Native American	–0.19***	0.01	–38.80	0.000	–0.20	–0.18
Race: Asian	0.04***	0.00	7.43	0.000	0.03	0.04
Race: Black or African American	0.22***	0.00	116.35	0.000	0.21	0.22
Race: White	0.17***	0.00	109.48	0.000	0.17	0.17
Race: Pacific Islander	0.06***	0.01	7.54	0.000	0.04	0.07
Race: other race	–0.55***	0.03	–19.12	0.000	–0.60	–0.49
Cumulative enrollments	0.13***	0.00	230.01	0.000	0.13	0.13
Employed at enrollment	0.79***	0.00	474.18	0.000	0.79	0.79
Military separation	0.73***	0.01	114.56	0.000	0.71	0.74
SNAP recipient	–0.58***	0.00	–175.22	0.000	–0.59	–0.57
Welfare Transition	–0.52***	0.00	–142.61	0.000	–0.53	–0.51
Reemployment Assistance	1.05***	0.00	639.43	0.000	1.05	1.06
Days from beginning of quarter to Enrollment	–0.01***	0.00	–5.16	0.000	–0.02	–0.01
Days from beginning of study Window to enrollment	0.01***	0.00	5.00	0.000	0.01	0.02
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	3.48***	0.71	4.93	0.000	2.10	4.87
Q2 – October–December	2.28***	0.47	4.86	0.000	1.36	3.20
Q3 – January–March	1.12***	0.23	4.79	0.000	0.66	1.57
Program type (reference = WIOA)						
Wagner-Peyser	–0.10***	0.00	–34.61	0.000	–0.10	–0.09
Local board variables (reference = local board 24)						
Local board 1	–0.40***	0.00	–92.42	0.000	–0.40	–0.39
Local board 2	–0.44***	0.01	–79.91	0.000	–0.45	–0.43
Local board 3	–0.46***	0.01	–67.02	0.000	–0.47	–0.44
Local board 4	–0.22***	0.00	–46.59	0.000	–0.23	–0.21
Local board 5	–0.22***	0.00	–44.71	0.000	–0.23	–0.21

[147] As this sensitivity study used a contemporaneous comparison group as its contrast, we did not control for economic conditions.

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
Local board 6	–0.46***	0.01	–74.66	0.000	–0.48	–0.45
Local board 7	–0.35***	0.01	–49.69	0.000	–0.36	–0.34
Local board 8	–0.09***	0.00	–22.32	0.000	–0.09	–0.08
Local board 9	–0.22***	0.00	–43.45	0.000	–0.23	–0.21
Local board 10	–0.19***	0.00	–44.35	0.000	–0.20	–0.18
Local board 11	–0.11***	0.00	–28.57	0.000	–0.12	–0.10
Local board 12	–0.04***	0.00	–11.40	0.000	–0.04	–0.03
Local board 13	–0.30***	0.00	–75.61	0.000	–0.31	–0.29
Local board 14	0.13***	0.00	37.00	0.000	0.13	0.14
Local board 15	–0.08***	0.00	–24.93	0.000	–0.09	–0.08
Local board 16	–0.10***	0.00	–23.98	0.000	–0.11	–0.09
Local board 17	–0.08***	0.00	–20.51	0.000	–0.09	–0.08
Local board 18	0.04***	0.00	11.20	0.000	0.04	0.05
Local board 19	–0.16***	0.01	–25.26	0.000	–0.17	–0.15
Local board 20	–0.13***	0.00	–30.39	0.000	–0.14	–0.13
Local board 21	0.04***	0.00	10.92	0.000	0.03	0.05
Local board 22	–0.06***	0.00	–16.06	0.000	–0.07	–0.05
Local board 23	–0.42***	0.00	–136.14	0.000	–0.42	–0.41
Policy variable						
WIOA implementation	–4.64***	0.94	–4.93	0.000	–6.48	–2.79
Constant	0.29***	0.00	208.04	0.000	0.28	0.29

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

EIGHTH QUARTER OUTCOMES

Table N.7. Preferred Model Results, Eighth Quarter Outcomes, RQ1

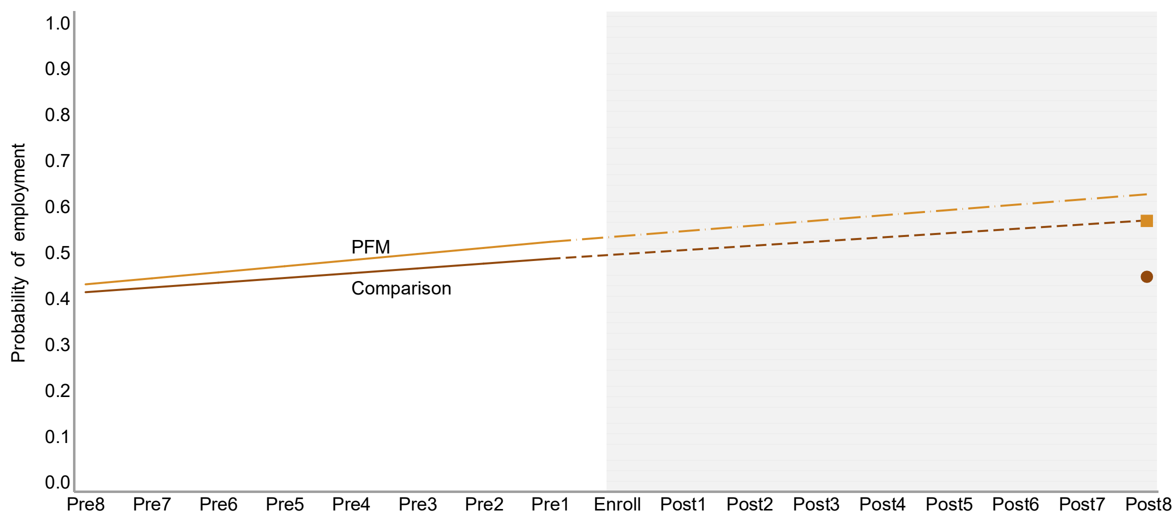
Variable	β	Standard Error	z	p-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.15***	0.00	46.28	0.000	0.14	0.15
Quarter counter	0.04***	0.00	124.83	0.000	0.04	0.04
Post variable	–0.50***	0.06	–8.93	0.000	–0.60	–0.39
Tx * qtr interaction	0.01***	0.00	21.33	0.000	0.01	0.01
Tx * post interaction	0.25***	0.01	33.41	0.000	0.24	0.27
Economic control variables						
Average weekly wages	0.00***	0.00	13.59	0.000	0.00	0.00
Number in labor force	0.00***	0.00	3.62	0.000	0.00	0.00
Number employed	0.00***	0.00	–9.20	0.000	0.00	0.00
Individual-level covariates						
High school graduate	0.38***	0.00	187.72	0.000	0.37	0.38
Gender	–0.02***	0.00	–15.91	0.000	–0.02	–0.02
Age at enrollment	0.01***	0.00	138.77	0.000	0.01	0.01
Disability status	–0.53***	0.00	–197.47	0.000	–0.53	–0.52
Veteran status	–0.44***	0.00	–175.40	0.000	–0.44	–0.43
Hispanic/Latino	0.00~	0.00	–1.81	0.070	–0.01	0.00

