Employment Programs to Support Reentry: Findings from the Reentry Project Grants Evaluation

Final Report

September 2024

Leah Shiferaw, Jonah Deutsch, Brett Fischer, Christian Geckeler, Anne Paprocki, Lea Folsom, Andrew Wiegand, Gina Lewis, and Brittany English

Submitted to:

U.S. Department of Labor
Chief Evaluation Office
200 Constitution Avenue, NW
Washington, DC 20210
Attention: Evan Murphy

Submitted by:

Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763

This page has been left blank for double-sided copying.

Employment Programs to Support Reentry: Findings from the Reentry Project Grant Evaluation

Final Report

September 2024

Leah Shiferaw, Jonah Deutsch, Brett Fischer, Christian Geckeler, Anne Paprocki, Lea Folsom, Andrew Wiegand, Gina Lewis, and Brittany English

Submitted to:
U.S. Department of Labor Chief Evaluation Office
200 Constitution Avenue, NW Washington, DC
Attention: Evan Murphy

Submitted by:
Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763

This page has been left blank for double-sided copying.

This report was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office (CEO) by Mathematica, under contracts #1605DC-18-A-0020/1605C2-20-F-00042 and #DOLQ129633249/1605DC-17-U-00106. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government. Some data used in this report were provided by the New York State Division of Criminal Justice Services (DCJS). The opinions, findings, and conclusions expressed in this publication are those of the authors and not those of DCJS. Neither New York State nor DCJS assumes liability for its contents or use thereof.

Suggested citation from this report:

Shiferaw, Leah, Jonah Deutsch, Brett Fischer, Christian Geckeler, Anne Paprocki, Lea Folsom, Andrew Wiegand, Gina Lewis, Brittany English. "Employment Programs to Support Reentry: Findings from the Reentry Project Grant Evaluation." Report submitted to the U.S. Department of Labor. Washington, DC. Mathematica, 2024. Available at https://www.dol.gov/agencies/oasp/evaluation/completedstudies/Reentry-Projects-Grant-Evaluation

Other study publications

1. Implementing Employment Programs to Support Reentry: Lessons from the Reentry Project Grants
2. DOL Evaluation Design Pre-Specification Plan: Reentry Programs
3. Advancing Employment Opportunities for Justice-Involved Individuals through Work-Based Learning: Experiences from Reentry Project Grantees
4. Adult and Young Adult Reentry Project Grants: Differences in Service Offerings and Implementation Challenges
5. Participants' Perspectives During Reentry Project Programs
6. Portrait of the Reentry Project Grantees
7. Connecting Reentry Project (RP) Participants to In-Demand Local Industries: Insights from RP Grant Programs
8. Common Indicators of Recidivism Used in Program and Policy Evaluations
9. Using Risk/Needs Assessments in Reentry Services
10. Data Privacy Seminars
11. Supporting Reentry Employment and Success: A Summary of the Evidence for Adults and Young Adults

Study publications are available at: https://www.dol.gov/agencies/oasp/evaluation/completedstudies/Reentry-Projects-Grant-Evaluation

Acknowledgements

We are thankful for the contributions of those who made this study possible. We benefitted from the valuable guidance and support provided by many individuals from the U.S. Department of Labor's (DOL's) Chief Evaluation Office throughout the study, especially Evan Murphy and Megan Lizik. We are also thankful to staff from the Reentry Employment Opportunities program within DOL's Employment and Training Administration for their support throughout the evaluation and their review of draft documents, and to staff from the Office of Policy Development and Research (OPDR), for providing data extracts for our analysis of administrative records.

We would also like to extend our most sincere appreciation to the Reentry Project grantees and their partners. We would further like to thank the state workforce and criminal justice agencies in Alabama, Florida, New Jersey, New York, Oregon, and Pennsylvania who provided data for this project. This report would not have been possible without their cooperation and insights. We would also like to extend our deep appreciation to the members of the Reentry Programs Technical Working Group, including Dr. Shawn Bushway, Dr. Harry Holzer, Debbie Mukamal, Dr. Omari Swinton, and Dr. Christy Visher for their time, commitment, and expertise. This report has benefited tremendously from their insight.

This report also benefitted from the contributions of many people at Mathematica and Social Policy Research Associates. We are thankful to all members of the study team who worked on the data collection, research, and data analysis that inform this report, including Peter Kress, Patrick Lavallee, Olivia Pham, Colleen Bullock, Mindy Wong, Sheldon Bond, Rhiannon Jones, Moriah Schwartz, and Liz Potamites. This report was improved thanks to detailed reviews by Jill Berk.

Contents

Exhibits

Executive Summary

Individuals released from incarceration often experience collateral consequences that can create barriers and impact access to education, employment, professional licensing, housing, public benefits, and other supports that help them successfully reenter society (U.S. Commission on Civil Rights 2019). Building upon prior reentry programs, the U.S. Department of Labor (DOL) awarded $243 million in Reentry Project (RP) grants between 2017 and 2019. These grants aimed to help organizations in high-crime communities implement comprehensive reentry programs to improve workforce and criminal justice outcomes for individuals with prior justice system involvement.

RP grantees included intermediary organizations that served large numbers of participants across multiple subgrantees and states as well as community-based organizations (CBOs) that served smaller numbers of participants in a single location. Grant amounts ranged from $4 million to $4.5 million for larger intermediaries and $500,000 to $1.5 million for CBOs (DOL 2018, 2019). The grants lasted 36 to 39 months and included a three-month planning period, a 24-month operational period, and a nine- to 12-month follow-up period. RP grants chose to serve either adults (individuals over age 24) or young adults (individuals ages 18 to 24) who were involved in the criminal justice system and/or who may be reentering society after incarceration. Program design varied, and DOL encouraged grantees to draw on evidence-informed or promising practices around employment-focused services.

To evaluate the RP programs, DOL's Chief Evaluation Office (CEO) contracted with Mathematica and Social Policy Research Associates (SPR) to examine the implementation and impacts of the programs funded during the 2017, 2018, and 2019 grant cycles. This report presents results from the RP impact study. Chapter 1 provides a detailed background on the RP grants and the guiding research questions for the impact study. Chapter 2 summarizes key findings from the RP implementation study that helped inform the study team's interpretation of the impact study findings. (For the full set of implementation findings, see Geckeler et al. 2023.) Chapter 3 describes the research design that the study team used to estimate the impact of RP on participant outcomes. Chapter 4 presents findings from the impact study and discusses our interpretation of these results. Chapter 5 concludes the report.

A. Insights from the implementation study

Understanding how RP grantees implemented their programs provides important context for interpreting findings from the main impact study. The implementation study examined 16 intermediary grantees and 68 CBO grantees from the 2018 and 2019 grant cycles using grantee surveys, virtual site visits, Workforce Integrated Performance System (WIPS) records, and document reviews (Geckeler et al. 2023). Grantees intended to combine structured employment experiences through models such as registered apprenticeship, work-based learning, and career pathways, with case management to facilitate the transition to unsubsidized employment. However, many grantees faced implementation challenges related to enrollment, retainment, and service delivery.

- The 2018 and 2019 grantees enrolled 17,361 RP participants (9,098 adult grantee participants and 8,263 young adult participants). Nevertheless, due in part to the COVID-19 pandemic, many grantees did not meet their enrollment goals, with 2019 grantees falling shorter than 2018 grantees and young adult grantees falling shorter than adult grantees.