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Race: Haitian	0.16***	0.00	36.47	0.000	0.15	0.17
Race: Native American	−0.22***	0.01	−44.40	0.000	−0.23	−0.21
Race: Asian	−0.06***	0.00	−12.65	0.000	−0.07	−0.05
Race: Black or African American	0.04***	0.00	22.46	0.000	0.04	0.05
Race: White	0.14***	0.00	86.39	0.000	0.14	0.15
Race: Pacific Islander	0.10***	0.01	13.34	0.000	0.09	0.12
Race: other race	−0.68***	0.03	−21.98	0.000	−0.74	−0.62
Cumulative enrollments	0.14***	0.00	324.60	0.000	0.14	0.14
Employed at enrollment	0.92***	0.00	551.01	0.000	0.92	0.92
Military separation	0.77***	0.01	122.49	0.000	0.76	0.79
SNAP recipient	−0.47***	0.00	−129.18	0.000	−0.47	−0.46
Welfare Transition	−0.49***	0.00	−129.03	0.000	−0.50	−0.49
Reemployment Assistance	1.68***	0.00	923.39	0.000	1.68	1.68
Days from beginning of quarter to Enrollment	0.00***	0.00	−38.35	0.000	0.00	0.00
Days from beginning of study Window to enrollment	0.00***	0.00	80.46	0.000	0.00	0.00
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.01***	0.00	6.77	0.000	0.01	0.02
Q2 – October–December	−0.04***	0.00	−21.91	0.000	−0.05	−0.04
Q3 – January–March	−0.06***	0.00	−23.20	0.000	−0.06	−0.05
Program type (reference = WIOA)						
Wagner-Peyser	0.08***	0.00	29.66	0.000	0.08	0.09
Local board variables (reference = local board 24)						
Local board 1	−0.44***	0.03	−16.65	0.000	−0.49	−0.39
Local board 2	−0.58***	0.03	−17.20	0.000	−0.65	−0.51
Local board 3	−0.66***	0.04	−16.60	0.000	−0.74	−0.58
Local board 4	−0.29***	0.04	−8.28	0.000	−0.36	−0.22
Local board 5	−0.42***	0.03	−14.36	0.000	−0.47	−0.36
Local board 6	−0.70***	0.04	−17.79	0.000	−0.77	−0.62
Local board 7	−0.58***	0.04	−14.66	0.000	−0.66	−0.50
Local board 8	−0.01	0.02	−0.34	0.737	−0.03	0.02
Local board 9	−0.40***	0.03	−12.47	0.000	−0.46	−0.34
Local board 10	−0.41***	0.03	−14.91	0.000	−0.46	−0.36
Local board 11	−0.18***	0.02	−8.77	0.000	−0.23	−0.14
Local board 12	0.44***	0.05	8.72	0.000	0.34	0.54
Local board 13	−0.48***	0.02	−21.27	0.000	−0.52	−0.44
Local board 14	0.14***	0.01	17.29	0.000	0.12	0.15
Local board 15	0.06***	0.01	5.65	0.000	0.04	0.08
Local board 16	−0.23***	0.02	−10.86	0.000	−0.27	−0.19
Local board 17	−0.30***	0.02	−14.30	0.000	−0.34	−0.26
Local board 18	0.02	0.02	1.42	0.157	−0.01	0.06
Local board 19	−0.35***	0.04	−9.42	0.000	−0.42	−0.28
Local board 20	−0.32***	0.02	−14.45	0.000	−0.37	−0.28

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
Local board 21	0.12***	0.01	10.72	0.000	0.10	0.14
Local board 22	0.23***	0.03	7.21	0.000	0.17	0.29
Local board 23	0.01	0.06	0.24	0.807	–0.10	0.13
Policy variable						
WIOA implementation	–0.06***	0.00	–24.08	0.000	–0.07	–0.06
Constant	0.41***	0.01	68.95	0.000	0.40	0.43

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Figure N.1. Pre- to Post-Program Change in Probability of Employment, Eight Quarters Post-Enrollment



APPENDIX O. RESULTS: RESEARCH QUESTION 2

BENCHMARK

Table O.1. Preferred Model Results, Benchmark, RQ2

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.05***	0.00	35.29	0.000	0.05	0.06
Quarter counter	–0.01***	0.00	–56.18	0.000	–0.01	–0.01
Post variable	–0.01	0.02	–0.53	0.594	–0.06	0.03
Tx * qtr interaction	0.01***	0.00	33.30	0.000	0.01	0.01
Tx * post interaction	–0.03***	0.00	–10.61	0.000	–0.03	–0.02
Economic control variables						
Average weekly wages	0.00***	0.00	17.44	0.000	0.00	0.00
Number in labor force	0.00	0.00	–0.72	0.473	0.00	0.00
Number employed	0.00~	0.00	–1.75	0.080	0.00	0.00
Individual-level covariates						
High school graduate	0.32***	0.00	266.23	0.000	0.32	0.32
Gender	0.17***	0.00	298.29	0.000	0.17	0.17
Age at enrollment	0.02***	0.00	675.17	0.000	0.02	0.02
Disability status	–0.17***	0.00	–111.88	0.000	–0.17	–0.16
Veteran status	0.07***	0.00	56.17	0.000	0.06	0.07
Hispanic/Latino	–0.09***	0.00	–120.58	0.000	–0.09	–0.09
Race: Haitian	–0.07***	0.00	–40.65	0.000	–0.07	–0.07
Race: Native American	–0.06***	0.00	–26.70	0.000	–0.07	–0.06
Race: Asian	0.06***	0.00	26.66	0.000	0.06	0.07
Race: Black or African American	–0.18***	0.00	–194.41	0.000	–0.18	–0.18
Race: White	0.05***	0.00	56.94	0.000	0.05	0.05
Race: Pacific Islander	0.01*	0.00	2.31	0.021	0.00	0.02
Race: other race	–0.01	0.02	–0.66	0.508	–0.05	0.02
Cumulative enrollments	–0.09***	0.00	–484.58	0.000	–0.09	–0.09
Employed at enrollment	0.12***	0.00	194.23	0.000	0.12	0.12
Military separation	0.24***	0.00	112.13	0.000	0.23	0.24
SNAP recipient	–0.28***	0.00	–168.75	0.000	–0.28	–0.28
Welfare Transition	–0.21***	0.00	–117.19	0.000	–0.21	–0.20
Reemployment Assistance	0.22***	0.00	347.46	0.000	0.21	0.22
Days from beginning of quarter to enrollment	0.00***	0.00	–14.94	0.000	0.00	0.00
Days from beginning of study window to enrollment	0.00***	0.00	31.10	0.000	0.00	0.00
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.01***	0.00	13.47	0.000	0.01	0.01
Q2 – October–December	0.02***	0.00	19.25	0.000	0.02	0.02
Q3 – January–March	0.02***	0.00	18.63	0.000	0.02	0.02
Program type (reference = WIOA)						
Wagner-Peyser	–0.11***	0.00	–89.05	0.000	–0.11	–0.11

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Local board variables (Reference = Local board 24)						
Local board 1	–0.13***	0.01	–11.84	0.000	–0.16	–0.11
Local board 2	–0.16***	0.01	–11.20	0.000	–0.19	–0.13
Local board 3	–0.16***	0.02	–9.44	0.000	–0.19	–0.13
Local board 4	–0.12***	0.02	–7.94	0.000	–0.15	–0.09
Local board 5	–0.07***	0.01	–5.53	0.000	–0.09	–0.04
Local board 6	–0.10***	0.02	–5.72	0.000	–0.13	–0.06
Local board 7	–0.22***	0.02	–12.83	0.000	–0.25	–0.18
Local board 8	0.08***	0.01	11.87	0.000	0.06	0.09
Local board 9	–0.08***	0.01	–5.83	0.000	–0.11	–0.05
Local board 10	–0.16***	0.01	–13.86	0.000	–0.19	–0.14
Local board 11	–0.05***	0.01	–5.63	0.000	–0.07	–0.03
Local board 12	0.15***	0.02	7.11	0.000	0.11	0.20
Local board 13	–0.05***	0.01	–4.72	0.000	–0.06	–0.03
Local board 14	0.15***	0.00	40.47	0.000	0.14	0.15
Local board 15	0.13***	0.00	28.37	0.000	0.12	0.14
Local board 16	–0.02*	0.01	–2.03	0.042	–0.04	0.00
Local board 17	–0.10***	0.01	–11.57	0.000	–0.12	–0.09
Local board 18	0.00	0.01	0.50	0.619	–0.01	0.02
Local board 19	–0.14***	0.02	–8.88	0.000	–0.17	–0.11
Local board 20	–0.08***	0.01	–8.47	0.000	–0.10	–0.06
Local board 21	0.14***	0.00	28.64	0.000	0.13	0.15
Local board 22	0.22***	0.01	15.94	0.000	0.19	0.25
Local board 23	0.19***	0.03	7.15	0.000	0.13	0.24
Policy variable						
WIOA implementation[148]	–0.01***	0.00	–9.42	0.000	–0.01	–0.01
Constant	8.68***	0.00	3179.01	0.000	8.67	8.68
ZINB variables						
Employed at enrollment	–0.50***	0.00	–326.90	0.000	–0.50	–0.50
Average wages prior to enrollment	0.00***	0.00	1037.75	0.000	0.00	0.00
Constant	–1.18***	0.00	–773.67	0.000	–1.18	–1.17

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

[148] To control for the WIOA policy implementation, we use the date that the legislation was mandated: July 1, 2016.

Table O.2. Model-Based Predicted Wages

Time	Treatment Group	Comparison Group
Quarter prior to enrollment		
8	$4,764.93	$4,788.75
7	$4,760.42	$4,743.80
6	$4,755.91	$4,699.28
5	$4,751.40	$4,655.18
4	$4,746.90	$4,611.49
3	$4,742.40	$4,568.21
2	$4,737.91	$4,525.34
1	$4,733.42	$4,482.86
Fourth quarter after enrollment	$ 4,711.04	$ 4,276.41

SENSITIVITY STUDIES

Table O.3. Preferred Model Results, OLS, RQ2

Variable	β	Standard Error	*t*	*p*-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	386.02***	7.76	49.72	0.000	370.80	401.24
Quarter counter	25.07***	0.84	29.76	0.000	23.42	26.72
Post variable	–1419.43***	121.96	–11.64	0.000	–1658.46	–1180.39
Tx * qtr interaction	46.98***	1.31	35.89	0.000	44.42	49.55
Tx * post interaction	–59.87***	14.15	–4.23	0.000	–87.60	–32.14
Economic control variables						
Average weekly wages	0.84***	0.06	13.20	0.000	0.72	0.97
Number in labor force	0.00***	0.00	8.01	0.000	0.00	0.00
Number employed	0.00***	0.00	–10.79	0.000	0.00	0.00
Individual-level covariates						
High school graduate	1322.77***	3.16	418.05	0.000	1316.57	1328.97
Gender	693.66***	2.95	234.80	0.000	687.87	699.45
Age at enrollment	62.73***	0.11	572.68	0.000	62.52	62.94
Disability status	–1315.73***	5.76	–228.27	0.000	–1327.02	–1304.43
Veteran status	–479.64***	6.73	–71.30	0.000	–492.82	–466.45
Hispanic/Latino	–637.30***	3.79	–168.35	0.000	–644.72	–629.88
Race: Haitian	–267.78***	6.63	–40.36	0.000	–280.78	–254.78
Race: Native American	–463.24***	9.69	–47.80	0.000	–482.24	–444.25
Race: Asian	316.29***	13.53	23.38	0.000	289.78	342.80
Race: Black or African American	–485.08***	4.06	–119.47	0.000	–493.03	–477.12
Race: White	565.35***	3.92	144.18	0.000	557.66	573.03
Race: Pacific Islander	272.24***	18.85	14.44	0.000	235.28	309.19
Race: other race	–963.76***	49.01	–19.67	0.000	–1059.81	–867.71
Cumulative enrollments	–199.66***	0.78	–254.58	0.000	–201.20	–198.13
Employed at enrollment	1415.10***	3.33	425.09	0.000	1408.57	1421.62
Military separation	2158.95***	17.88	120.72	0.000	2123.90	2194.00