- Grantees also struggled to retain participants, especially young adults. Program staff noted that they struggled to keep some participants motivated and engaged, especially when competing with the short-term labor market opportunities that were available to participants. Staff reported that they helped participants shift their mindset about what was possible for them given their past experiences so that they could remain engaged in RP grant programs.
- Although many programs intended to offer work-based learning experiences, only a few participants received them. For example, based on WIPS records, only 1.3 percent of participants received registered apprenticeship programming and only 2.3 percent of participants received on-the-job training. Overall, about 72 percent of RP participants from 2018 and 2019 received education or training services, with 43 percent receiving occupational skills training. Nearly all grantees reported offering case management and an array of support services, but we do not have data to measure receipt of these services.
- The COVID-19 pandemic affected sites' abilities to offer education, training, and work-based learning opportunities to their participants. Grantees reported closures of training facilities and difficulties maintaining connections with employer partners or training providers. In some instances, employers were no longer willing to accept any individuals for work-based learning who were not already on their personnel roster.

B. Impact study overview and findings

The impact study aimed to estimate the causal effect of RP services by answering the following confirmatory research questions:

1. What is the impact of RP on the likelihood of being convicted of a crime over the 10 quarters after enrollment compared with Wagner-Peyser employment services?
2. What is the impact of RP on the likelihood of being employed in the 9th and 10th quarters after enrollment compared with Wagner-Peyser employment services?
3. What is the impact of RP on participants' earnings in the 9th and 10th quarters after enrollment compared with Wagner-Peyser employment services?

We compared program participants to a matched group of similar individuals who also sought out employment assistance but did not have access to RP services. We interpreted differences in post-program outcomes across these program and comparison groups as the impact of RP services.

Methods

To estimate the impact of the RP program on criminal recidivism, employment, and earnings, we constructed a comparison group consisting of participants in Wagner-Peyser employment services programs. Intuitively, Wagner-Peyser represents an alternative to RP, providing lighter-touch services (for example, access to a computer and job postings website) to people who, like those who enroll in RP, request help in securing employment. In the absence of RP, many individuals with criminal justice backgrounds looking for employment assistance may very well go to American Job Centers or otherwise enroll in Wagner-Peyser.

From the sizeable pool of potential comparison group members, we selected Wagner-Peyser participants who shared key demographic characteristics and features of their criminal justice backgrounds with RP participants. We used propensity score methods that enabled us to further refine our comparison group by incorporating information on a larger set of pre-program characteristics. We describe this matched comparison design in more detail in the Technical Appendix.

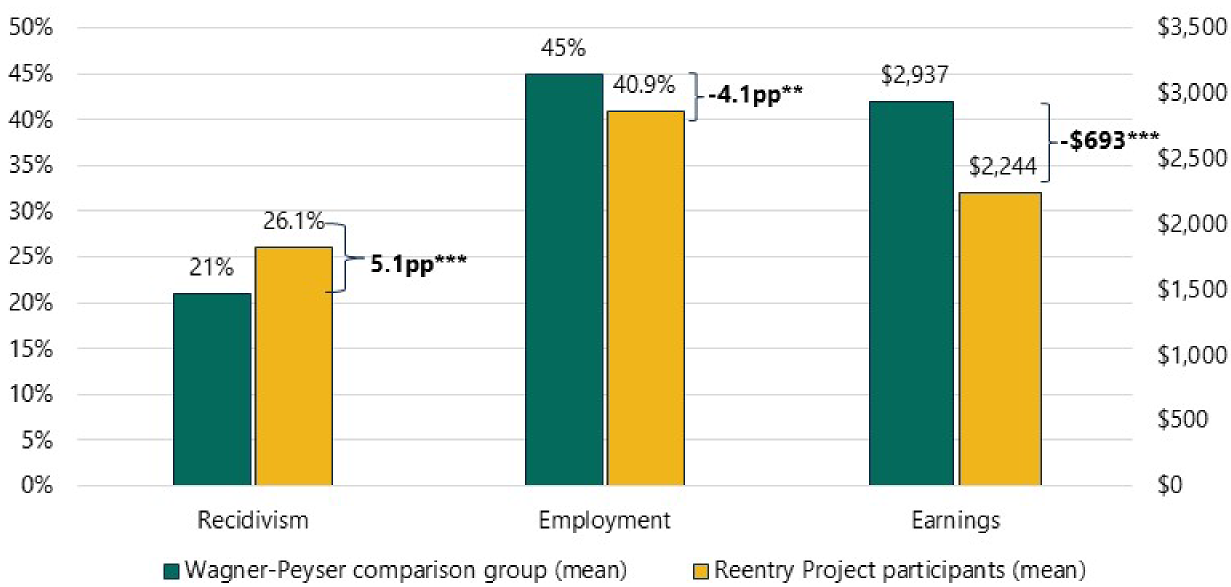
Data constraints limited the scope of the impact study. Given varying state restrictions on access to workforce and criminal justice data, we could only collect suitable data on program and comparison group members from 6 of the 34 states with RP grants: Alabama, Florida, New Jersey, New York, Oregon, and Pennsylvania. We used administrative earnings, employment, and criminal justice records to measure outcomes. Due to earnings data retention requirements, we could not obtain pre-program earnings and employment information for most members of our analytic sample.

Impact study findings

RP participants had worse post-program criminal justice and labor market outcomes relative to the matched Wagner-Peyser comparison group. In particular:

- RP participants were 5.1 percentage points more likely to have a new criminal conviction in the 10 quarters after program entry compared to Wagner-Peyser matched comparison group members.
- In the 9th and 10th quarters after enrollment, RP participants were 4.1 percentage points less likely to be employed than matched Wagner-Peyser comparison group members.
- RP participants earned $693 less in the 9th and 10th quarters after enrollment compared to matched Wagner-Peyser comparison group members, who earned $2,937 on average during that period.

Exhibit ES.1. Impact of Reentry Project on recidivism, employment, and earnings outcomes



Source: National Directory of New Hires (NDNH) data and state administrative court records matched to Workforce Integrated Performance System (WIPS) data. Sample includes data from 2018–2023.

Exhibit ES.1 (*continued*)

Notes: Employment is defined as having any earnings in a given quarter. Estimates based on a total sample size of 1,198 RP participants and 16,032 Wagner-Peyser participants. Wagner-Peyser group means are unadjusted; Reentry Project group means are adjusted means equal to the Wagner-Peyser group mean plus the estimated impact. For a detailed description of estimation methods, please see the Technical Appendix.

** p-value < 0.05

*** p-value < 0.01

pp = percentage points.

C. Interpreting results of the impact study

Our impact estimates indicate that RP participants did not fare as well in criminal justice and labor market outcomes compared to Wagner-Peyser participants. These findings support many potential interpretations, but we focus on two overarching narratives: (1) RP might lead to relatively worse outcomes for participants, relative to Wagner-Peyser, and (2) our estimation approach might fail to address underlying differences that generate worse outcomes for RP participants. We explore each of these interpretations in turn.

Contextualizing findings in the literature

At least two past randomized control trials (RCTs) found substantial adverse impacts of reentry services programs on participant recidivism. Specifically, Wiegand and Sussell (2016) report that participation in Reintegration of Ex-Offenders (RExO) increased the probability of future criminal convictions after program enrollment, while D'Amico and Kim (2018), who evaluated the Second Chance Act (SCA), found an overall increase in the number of future convictions after program enrollment for program participants compared to control group members.