Variable	β	Standard Error	*t*	*p*-value	95% Confidence Interval	
SNAP recipient	−1326.94***	4.44	−298.90	0.000	−1335.64	−1318.24
Welfare Transition	−814.43***	4.47	−182.07	0.000	−823.20	−805.66
Reemployment Assistance	3592.52***	4.37	822.25	0.000	3583.95	3601.08
Days from beginning of quarter to enrollment	−1.59***	0.06	−28.15	0.000	−1.71	−1.48
Days from beginning of study window to enrollment	0.63***	0.01	56.25	0.000	0.61	0.66
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	55.62***	4.76	11.67	0.000	46.29	64.96
Q2 – October–December	31.24***	4.50	6.94	0.000	22.42	40.06
Q3 – January–March	43.90***	4.75	9.25	0.000	34.59	53.21
Program type (reference = Wagner-Peyser)						
WIOA	784.66***	6.71	117.02	0.000	771.52	797.80
Local board variables (reference = local board 12)						
Local board 1	−639.45***	164.33	−3.89	0.000	−961.54	−317.36
Local board 2	−795.82***	179.77	−4.43	0.000	−1148.16	−443.48
Local board 3	−687.18***	192.58	−3.57	0.000	−1064.63	−309.73
Local board 4	−327.39~	183.01	−1.79	0.074	−686.09	31.31
Local board 5	−182.51	169.67	−1.08	0.282	−515.06	150.03
Local board 6	−393.02*	191.36	−2.05	0.040	−768.08	−17.97
Local board 7	−781.64***	191.82	−4.07	0.000	−1157.59	−405.69
Local board 8	0.98	78.28	0.01	0.990	−152.46	154.41
Local board 9	−215.31	176.24	−1.22	0.222	−560.73	130.12
Local board 10	−759.85***	166.49	−4.56	0.000	−1086.16	−433.54
Local board 11	−184.73	152.71	−1.21	0.226	−484.04	114.57
Local board 13	−263.35~	155.93	−1.69	0.091	−568.97	42.27
Local board 14	923.71***	122.96	7.51	0.000	682.71	1164.72
Local board 15	445.32***	89.09	5.00	0.000	270.70	619.94
Local board 16	−80.51	153.30	−0.53	0.599	−380.98	219.96
Local board 17	−402.19*	152.55	−2.64	0.008	−701.19	−103.19
Local board 18	349.01*	145.15	2.40	0.016	64.52	633.50
Local board 19	−432.83*	186.52	−2.32	0.020	−798.40	−67.25
Local board 20	−433.64*	155.36	−2.79	0.005	−738.14	−129.15
Local board 21	637.83***	88.84	7.18	0.000	463.70	811.96
Local board 22	565.00***	41.49	13.62	0.000	483.67	646.32
Local board 23	−549.32***	23.35	−23.53	0.000	−595.09	−503.56
Local board 24	−140.87	108.59	−1.30	0.195	−353.70	71.95
Policy variable						
WIOA implementation	−119.26***	6.38	−18.68	0.000	−131.77	−106.75
Constant	4048.53***	13.87	291.91	0.000	4021.35	4075.71

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table O.4. Preferred Model Results, No Covariates, RQ2

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.03***	0.00	25.37	0.000	0.03	0.03
Quarter counter	–0.01***	0.00	–60.31	0.000	–0.01	–0.01
Post variable	–0.06***	0.00	–27.21	0.000	–0.06	–0.05
Tx * qtr interaction	0.01***	0.00	27.47	0.000	0.01	0.01
Tx * post interaction	–0.02***	0.00	–7.39	0.000	–0.03	–0.02
Constant	8.80***	0.00	11000.00	0.000	8.80	8.81
ZINB variables						
Employed at enrollment	–0.49***	0.00	–317.01	0.000	–0.49	–0.49
Average wages prior to enrollment	0.00***	0.00	–1052.33	0.000	0.00	0.00
Constant	–1.23***	0.00	–776.08	0.000	–1.23	–1.23

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table O.5. Preferred Model Results, No Weighting, RQ2

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.11***	0.00	74.26	0.000	0.11	0.12
Quarter counter	–0.01***	0.00	–73.15	0.000	–0.01	–0.01
Post variable	–0.02	0.02	–0.68	0.496	–0.06	0.03
Tx * qtr interaction	0.01***	0.00	53.81	0.000	0.01	0.01
Tx * post interaction	–0.07***	0.00	–25.96	0.000	–0.08	–0.06
Economic control variables						
Average weekly wages	0.00***	0.00	18.00	0.000	0.00	0.00
Number in labor force	0.00***	0.00	5.77	0.000	0.00	0.00
Number employed	0.00***	0.00	–10.20	0.000	0.00	0.00
Individual-level covariates						
High school graduate	0.32***	0.00	277.09	0.000	0.32	0.32
Gender	0.17***	0.00	304.31	0.000	0.16	0.17
Age at enrollment	0.02***	0.00	702.37	0.000	0.02	0.02
Disability status	–0.16***	0.00	–116.11	0.000	–0.17	–0.16
Veteran status	0.07***	0.00	62.59	0.000	0.07	0.07
Hispanic/Latino	–0.09***	0.00	–124.98	0.000	–0.09	–0.09
Race: Haitian	–0.07***	0.00	–39.33	0.000	–0.08	–0.07
Race: Native American	–0.07***	0.00	–30.16	0.000	–0.07	–0.07
Race: Asian	0.06***	0.00	26.14	0.000	0.06	0.07
Race: Black or African American	–0.18***	0.00	–199.10	0.000	–0.18	–0.17
Race: White	0.05***	0.00	59.07	0.000	0.05	0.05
Race: Pacific Islander	0.01*	0.00	2.21	0.027	0.00	0.01
Race: other race	–0.01	0.02	–0.40	0.685	–0.04	0.03
Cumulative enrollments	–0.09***	0.00	–498.93	0.000	–0.09	–0.09
Employed at enrollment	0.12***	0.00	199.94	0.000	0.12	0.12
Military separation	0.24***	0.00	119.61	0.000	0.24	0.25
SNAP recipient	–0.29***	0.00	–194.95	0.000	–0.29	–0.28