Both of these prior studies provide experimental evidence that reentry services might increase reoffending behavior, and thus suggest that it is plausible for Reentry Project participation to have had an adverse effect on recidivism. Nonetheless, although the notion that reentry services may not be effective at preventing future crime is not new (see, for example, Doleac 2019), in the context of the existing evidence, our results are outliers. In reviews of the evidence on the effectiveness of similar reentry service programs, Lacoe and Betesh (2019) and Cortina et al. (forthcoming) documented mixed impacts of other reentry services programs. However, prior studies typically report either null effects of reentry services on employment and recidivism, or improvements in these outcomes.

Potential bias from limited pre-program data

Due to data limitations, our matching approach may have led us to compare RP and Wagner-Peyser participants with observably similar but, in reality, very distinct prior justice involvement and employment backgrounds. Our impact study had to overcome two critical sources of missing data: (1) a lack of granular sentencing, incarceration, and probation records; and (2) a lack of pre-program employment and earnings information. Absent such data, there may be fundamental differences between RP and Wagner-Peyser participants even after creating a matched comparison group. In particular, because of program eligibility requirements, most adult RP participants likely had recently received serious criminal sentences involving incarceration or supervised probation; Wagner-Peyser has no such condition. We were able to restrict our pool of comparison group members to Wagner-Peyser participants with recent criminal charges with similar case characteristics (for example, whether the case included a felony offense or whether the person had a history of other criminal cases). However, we did not have the data to limit the pool to individuals who would have met the Reentry Projects' eligibility criteria related to incarceration or supervised probation.

This may have resulted in a matched comparison group with very different background characteristics than the RP participants in our analytic sample. Before enrollment, the average RP participant might have been more likely than matched comparison group members to be incarcerated or under supervised probation, which might have limited their employability. After enrollment, RP participants may have been more likely to re-offend and less likely to secure gainful employment than the average comparison group member, given the relative seriousness of their prior justice involvement. In other words, lacking detailed data on pre-program criminal sentencing (and earnings), we may have arrived at a matched sample in which RP participants had systematically lower earnings potential before and after enrollment, and a higher risk of recidivating after enrollment, than matched Wagner-Peyser participants.

An analysis of variation in our impact estimates by severity of prior justice involvement supports this hypothesis. Although we cannot reliably observe incarceration or probation details, we can observe characteristics of individuals' recent criminal cases that are likely correlated with more serious sentences. Indeed, RP participants who had more serious prior justice involvement showed smaller differences in outcomes compared to similar matched Wagner-Peyser participants. That is, when we focused on RP and Wagner-Peyser participants with the greatest likelihood of pre-program incarceration or supervised probation, and the highest risk of post-program recidivism, we found relatively small and generally not statistically significant—albeit, non-zero—differences in outcomes. This pattern could suggest that missing data might lead us towards biased estimates.

Implications for policy and future research

The results from this impact evaluation provide several implications for policy and future research. RP grant implementation study findings suggest that RP participants often did not receive the structured employment-focused services that were intended as part of the RP grant program model, meaning that the impact findings likely do not reflect the impacts of these key program components. Existing evidence on employment-focused reentry programs has reinforced the importance of intensive employment and training services paired with wraparound supports for improving labor market and recidivism outcomes for populations with prior justice involvement (Lacoe and Betesh 2019; Wiegand and Sussell 2016). These program components are central to the program logic model for the RP grants, but participants likely did not receive these services as intended. These findings suggest that there would be returns to focusing programming and policy to address program implementation challenges experienced by grantees to ensure they are able to offer the intended program model and participants receive those services. To evaluate the impact of the intended program components, future research should consider assessing implementation fidelity, prior to beginning an impact study and using comprehensive measures of service receipt, to test that implementation of the program model is occurring as intended before evaluating it.

The potential sources of bias in our impact estimates are unique to a matched comparison design. An RCT approach, for instance, would obviate the need for additional pre-program administrative data (for example, jail or juvenile justice data), because treatment and comparison groups would both be formed from individuals eligible for the program and be identical on both observable and unobservable characteristics. However, RCTs are infeasible in many contexts. We explored the possibility of conducting an RCT for the RP grant evaluation but ultimately concluded that an experimental study design was not feasible due to low enrollment in the grant programs, which precluded us from ethically randomizing applicants to a control condition. Alternative approaches, such as piloting enhanced services, may allow for a randomized design in settings without oversubscription. Given the challenges in implementing an RCT design, running an RCT successfully may also require incentives or compensation for grantees to participate in such research and clear requirements for them to do so.

I. Introduction

Individuals released from incarceration often experience collateral consequences that can create barriers and impact access to education, employment, professional licensing, housing, public benefits, and other supports that help them successfully reenter society (U.S. Commission on Civil Rights 2019). More than 40 percent of prison and jail inmates lack a high school education (Denney et al. 2014) and many individuals involved in the justice system experience substance abuse and mental or physical disabilities (Bronson and Berzofsky 2017). These obstacles, along with the stigma of having a criminal record and limits on the types of jobs they can obtain due to restrictions on occupational licensing for people with criminal records, make finding a job difficult (Pager 2003; Holzer et al. 2004; Raphael 2014; Council of State Governments Justice Center 2020). Among those individuals released from state prisons, 45 percent are without employment one year following release (Looney and Turner 2018). Quantitative and qualitative evidence demonstrates that employment is an important component of successful reentry for individuals who have justice system involvement because it provides a needed source of income and serves as a positive activity that can help individuals establish healthy routines and reduce the likelihood that they will engage in future risky behaviors (Bellotti et al. 2018; Lacoe and Betesh 2019; Ramakers et al. 2017).

More than 608,000 individuals were released from state and federal prisons in 2019 (Carson 2020) and more than 10.3 million were admitted to local jails with an average stay of under one month before release (Zeng and Minton 2021). There is also a considerable number of people that have been convicted without having been incarcerated. As a result, there is substantial need for support to help justice system involved individuals prepare for, find, and retain long-term employment. To improve workforce and criminal justice outcomes for individuals with justice involvement, the U.S. Department of Labor (DOL) awarded $243 million in Reentry Project (RP) grants to 116 grantees between 2017 and 2019. RP grants aim to serve either adults (ages 25 and up) or young adults (individuals ages 18 to 24) who have been involved in the criminal justice system and/or who may be reentering society after incarceration.

DOL's Chief Evaluation Office (CEO) contracted with Mathematica and Social Policy Research Associates in September 2017 to conduct an evaluation of Reentry Project (RP) grants. This report presents results from the RP impact analysis and draws on findings from the previously completed implementation study to provide context for these results (Geckeler et al. 2023). This chapter provides background on DOL's history of reentry employment initiatives, describes the RP grants, and provides an overview of the impact findings included in this report.

A. Overview of the Reentry Project (RP) grants

Building on DOL's prior reentry efforts, the RP grant program encouraged organizations to implement comprehensive reentry programs to support justice system involved adults and young adults to successfully engage in their communities and avoid recidivism (DOL 2017). The funding opportunity announcement (FOA) established the grant's requirements, including specifying eligible organizations, eligible populations, and service delivery requirements aimed at improving participants' employment outcomes after reentry (DOL 2017). RP grants were the latest in a series of DOL-funded reentry grant initiatives.