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
Welfare Transition	–0.21***	0.00	–124.48	0.000	–0.22	–0.21
Reemployment Assistance	0.22***	0.00	361.93	0.000	0.21	0.22
Days from beginning of quarter to enrollment	0.00***	0.00	–15.33	0.000	0.00	0.00
Days from beginning of study window to enrollment	0.00***	0.00	35.60	0.000	0.00	0.00
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.01***	0.00	15.81	0.000	0.01	0.02
Q2 – October–December	0.02***	0.00	19.23	0.000	0.01	0.02
Q3 – January–March	0.02***	0.00	17.97	0.000	0.01	0.02
Program type (reference = WIOA)						
Wagner-Peyser	–0.12***	0.00	–97.60	0.000	–0.12	–0.11
Local board variables (reference = local board 24)						
Local board 1	–0.11***	0.01	–10.29	0.000	–0.14	–0.09
Local board 2	–0.13***	0.01	–9.46	0.000	–0.16	–0.11
Local board 3	–0.13***	0.02	–7.86	0.000	–0.16	–0.10
Local board 4	–0.10***	0.01	–6.74	0.000	–0.13	–0.07
Local board 5	–0.05***	0.01	–3.76	0.000	–0.07	–0.02
Local board 6	–0.07***	0.02	–4.33	0.000	–0.10	–0.04
Local board 7	–0.19***	0.02	–11.46	0.000	–0.22	–0.16
Local board 8	0.07***	0.01	11.05	0.000	0.06	0.08
Local board 9	–0.06***	0.01	–4.47	0.000	–0.09	–0.03
Local board 10	–0.15***	0.01	–12.73	0.000	–0.17	–0.12
Local board 11	–0.04***	0.01	–4.52	0.000	–0.06	–0.02
Local board 12	0.13***	0.02	6.29	0.000	0.09	0.17
Local board 13	–0.02*	0.01	–2.53	0.012	–0.04	–0.01
Local board 14	0.15***	0.00	42.01	0.000	0.14	0.16
Local board 15	0.13***	0.00	29.18	0.000	0.12	0.14
Local board 16	–0.01	0.01	–0.89	0.371	–0.03	0.01
Local board 17	–0.09***	0.01	–10.04	0.000	–0.11	–0.07
Local board 18	0.02*	0.01	2.07	0.038	0.00	0.03
Local board 19	–0.12***	0.02	–7.41	0.000	–0.15	–0.08
Local board 20	–0.06***	0.01	–6.75	0.000	–0.08	–0.05
Local board 21	0.14***	0.00	28.91	0.000	0.13	0.15
Local board 22	0.21***	0.01	15.37	0.000	0.18	0.23
Local board 23	0.15***	0.03	5.95	0.000	0.10	0.20
Policy variable						
WIOA implementation	–0.02***	0.00	–13.47	0.000	–0.02	–0.01
Constant	8.65***	0.00	3247.87	0.000	8.65	8.66
ZINB Variables						
Employed at enrollment	–0.50***	0.00	–342.09	0.000	–0.51	–0.50
Average wages prior to enrollment	0.00***	0.00	1090.41	0.000	0.00	0.00
Constant	–1.17***	0.00	–809.87	0.000	–1.17	–1.16

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Table O.6. Preferred Model Results, Contemporaneous Comparison Group, RQ2[149]

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	–0.09***	0.00	–75.62	0.000	–0.09	–0.09
Quarter counter	0.03***	0.00	136.12	0.000	0.03	0.03
Post variable	–0.06***	0.00	–29.39	0.000	–0.06	–0.06
Tx * qtr interaction	–0.02***	0.00	–80.95	0.000	–0.02	–0.02
Tx * post interaction	0.00	0.00	0.49	0.628	0.00	0.01
Individual-level covariates						
High school graduate	0.36***	0.00	278.94	0.000	0.36	0.36
Gender	0.17***	0.00	265.54	0.000	0.17	0.17
Age at enrollment	0.01***	0.00	443.47	0.000	0.01	0.01
Disability status	–0.19***	0.00	–104.86	0.000	–0.19	–0.19
Veteran status	0.06***	0.00	41.61	0.000	0.05	0.06
Hispanic/Latino	–0.10***	0.00	–112.05	0.000	–0.10	–0.09
Race: Haitian	–0.06***	0.00	–29.43	0.000	–0.07	–0.06
Race: Native American	–0.06***	0.00	–21.19	0.000	–0.07	–0.06
Race: Asian	0.06***	0.00	24.71	0.000	0.06	0.07
Race: Black or African American	–0.19***	0.00	–173.42	0.000	–0.19	–0.19
Race: White	0.03***	0.00	34.82	0.000	0.03	0.04
Race: Pacific Islander	0.00	0.00	–0.42	0.672	–0.01	0.01
Race: other race	–0.01	0.02	–0.41	0.684	–0.04	0.03
Cumulative enrollments	–0.07***	0.00	–256.08	0.000	–0.07	–0.07
Employed at enrollment	0.15***	0.00	204.68	0.000	0.14	0.15
Military separation	0.25***	0.00	103.86	0.000	0.24	0.25
SNAP recipient	–0.37***	0.00	–188.46	0.000	–0.37	–0.36
Welfare Transition	–0.24***	0.00	–103.40	0.000	–0.25	–0.24
Reemployment Assistance	0.18***	0.00	245.63	0.000	0.18	0.18
Days from beginning of quarter to enrollment	0.01***	0.00	9.81	0.000	0.01	0.02
Days from beginning of study window to enrollment	–0.01***	0.00	–9.97	0.000	–0.02	–0.01
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	–3.96***	0.39	–10.05	0.000	–4.73	–3.18
Q2 – October–December	–2.62***	0.26	–10.02	0.000	–3.14	–2.11
Q3 – January–March	–1.28***	0.13	–9.85	0.000	–1.53	–1.03
Program type (reference = WIOA)						
Wagner-Peyser	–0.09***	0.00	–70.58	0.000	–0.10	–0.09
Local board variables (reference = local board 24)						
Local board 1	–0.08***	0.00	–33.97	0.000	–0.08	–0.07
Local board 2	–0.09***	0.00	–33.22	0.000	–0.10	–0.09
Local board 3	–0.12***	0.00	–37.16	0.000	–0.13	–0.11
Local board 4	–0.06***	0.00	–24.49	0.000	–0.06	–0.05
Local board 5	0.02***	0.00	7.17	0.000	0.01	0.02

[149] As this sensitivity study used a contemporaneous comparison group as its contrast, we did not control for economic conditions.

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
Local board 6	–0.04***	0.00	–11.55	0.000	–0.04	–0.03
Local board 7	–0.15***	0.01	–29.47	0.000	–0.16	–0.14
Local board 8	0.05***	0.00	22.63	0.000	0.04	0.05
Local board 9	0.02***	0.00	7.72	0.000	0.02	0.03
Local board 10	–0.12***	0.00	–48.15	0.000	–0.12	–0.11
Local board 11	–0.02***	0.00	–8.18	0.000	–0.02	–0.01
Local board 12	0.02***	0.00	12.32	0.000	0.02	0.03
Local board 13	–0.01*	0.00	–2.43	0.015	–0.01	0.00
Local board 14	0.22***	0.00	106.95	0.000	0.22	0.22
Local board 15	0.15***	0.00	78.94	0.000	0.14	0.15
Local board 16	0.01**	0.00	3.06	0.002	0.00	0.01
Local board 17	–0.06***	0.00	–26.85	0.000	–0.06	–0.05
Local board 18	0.04***	0.00	18.53	0.000	0.04	0.05
Local board 19	–0.08***	0.00	–24.05	0.000	–0.09	–0.07
Local board 20	–0.04***	0.00	–18.67	0.000	–0.05	–0.04
Local board 21	0.17***	0.00	77.98	0.000	0.16	0.17
Local board 22	0.15***	0.00	76.37	0.000	0.15	0.15
Local board 23	0.08***	0.00	41.98	0.000	0.08	0.08
Policy variable						
WIOA implementation	5.27***	0.52	10.04	0.000	4.24	6.30
Constant	8.89***	0.00	11000.00	0.000	8.89	8.90
ZINB variables						
Employed at enrollment	–0.31***	0.00	–150.59	0.000	–0.31	–0.30
Average wages prior to enrollment	0.00***	0.00	–988.37	0.000	0.00	0.00
Constant	–1.43***	0.00	–685.24	0.000	–1.43	–1.43

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

EIGHTH QUARTER OUTCOMES

Table O.7. Preferred Model Results, Eighth Quarter Outcomes, RQ2

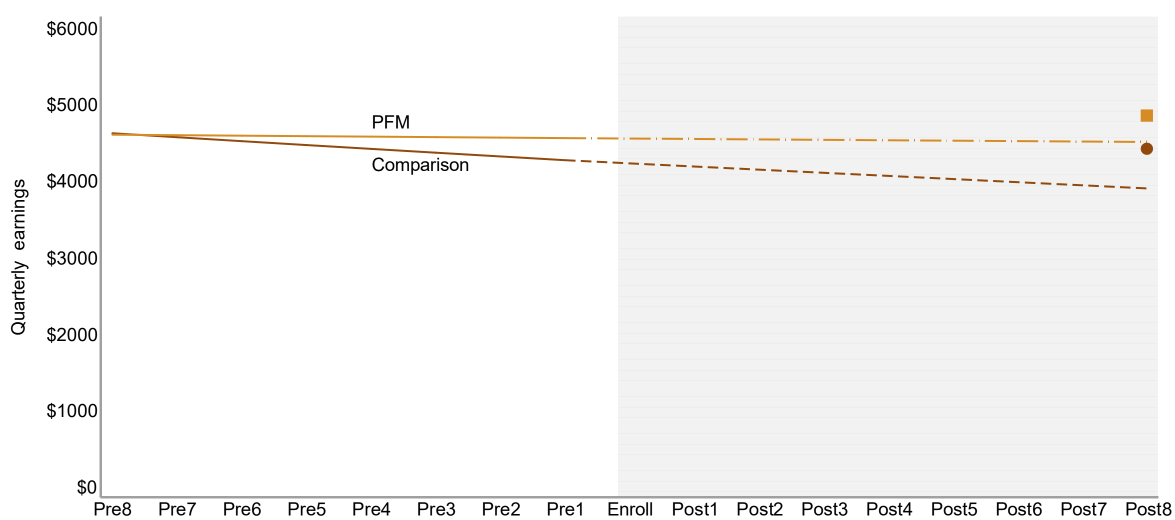
Variable	β	Standard Error	z	p-value	95% Confidence Interval	
CSITS variables						
Tx group indicator	0.07***	0.00	38.25	0.000	0.06	0.07
Quarter counter	–0.01***	0.00	–60.47	0.000	–0.01	–0.01
Post variable	0.13***	0.03	4.17	0.000	0.07	0.18
Tx * qtr interaction	0.01***	0.00	34.36	0.000	0.01	0.01
Tx * post interaction	–0.05***	0.00	–12.58	0.000	–0.06	–0.04
Economic control variables						
Average weekly wages	0.00	0.00	0.65	0.515	0.00	0.00
Number in labor force	0.00*	0.00	–2.45	0.014	0.00	0.00
Number employed	0.00	0.00	0.81	0.416	0.00	0.00
Individual-level covariates						
High school graduate	0.31***	0.00	229.99	0.000	0.31	0.32

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Gender	0.17***	0.00	257.83	0.000	0.17	0.17
Age at enrollment	0.02***	0.00	592.18	0.000	0.02	0.02
Disability status	–0.16***	0.00	–92.49	0.000	–0.16	–0.16
Veteran status	0.07***	0.00	50.04	0.000	0.07	0.07
Hispanic/Latino	–0.08***	0.00	–93.82	0.000	–0.08	–0.08
Race: Haitian	–0.07***	0.00	–35.46	0.000	–0.08	–0.07
Race: Native American	–0.07***	0.00	–26.79	0.000	–0.08	–0.07
Race: Asian	0.06***	0.00	22.71	0.000	0.06	0.07
Race: Black or African American	–0.17***	0.00	–163.96	0.000	–0.18	–0.17
Race: White	0.05***	0.00	51.79	0.000	0.05	0.05
Race: Pacific Islander	0.02***	0.00	5.87	0.000	0.02	0.03
Race: other race	0.01	0.02	0.42	0.673	–0.03	0.05
Cumulative enrollments	–0.08***	0.00	–410.31	0.000	–0.09	–0.08
Employed at enrollment	0.12***	0.00	176.63	0.000	0.12	0.13
Military separation	0.25***	0.00	104.48	0.000	0.25	0.26
SNAP recipient	–0.28***	0.00	–133.92	0.000	–0.28	–0.27
Welfare Transition	–0.20***	0.00	–90.08	0.000	–0.21	–0.20
Reemployment Assistance	0.21***	0.00	297.82	0.000	0.21	0.21
Days from beginning of quarter to enrollment	0.00***	0.00	–14.85	0.000	0.00	0.00
Days from beginning of study window to enrollment	0.00***	0.00	22.53	0.000	0.00	0.00
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.00***	0.00	3.76	0.000	0.00	0.01
Q2 – October–December	0.01***	0.00	6.58	0.000	0.00	0.01
Q3 – January–March	0.02***	0.00	16.97	0.000	0.02	0.03
Program type (reference = WIOA)						
Wagner-Peyser	–0.11***	0.00	–74.63	0.000	–0.11	–0.11
Local board variables (reference = local board 24)						
Local board 1	–0.15***	0.01	–10.76	0.000	–0.18	–0.12
Local board 2	–0.18***	0.02	–10.36	0.000	–0.22	–0.15
Local board 3	–0.20***	0.02	–9.57	0.000	–0.24	–0.16
Local board 4	–0.15***	0.02	–8.28	0.000	–0.19	–0.12
Local board 5	–0.08***	0.02	–5.21	0.000	–0.11	–0.05
Local board 6	–0.14***	0.02	–6.77	0.000	–0.18	–0.10
Local board 7	–0.26***	0.02	–12.62	0.000	–0.30	–0.22
Local board 8	0.10***	0.01	12.23	0.000	0.08	0.11
Local board 9	–0.08***	0.02	–4.67	0.000	–0.11	–0.05
Local board 10	–0.20***	0.01	–13.64	0.000	–0.23	–0.17
Local board 11	–0.08***	0.01	–6.84	0.000	–0.10	–0.05
Local board 12	0.17***	0.03	6.45	0.000	0.12	0.23
Local board 13	–0.03*	0.01	–2.34	0.019	–0.05	0.00
Local board 14	0.16***	0.00	37.84	0.000	0.15	0.17
Local board 15	0.16***	0.01	28.03	0.000	0.15	0.18
Local board 16	–0.06***	0.01	–4.90	0.000	–0.08	–0.03