1. Building on prior reentry employment initiatives

For two decades, DOL has invested in reentry services by providing substantial funding toward programs helping justice system involved individuals find employment and avoid recidivism. The RP grants represent one investment in a series of DOL grant initiatives supporting reentry programming (see Exhibit I.1). DOL has funded additional reentry efforts, including the Pathway Home grants program, which provides job preparation, career exploration and planning, and other supportive services pre- and post-release. Additionally, in partnership with the U.S. Department of Justice, Bureau of Prisons, the Partners for Reentry Opportunities in Workforce Development (PROWD) grants program provides employment-related reentry services to people in minimum- and low-security federal prisons, residential reentry centers, and the communities they return to upon release (DOL 2022).

Exhibit I.1. U.S. Department of Labor grant initiatives supporting reentry programming from 2010 to 2022

Source: DOL 2022b.

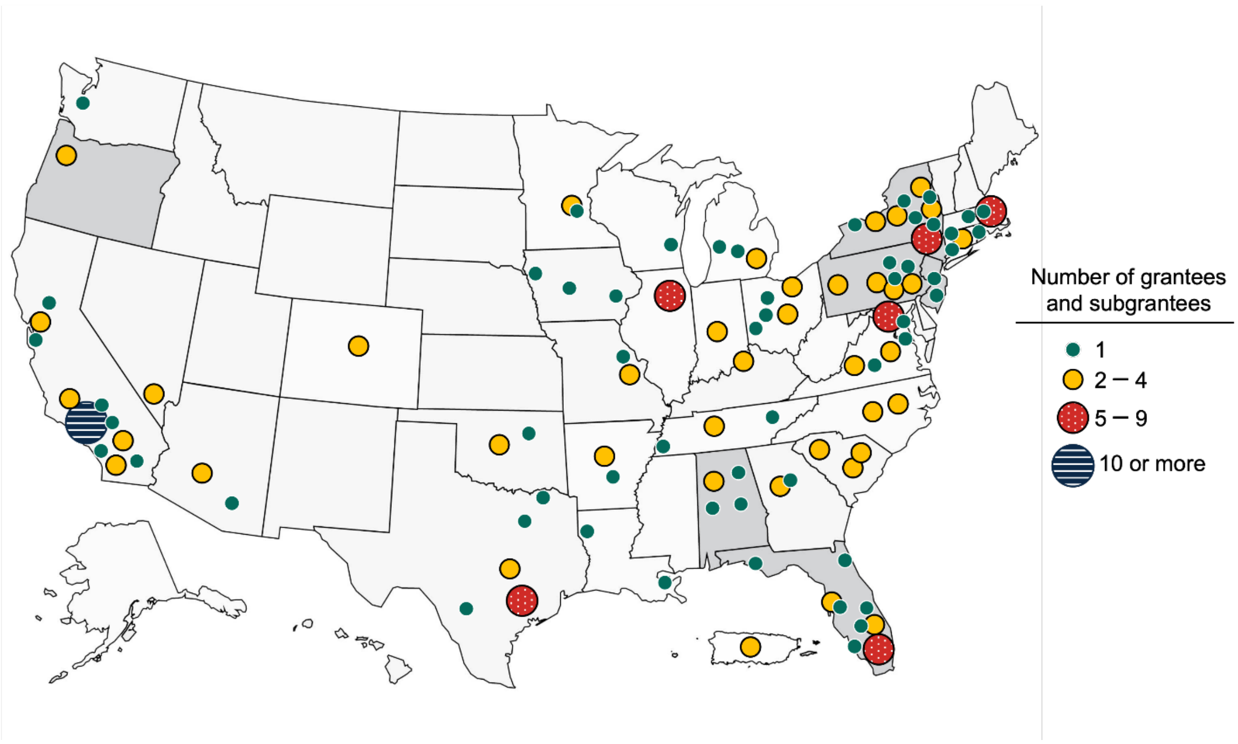
LEAP = Linking Employment Activities Pre-Release, PROWD = Partners for Reentry Opportunities in Workforce Development.

2. RP grantee characteristics, funding, and populations

Between 2017 and 2019, DOL awarded $243 million in RP grants to 91 community-based organizations (CBOs) and 25 intermediary organizations. Intermediary organizations served large numbers of participants across multiple subgrantees and states, and CBOs served a smaller number of participants in a single location (DOL 2018, 2019). Grant awards for intermediary organizations ranged from $3,996,685 to $4,500,000, with an average award amount of $4,462,217. Grant awards for CBOs ranged from $560,000 to $1,500,000, and the average award amount was $1,424,159. Grantee expected enrollment ranged from 70 to 705 participants, with an average enrollment expectation of 268 participants per grant. Grantees could not exceed an $8,000 cost-per-participant for the duration of the grant which included administrative, planning, and follow-up costs (DOL 2017, 2018, 2019).

Grant recipients were located in high-crime, high-poverty communities across the United States, with the majority (78 percent) of recipients implementing their programs in urban or suburban areas.[1] The intermediary and CBOs were awarded the RP grants between 2017 and 2019 and provided programs in 34 states, the District of Columbia, and Puerto Rico (Exhibit I.2).

Exhibit I.2. Locations of 2017–2019 Reentry Project grant programs



Source: Grantee applications and grantee surveys.

Note: States shaded in darker grey were included in the Reentry Project impact study.

In addition to applying as either an intermediary or CBO grantee, the grant eligibility criteria required applicants to select a target population: (1) adults (ages 25 or older) who were incarcerated in the adult criminal justice system and released from prison or jail within 180 days; or (2) young adults (ages 18 to 24) who were involved in the juvenile or adult justice system (up to 10 percent of participants could have been those who dropped out of high school without criminal justice involvement). Exhibit I.3 provides the

[1] In the 2018 RP FOA, high-poverty communities are defined as communities with poverty rates of at least 25 percent as exhibited through the use of American Community Survey (ACS) data, and high-crime communities are defined as "communities with crime rates within the targeted area that are higher than the rate for the overall city (for urban areas) or of non-metropolitan counties in the state (for rural areas)."

number of grantee program sites that served adult and young adult populations.[2] A little more than half of the grants were used to provide services to adult populations.

Exhibit I.3. Number of program sites by grant award year, population of interest, and grantee type

Grantee type	PY2017	PY2018	PY2019	Total
RP young adult	13	21	18	52
Intermediaries	4	7	6	17
CBOs	9	14	12	35
RP adult	19	21	24	64
Intermediaries	5	2	1	8
CBOs	14	19	23	56
Total	32	42	42	116

Source: Grant applications and clarifying calls

Note: Intermediary counts include counts of subgrantees. Some CBOs and subgrantees received both young adult and adult grants, as well as grants from multiple PYs.

CBOs = community-based organizations; PY = program year; RP = Reentry Project.