Variable	β	Standard Error	z	p-value	95% Confidence Interval	
Local board 17	–0.12***	0.01	–10.99	0.000	–0.14	–0.10
Local board 18	–0.01	0.01	–0.88	0.380	–0.03	0.01
Local board 19	–0.20***	0.02	–10.12	0.000	–0.24	–0.16
Local board 20	–0.09***	0.01	–7.89	0.000	–0.12	–0.07
Local board 21	0.18***	0.01	28.92	0.000	0.16	0.19
Local board 22	0.25***	0.02	14.70	0.000	0.22	0.28
Local board 23	0.22***	0.03	6.84	0.000	0.16	0.28
Policy variable						
WIOA implementation	0.00	0.00	–0.08	0.936	0.00	0.00
Constant	8.64***	0.00	2678.53	0.000	8.64	8.65
ZINB variables						
Employed at enrollment	–0.52***	0.00	–294.89	0.000	–0.52	–0.52
Average wages prior to enrollment	0.00***	0.00	–893.60	0.000	0.00	0.00
Constant	–1.12***	0.00	–657.41	0.000	–1.12	–1.11

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

Figure O.1. *Pre- to Post-Program Change in Quarterly Wages, Eight Quarters Post-Enrollment*



APPENDIX P. RESULTS: RESEARCH QUESTION 3

BENCHMARK

Table P.1. Preferred Model Results, RQ3

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Time (Quarter)						
1	−1.06***	0.00	−248.92	0.000	−1.07	−1.05
2	−1.61***	0.00	−354.07	0.000	−1.61	−1.60
3	−2.01***	0.00	−403.48	0.000	−2.02	−2.00
4	−2.25***	0.01	−403.84	0.000	−2.26	−2.24
5	−2.47***	0.01	−388.58	0.000	−2.48	−2.45
6	−2.67***	0.01	−354.05	0.000	−2.69	−2.66
7	−2.83***	0.01	−319.44	0.000	−2.85	−2.81
8	−2.97***	0.01	−275.57	0.000	−2.99	−2.94
9	−3.12***	0.01	−232.28	0.000	−3.14	−3.09
10	−3.29***	0.02	−184.25	0.000	−3.32	−3.25
11	−3.42***	0.03	−133.62	0.000	−3.47	−3.37
Treatment group indicator	0.02**	0.01	2.86	0.004	0.01	0.03
Economic control variables						
Average weekly wages	0.00	0.00	1.61	0.107	0.00	0.00
Number in labor force	0.00***	0.00	−38.23	0.000	0.00	0.00
Number employed	0.00***	0.00	38.09	0.000	0.00	0.00
Individual-level covariates						
High school graduate	0.32***	0.00	66.83	0.000	0.31	0.33
Gender	0.06***	0.00	21.57	0.000	0.06	0.07
Age at enrollment	−0.02***	0.00	−166.77	0.000	−0.02	−0.02
Disability status	−0.47***	0.01	−75.93	0.000	−0.48	−0.45
Veteran status	−0.04***	0.01	−6.77	0.000	−0.05	−0.03
Hispanic/Latino	0.18***	0.00	47.99	0.000	0.17	0.19
Race: Haitian	0.17***	0.01	16.64	0.000	0.15	0.19
Race: Native American	−0.02~	0.01	−1.79	0.074	−0.04	0.00
Race: Asian	0.13***	0.01	10.72	0.000	0.10	0.15
Race: Black or African American	0.34***	0.00	72.66	0.000	0.33	0.34
Race: White	0.25***	0.00	60.41	0.000	0.24	0.26
Race: Pacific Islander	0.08***	0.02	4.28	0.000	0.05	0.12
Race: other race	−0.27***	0.07	−3.71	0.000	−0.42	−0.13
Cumulative enrollments	0.07***	0.00	67.97	0.000	0.07	0.07
Employed at enrollment	0.00	0.01	−0.22	0.827	−0.01	0.01
Military separation	0.05**	0.02	2.88	0.004	0.02	0.09
SNAP recipient	−0.06***	0.01	−9.19	0.000	−0.07	−0.05
Welfare Transition	−0.01	0.01	−1.34	0.182	−0.02	0.00
Reemployment Assistance	0.36***	0.00	79.45	0.000	0.36	0.37
Days from beginning of quarter to enrollment	0.00***	0.00	65.12	0.000	0.00	0.00
Days from beginning of study window to enrollment	0.00***	0.00	−4.28	0.000	0.00	0.00

Variable	β	Standard Error	*z*	*p*-value	95% Confidence Interval	
Quarter of enrollment (reference = Q4 – April–June)						
Q1 – July–September	0.05***	0.00	12.57	0.000	0.04	0.06
Q2 – October–December	–0.04***	0.00	–9.58	0.000	–0.05	–0.03
Q3 – January–March	–0.05***	0.00	–12.61	0.000	–0.06	–0.05
Program type (reference = WIOA)						
Wagner-Peyser	–0.16***	0.01	–24.10	0.000	–0.17	–0.15
Local board variables (reference = local board 24)						
Local board 1	–1.94***	0.05	–38.38	0.000	–2.04	–1.84
Local board 2	–2.30***	0.06	–35.87	0.000	–2.43	–2.17
Local board 3	–2.67***	0.08	–35.00	0.000	–2.82	–2.52
Local board 4	–2.30***	0.07	–34.24	0.000	–2.43	–2.16
Local board 5	–2.05***	0.06	–37.13	0.000	–2.15	–1.94
Local board 6	–2.72***	0.07	–36.34	0.000	–2.87	–2.57
Local board 7	–2.56***	0.08	–33.86	0.000	–2.70	–2.41
Local board 8	0.62***	0.03	21.85	0.000	0.57	0.68
Local board 9	–2.26***	0.06	–37.24	0.000	–2.38	–2.14
Local board 10	–1.83***	0.05	–34.57	0.000	–1.94	–1.73
Local board 11	–1.48***	0.04	–36.15	0.000	–1.56	–1.40
Local board 12	2.87***	0.09	30.29	0.000	2.68	3.05
Local board 13	–1.65***	0.04	–38.54	0.000	–1.74	–1.57
Local board 14	–0.71***	0.02	–42.50	0.000	–0.75	–0.68
Local board 15	0.07**	0.02	3.27	0.001	0.03	0.11
Local board 16	–1.49***	0.04	–35.84	0.000	–1.57	–1.41
Local board 17	–1.43***	0.04	–35.33	0.000	–1.51	–1.35
Local board 18	–1.13***	0.03	–33.55	0.000	–1.20	–1.07
Local board 19	–2.21***	0.07	–31.03	0.000	–2.35	–2.07
Local board 20	–1.47***	0.04	–34.26	0.000	–1.56	–1.39
Local board 21	0.33***	0.02	14.75	0.000	0.28	0.37
Local board 22	1.63***	0.06	27.22	0.000	1.52	1.75
Local board 23	3.28***	0.11	28.56	0.000	3.05	3.50
Policy variable						
WIOA implementation[150]	–0.11***	0.01	–18.16	0.000	–0.12	–0.09

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.1$.

[150] To control for the WIOA policy implementation, we use the date that the legislation was mandated: July 1, 2016.

Table P.2. Regression-Adjusted Conditional Probability Estimates for Treatment and Comparison Groups

	Proportion Employed		Proportion Remaining Unemployed	
Quarter After Enrollment	***Treatment Group***	***Comparison Group***	***Treatment Group***	***Comparison Group***
1	0.269	0.274	0.731	0.726
2	0.176	0.175	0.603	0.599
3	0.124	0.122	0.528	0.525
4	0.099	0.097	0.475	0.474
5	0.082	0.080	0.436	0.436
6	0.068	0.067	0.407	0.407
7	0.059	0.059	0.383	0.383
8	0.051	0.052	0.363	0.363
9	0.045	0.045	0.347	0.347
10	0.039	0.039	0.333	0.333
11	0.036	0.036	0.321	0.321

Table P.3. Descriptive Characteristics of Full Sample, Treatment Group, and Comparison Group

	Full Sample	Treatment	Comparison
Number in sample	1,214,269	457,449	756,820
Number employed	656,437	245,995	410,442
Percent employed	54.1%	53.8%	54.2%
Average time to employment (in quarters)	3.62	3.55	3.67

Table P.4. Life Table, Full Sample

		Number		Proportion	
Quarter After Enrollment	**Sample Size**	***Employed***	***Censored***	***Employed***	***Still Unemployed***
1	1,214,269	310,791	124,863	0.26	0.74
2	778,615	136,704	40,328	0.18	0.61
3	601,583	74,380	47,789	0.12	0.54
4	479,414	47,283	46,625	0.10	0.49
5	385,506	31,327	45,928	0.08	0.45
6	308,251	20,796	40,095	0.07	0.42
7	247,360	14,551	45,760	0.06	0.39
8	187,049	9,592	43,553	0.05	0.37
9	133,904	6,036	42,406	0.05	0.35
10	85,462	3,359	37,186	0.04	0.34
11	44,917	1,618	43,299	0.04	0.33

Note: The median lifetime, or the time at which half of the group achieves employment, occurs between quarters three and four after enrollment, at 3.72 quarters, and is represented by the dashed line. In discrete-time hazard model terminology, the proportion achieving employment is referred to as the "hazard function"; the proportion still unemployed is referred to as the "survivor function."

Table P.5. Life Table, Treatment Group

		Number		Proportion	
Quarter After Enrollment	**Sample Size**	***Employed***	***Censored***	***Employed***	***Still Unemployed***
1	457,449	122,547	44,910	0.27	0.73
2	289,992	53,939	13,499	0.19	0.60
3	222,554	25,988	18,842	0.12	0.53
4	177,724	15,924	21,027	0.09	0.48
5	140,773	10,147	21,134	0.07	0.45
6	109,492	6,684	16,267	0.06	0.42
7	86,541	4,337	16,237	0.05	0.40
8	65,967	2,967	15,139	0.05	0.38
9	47,861	1,925	15,770	0.04	0.36
10	30,166	1,059	13,230	0.04	0.35
11	15,877	478	15,399	0.03	0.34

Note: The median lifetime, or the time at which half of the group achieves employment, occurs between quarter three and four after enrollment, at 3.55 quarters, and is represented by the dashed line. In discrete-time hazard model terminology, the proportion achieving employment is referred to as the "hazard function"; the proportion still unemployed is referred to as the "survivor function."

Table P.6. Life Table, Comparison Group

		Number		Proportion	
Quarter After Enrollment	**Sample Size**	***Employed***	***Censored***	***Employed***	***Still Unemployed***
1	756,820	188,244	79,953	0.25	0.75
2	488,623	82,765	26,829	0.17	0.62
3	379,029	48,392	28,947	0.13	0.54
4	301,690	31,359	25,598	0.10	0.49
5	244,733	21,180	24,794	0.09	0.45
6	198,759	14,112	23,828	0.07	0.41
7	160,819	10,214	29,523	0.06	0.39
8	121,082	6,625	28,414	0.06	0.37
9	86,043	4,111	26,636	0.05	0.35
10	55,296	2,300	23,956	0.04	0.33
11	29,040	1,140	27,900	0.04	0.32

Note: The median lifetime, or the time at which half of the group achieves employment, occurs between quarter three and four after enrollment, at 3.79 quarters, and is represented by the dashed line. In discrete-time hazard model terminology, the proportion achieving employment is referred to as the "hazard function"; the proportion still unemployed is referred to as the "survivor function."

Figure P.1. Unadjusted Conditional Probability of Becoming Employed, Full Sample

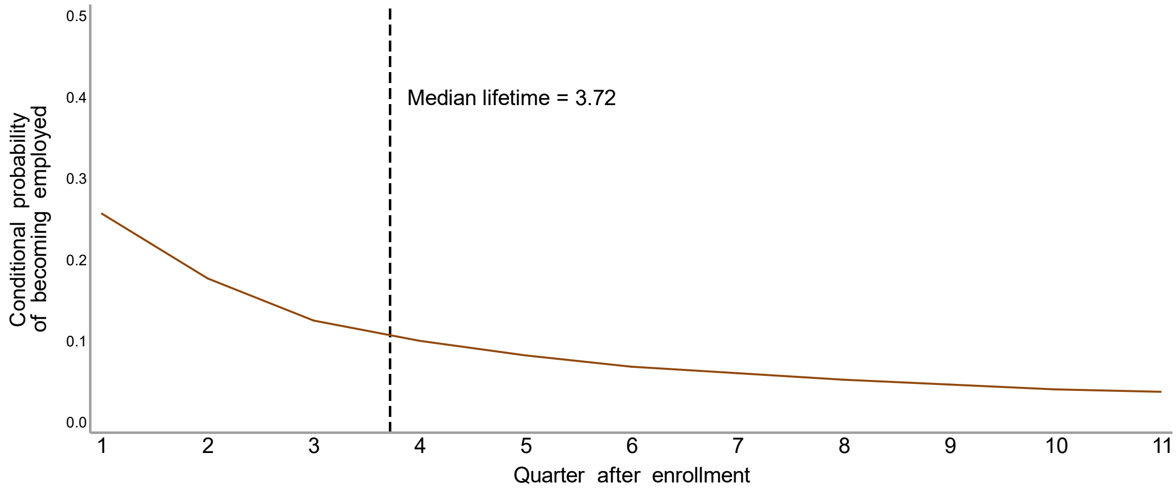


Figure P.2. Unadjusted Proportion of Individuals Who Remain Unemployed, Full Sample