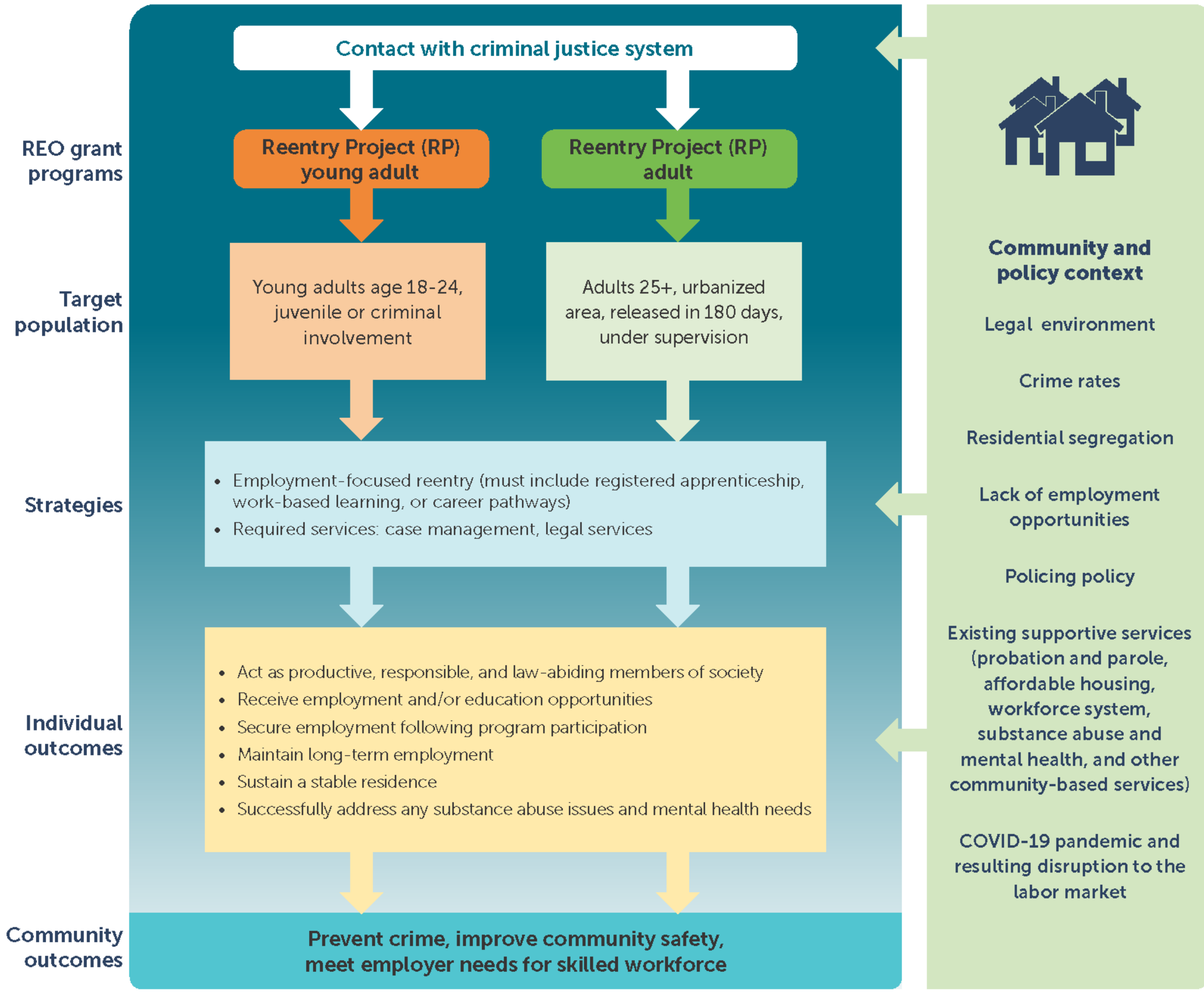
3. Program design and services

Although DOL provided organizations with substantial flexibility in their program design, the FOA encouraged applicants to build their projects using evidence-informed or promising practices in employment-focused services, as well as case management (DOL 2017, 2018). As illustrated in the logic model shown in Exhibit I.4, RP grantees intended to combine structured employment experiences—through models such as registered apprenticeship, work-based learning, and career pathways—with case management to facilitate the transition to unsubsidized employment.

The services that grantees offered varied depending on the grant stream and target group, but all grantees offered an array of services, including career preparedness, employment-focused services, and case management. In addition, all RP grantees were required to use at least one of the following employment strategies: registered apprenticeship, work-based learning, and career pathways. More information about grantee recruitment and enrollment efforts and the services they provided to participants is described in the RP implementation report and summarized in Chapter 2 (Geckeler et al. 2023). Grantees were not required to target specific industries, but findings from the virtual site visits with select 2018 and 2019 grantees showed that grantees tended to focus their programs on the major industries in their area that were friendly to hiring justice-involved individuals. These industries included construction (15 sites), culinary and hospitality (seven sites), manufacturing (eight sites), warehousing (seven sites), and transportation (six sites).

[2] Program site refers to the location where RP grant-funded services were implemented. For example, an intermediary grant counts as a single grant, but subgrantees deliver the services across several sites.

Exhibit I.4. Reentry Project logic model

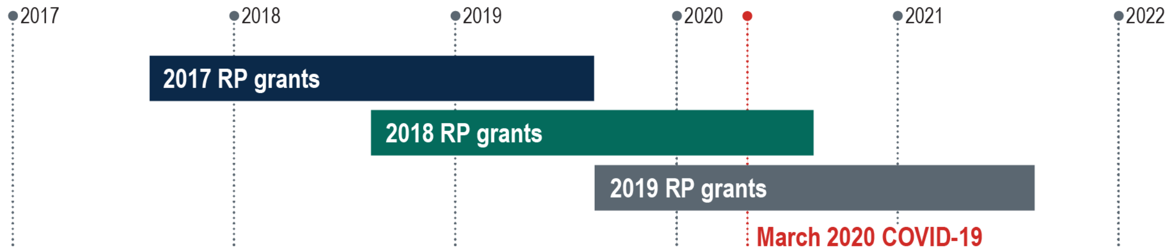


REO = Reentry Employment Opportunities.

4. Period of performance

The grants were 36 to 39 months long, including a three-month planning period, 24-months of enrollment and service provision, and a nine or 12-month follow-up period to assess participants' employment and criminal justice outcomes (DOL 2017, 2018, 2019). Exhibit 1.5 illustrates the 24-month planned period of operation or service delivery for all grantees and the timing of the grants against the COVID-19 pandemic. All grantees in the implementation study were operating when the COVID-19 pandemic began in March 2020, although some grantees were near the end of program operations and in the follow-up period. During site visits with 2018 and 2019 RP grantees, respondents stated that the COVID-19 pandemic altered program operations, including enrollment, service delivery, and participant outcomes. Over a third of program sites (11 of 27) shared during site visits that the pandemic initially led to high unemployment rates and limited job opportunities. By the time of site visits in early 2022, the job market had improved and the same program sites shared that there were more job opportunities for their participants as employers began to relax their hiring requirements. However, as the local economy improved, two sites explained that they found it difficult to recruit participants because they no longer needed training to secure employment.

Exhibit I.5. 2017–2019 Reentry Project (RP) grants planned period of operation



Source: DOL 2017, 2018, 2019.

Note: Information is based on the anticipated start date in the FOAs for RP grantees. Some 2018 and 2019 RP grantees received no cost extensions due to the COVID-19 pandemic that enabled them to continue enrolling participants through fall 2022 and 2023.

B. Evidence on similar programs

Rigorous studies on the impact of employment-focused interventions for individuals involved with the justice system have found a range of impacts on employment and recidivism outcomes. Although the service models varied, these studies examined the impacts of programs that most typically offered either work readiness training or job search assistance through interventions delivered to individuals after release from incarceration.

Some of these past studies suggest that reentry services might help individuals secure employment and avoid future criminal involvement. For example, the quasi-experimental evaluations of Texas' Project RIO (Re-Integration of Offenders) (Finn 1998) and the ComALERT Prisoner Reentry Program (Jacobs and Western 2007) both showed increased employment rates as well as reduced recidivism rates among program participants. However, as discussed in Lacoe and Betesh (2019), the overall evidence on many employment-focused reentry programs is mixed, with several studies finding no significant benefits for participants' labor market outcomes or recidivism rates. In particular,

- The Re-integration of Ex-Offenders (RExO) study (Wiegand and Sussell 2016) found no statistically significant benefits of program participation for individuals' employment or recidivism.
- The evaluation of the Center for Employment Opportunities program for individuals released from prison (Redcross et al. 2012) found that program participants had a 6.3 percentage-point higher likelihood of employment in unsubsidized jobs but were 5.6 percentage points more likely to be convicted of a crime in the first three years after enrollment, relative to the control group.
- The Enhanced Transitional Jobs Demonstration study found that randomly assigned program participants had a 4 percentage-point higher employment rate and earned $700 more annually than the control group but were no less likely to re-offend on average.
- The Second Chance Act Adult Demonstration Program (D'Amico and Kim 2018) found that, relative to an experimental control group, program participants had a 4.6 percentage-point higher employment

rate, earned $900 more per quarter, but were 6.4 percentage points more likely to have a new criminal conviction after 18 months.

This study seeks to build on this research base by examining the effectiveness of RP grant programs that use employment-focused services, wraparound supports, and other evidence-based strategies to improve outcomes for justice-involved youth and adults.

C. Evaluating Reentry Project grants

In 2017, DOL's CEO contracted with Mathematica and Social Policy Research Associates to build evidence about effective strategies to serve people with prior justice involvement and facilitate their successful reentry into the community. The RP evaluation aims to determine the impacts of the program on labor market and criminal justice outcomes (*impact study*) and understand how the grant programs were implemented across a broad range of intermediaries and CBOs (*implementation study*). This report presents findings from the impact evaluation, which uses a quasi-experimental design to estimate impacts for RP participants associated with 2017, 2018, and 2019 grantees in six states.

D. Impact study research questions

The impact study estimates the extent to which RP programs improved participants' earnings, employment, and criminal justice outcomes. The study team compared the outcomes of RP participants enrolled between 2018 and 2021 to a matched comparison group of Wagner-Peyser participants who enrolled in the same period (see Chapter 3 for additional details on the study sample). Specifically, the study team analyzed the research questions listed in Exhibit I.6, as specified in the study's pre-specification plan.

Exhibit I.6. Research questions used for Reentry Project evaluation impact study

Confirmatory research questions
1. What is the impact of RP on the likelihood of being convicted of a crime over the 10 quarters after enrollment compared with Wagner-Peyser employment services?
2. What is the impact of RP on the likelihood of being employed in the 9th and 10th quarters after enrollment compared with Wagner-Peyser employment services?
3. What is the impact of RP on participants' earnings in the 9th and 10th quarters after enrollment compared with Wagner-Peyser employment services?
Exploratory research questions
1. What is the impact of RP on participants' arrest rates and incarceration rates over the 10 quarters after enrollment compared with Wagner-Peyser employment services?
2. What is the impact of RP on participants employment and earnings outcomes in the 4th and 5th quarters after enrollment compared with Wagner-Peyser employment services?
3. What is the impact of RP on participants' conviction, arrest, and incarceration rates over the 5 quarters after enrollment compared with Wagner-Peyser employment services?
4. What is the impact of RP on the frequency and severity of criminal justice outcomes in the 4th and 5th and 9th and 10th quarters after enrollment compared with Wagner-Peyser employment services?

Note: Confirmatory research questions describe the primary analyses, which will be used to assess impacts of program participation. Exploratory research questions describe secondary analyses that help explain the primary impact estimates.

Exhibit I.7. Subgroups used for Reentry Project evaluation exploratory analyses

Subgroups for exploratory analyses
• Adult versus young adult participants
• Participants of different races or ethnicities
• Participants of different gender
• Participants with lower versus higher frequency of prior criminal justice involvement
• Participants served by different types of grantees (intermediary grantees versus community-based organizations)
• Participants who received different types of services (case management only, case management and work-based learning, and so forth)
• Participants who enrolled before versus those who enrolled during the COVID-19 pandemic

E. Data sources

The impact study used three distinct types of data:

1. **Workforce Integrated Performance System (WIPS)**. The WIPS is a national database that contains data on participants in DOL-funded workforce programs (as well as some U.S. Department of Education-funded programs), including Wagner-Peyser employment services and the RP grants. The WIPS contains data on individual-level demographic characteristics, including age, gender, race, ethnicity, disability status, education, employment status at program enrollment, and English learner status. The WIPS also includes data on employment and training services received through DOL workforce programs. We used these data to form a matched comparison group and examine impacts for key demographic subgroups and subgroups defined by service receipt (see Chapter 3 and Section A of the Technical Appendix for additional details).
2. **National Directory of New Hires (NDNH)**. NDNH data are maintained by the Office of Child Support Services, Administration for Children and Families, U.S. Department of Health and Human Services. These data include information collected through states' unemployment insurance systems and describe quarterly employment and earnings. We used NDNH data to examine employment and earnings for the full post-enrollment period. We discuss the limitations of the NDNH data in Section A of the Technical Appendix.
3. **Criminal justice data**. State criminal justice data primarily provided information on criminal convictions both before enrollment (to use as matching variables and to form subgroups of interest) and after enrollment (to use as outcomes). Additional information on criminal justice data is included in Chapter 3 and in Section A of the Technical Appendix.

F. Sample description and characteristics

The impact study's population of interest consists of RP participants who enrolled between 2018 and 2021, along with comparison group members who enrolled in Wagner-Peyser services in the same period. The impact study is limited to participants in the six states for which we could obtain pre-program and outcome data on both RP participants and comparison group members.[3] The RP impact study participants were diverse in terms of their demographic backgrounds. As shown in Exhibit I.8, about 63 percent of participants were ages 18 to 29 at enrollment, 21 percent were ages 30 to 39, and 15 percent were ages 40 or older. About 21 percent of the participants self-identified as female. Twenty percent of participants identified as White; 73 percent identified as Black; and 6 percent were of another racial background. Fourteen percent identified as Hispanic, while 86 percent identified as non-Hispanic.

Exhibit I.8. Characteristics of Reentry Project participants at program entrance (July 2018 to December 2021)

Sample Size 3,090

AGES (YEARS)

18 to 24 years	49%
20 to 29 years	14%
30 to 39 years	21%
40 to 49 years	9%
50 to 59 years	5%
60 years and older	1%

GENDER

Female	21%
Male	79%

EDUCATION LEVEL

No HS completion	35%
HS equivalent	26%
HS graduate	34%
Any postsecondary	5%

RACE / ETHNICITY

White	20%
Black	73%
Other/Multiracial	6%
Hispanic	14%
Non-Hispanic	86%

Source: Workforce Integrated Performance System data on Reentry Project participants who enrolled between 2018 and 2021.

[3] Chapter 3 describes the selection of RP participants for the impact study in further detail, including how the RP participants in the impact study compare to all RP participants in terms of demographic characteristics and service receipt.

G. Limitations

There are several limitations to the impact study findings that are important to note. First, findings may not be generalizable. Data availability limited the RP participants that we could include in our analysis to a subset of RP participants in six of the 34 states in which the program operated. Although the characteristics of RP participants in the study and the services they received were generally similar to the full population of RP participants, these results are not necessarily generalizable to other RP grantees not included in the study sample. In addition, RP participants in other states and programs likely faced different labor market conditions and policy environments than those included in the study sample. Chapter 4 also discusses potential limitations of our quasi-experimental comparison group design, including unobservable differences between the treatment and matched comparison group that could lead to bias in our findings.

H. Structure of report

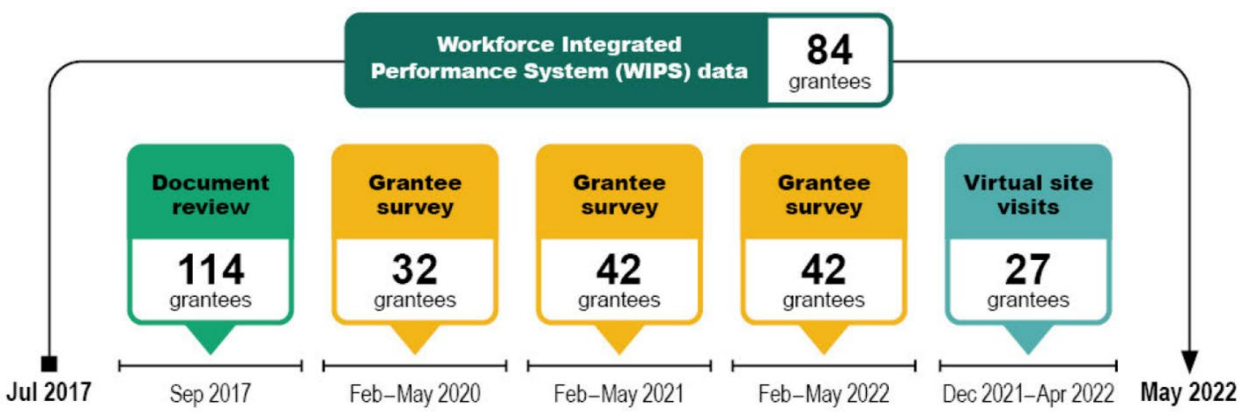
This report presents results from the impact analyses and important context for these results. Chapter 2 describes the implementation of 2018 and 2019 RP grants to provide context for interpreting the impact findings. Chapter 3 describes the impact study sample and design, and Chapter 4 discusses the findings. The report concludes by highlighting key takeaways from the impact analyses, context for the findings, and future directions for continuing research. The Technical Appendix provides details on the data, sample, methods, and sensitivity analyses.

II. Implementation of Reentry Project Grants

A recent meta-analysis of 127 interventions reviewed in the Pathways to Work Clearinghouse demonstrated that the efficacy of job training programs depends on what services are offered, how the program is implemented, and who the program serves (Shiferaw and Thal 2022). Additionally, although most prior studies of adult reentry programs do not find consistent, positive effects, this may be due to variation in program models, implementation quality, and study designs (Lacoe and Betesh 2019). Therefore, understanding the RP's implementation has important implications for interpreting any impacts observed as well as the policy implications of those impacts. This chapter presents a summary of key findings from the RP implementation study and identifies considerations for the impact study. The RP implementation study described how the 2018 and 2019 RP grant programs were implemented across a range of intermediaries. The full findings from the implementation study are available in a previously released report (Geckeler et al. 2023).[4]

The implementation study focused on 84 RP grantees from 2018 and 2019 and drew on: a grantee document review; virtual site visits with 27 of the 2018 and 2019 RP-funded grantees or subgrantees; a grantee survey; and Workforce Integrated Performance System (WIPS) data from program year (PY) 2018 Q1 to PY 2021 Q2, or July 1, 2018, to December 31, 2021 (Exhibit II.1).[5] While some data collection methods (such as the grantee survey and document review) also included 2017 grantees, the implementation study only included data from the 2018 and 2019 grantees in its analysis.

Exhibit II.1. Reentry Project implementation data collection timeline, July 2017 – May 2022



[4] The implementation study report and series of briefs can be found on DOL's website at: https://www.dol.gov/agencies/oasp/evaluation/completedstudies/Reentry-Projects-Grant-Evaluation.

[5] The selection process for the 27 visited sites was purposeful and considered a wide range of factors. While the evaluation team sought to include a diverse group of sites, the process was not random, and data collected were not fully representative of all sites. Selection criteria included timing (sites had to be operating until at least March 2022, as the site visits occurred in early 2022), a blend of intermediary subgrantee and community-based organization (CBO) grantee sites, a balance of sites operating young adult and adult RP grant programs, and sites representing geographic diversity. The study team also included some of the intermediary subgrantees and CBO grantees that were likely be included in the impact study. Additionally, based on conversations with DOL staff and review of grant applications, the study team also selected sites based on information that indicated they had implemented strategies of interest to DOL, such as offering apprenticeships or providing cognitive behavioral therapy.

The implementation study examined how 2018 and 2019 grantees identified and coordinated programming with subgrantees, processes for recruiting and enrolling participants, case management and service planning, and educating and training services offered and provided to RP participants. The implementation study includes a larger set of grantees than those in the impact study sample. For instance, the grantee document review, grantee survey, and WIPS records included data for almost all RP grantees for the years noted above (additionally, the grantee document review and grantee survey also included 2017 grantees, though data from this group was not included in implementation study analyses). Therefore, these sources and related analyses include grantees with participants in the impact study as well as many grantees that are not in the impact study analysis. Additionally, due to the data collection methods used in the impact study (described in detail in the next chapter), there are many grantees that only had a minority of their participants included in the impact study. For this reason, we do not present implementation study results pertaining only to the impact study grantees, as that category is not well defined.

According to the WIPS data, the 2018 and 2019 RP grantees enrolled a total of 17,361 participants in their RP programs. By comparison, 3,090 RP participants were eligible to be included in the impact study. See Chapter 3 for a discussion of impact study sample; see Chapter 3, Exhibit III.2 for a comparison of impact study participants to non-impact study participants.

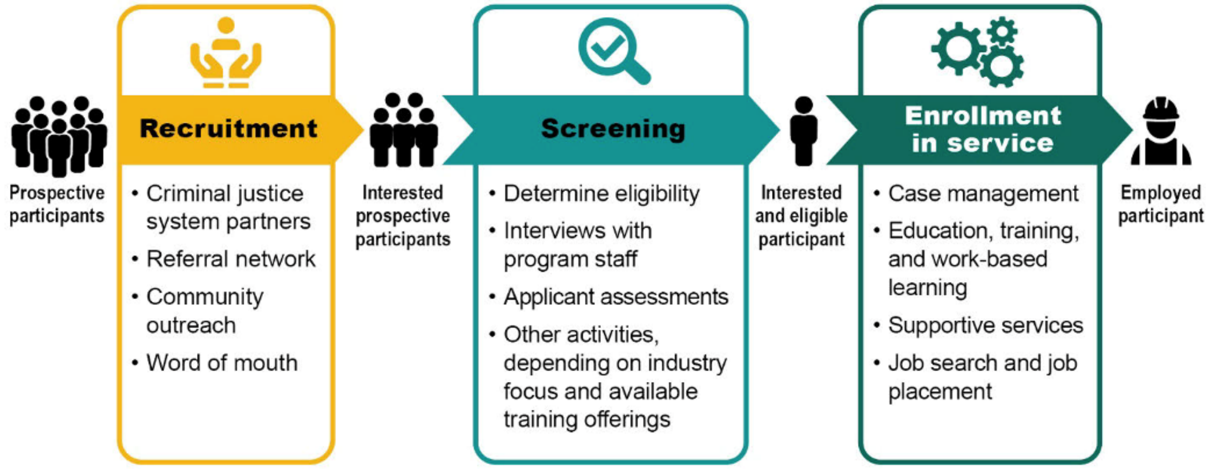
A. RP implementation study key findings

By providing insights on the services available to RP participants, populations served, variation in the services provided across and within RP grantees, grantee characteristics, as well as the challenges and successes grantees experienced, the implementation study provides important context for the impact study findings. An important, overarching context is that the COVID-19 pandemic occurred during the implementation study period for 2018 and 2019 grantees[6]. Grantees identified numerous challenges related to the pandemic, including constraints on their ability to conduct outreach, recruitment, and service delivery. These challenges are noted throughout this chapter when potentially relevant to interpreting impact study findings.

- After determining eligibility based on DOL's established criteria, RP grantees employed multiple strategies for screening potential participants to ensure their suitability for RP programming. As reported in the grantee survey, grantees utilized common screening activities including interviewing with program staff (95 percent of grantees), completing application forms (94 percent of grantees), and undergoing a criminal record review (83 percent of grantees) (Exhibit II.2). Screening activities, like interviews, were performed to assess whether a participant was willing to commit to programming and prepared for the demands of the program. Compared to adult grantees, grantees that served young adults more frequently reported assessing potential participants’ education levels and prior work experience as well as requiring interviews and application forms.

[6] RP grants were 36-39 months long, including a three-month planning period, 24 months of enrollment and service provision, and a nine or 12-month follow-up period to assess participants’ employment and criminal justice outcomes (DOL 2017, 2018, 2019). All grantees were operating at different grant phases, meaning some grantees were in early stages of operation while others were concluding operations, when the COVID-19 pandemic began in March 2020. Note that some 2018 and 2019 RP grantees received no cost extensions due to the COVID-19 pandemic that enabled them to continue enrolling participants through fall 2022 and 2023.

Exhibit II.2. Typical sequence for linking potential participants to RP services



Source: Virtual site visits (N = 27).

- Despite their outreach efforts, RP grantees reported challenges recruiting and enrolling participants. Most grantees (69 percent) indicated in the survey that recruiting participants was "somewhat" or "very" challenging. Virtual site visits provided further insights, with at least 13 sites reporting that recruitment became challenging during the COVID-19 pandemic due to court closures, restricted contact between referring partners, suspension of community outreach activities, and greater isolation that limited word-of-mouth referrals. Grantees also somewhat struggled with enrollment due to the COVID-19 pandemic. Despite enrolling a large number of participants, many 2018 and 2019 grantees were unable to reach the enrollment goals established with DOL. Exhibit II.3 below shows the progress grantees made toward reaching enrollment goals. As shown, a greater share of 2018 grantees (63 percent) reached enrollment targets compared to 2019 grantees (49 percent).

Exhibit II.3. Percentage of Reentry Project grantees meeting enrollment goals, by grant year

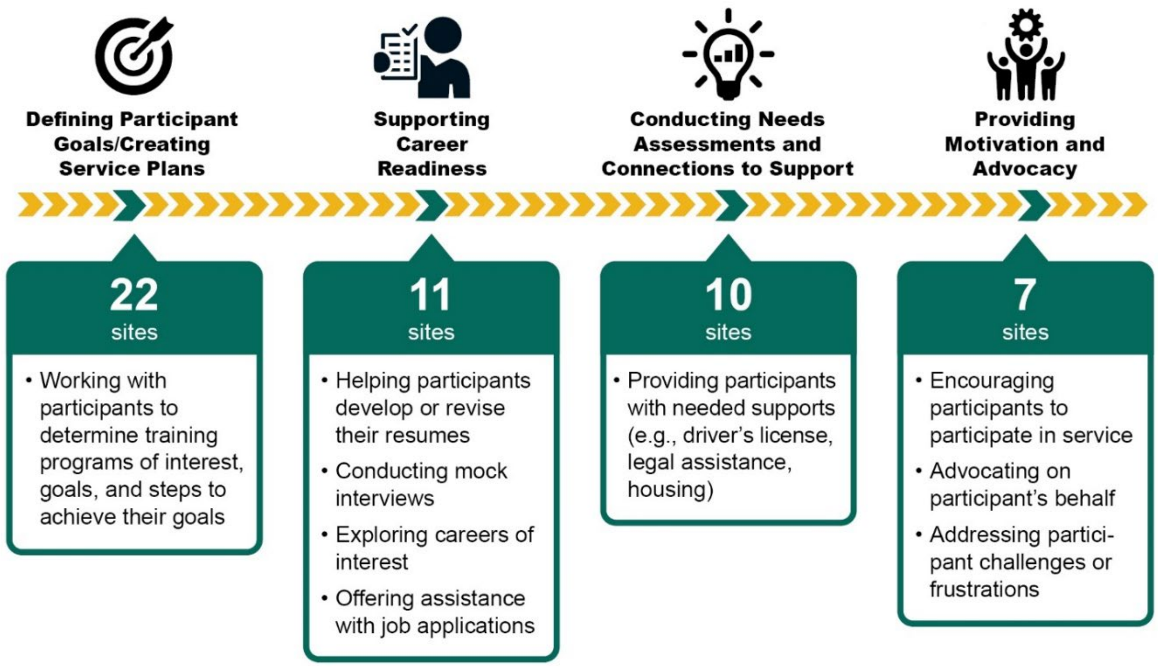
Percentage of enrollment target met	Percentage of 2018 grantees	Percentage of 2019 grantees
100 percent and above	63%	49%
90–99 percent	16%	12%
75–89 percent	8%	7%
50–74 percent	5%	15%
49 percent and below	8%	17%

Source: Workforce Integrated Performance System data (n= 79 RP grantees), July 1, 2018–December 31, 2021.
Note: Grantees outside of the impact study or that had missing WIPS data were excluded.

- Case management was an integral component of program service delivery. Ninety-seven percent of surveyed grantees from 2018 and 2019 had at least one case manager, with an average of 2.5 case managers per RP program. During virtual visits to nine of 27 sites, interviewed participants and program staff emphasized the importance of the case manager/participant relationship in motivating participant success. As illustrated in Exhibit II.4, during site visits program staff identified common goals for case management services.

- Intermediary grantees reported establishing set case management models that their subgrantees followed, while CBO grantees developed models to meet local needs. Intermediary organizations reported providing their subgrantees with guidance regarding their case management models to promote consistent experiences for participants across subgrantee locations. For example, to encourage overall uniformity throughout programming, three visited intermediaries stated that their 10 subgrantee CBOs all followed a standard model of service delivery. As highlighted through virtual visits, CBO grantees all described developing their case management models based on local community context, such as availability of other services in their communities, and participant needs.

Exhibit II.4. Reentry Project case management goals as identified by site visit respondents



Source: Virtual site visits (N = 27).

Note: Categories are not mutually exclusive and individual sites may have identified more than one goal.

- Case managers regularly connected with RP participants. RP grantees required case managers to meet with their clients from as frequently as weekly (11 of 27 visited sites) to as seldom as monthly (eight of 27 visited sites). Although grantees established standards to ensure that all participants received a baseline level of contact with their case managers, frontline staff from 10 site visit grantees described being in more frequent contact with participants than their programs required. Frequency of contact also varied over the course of program enrollment for at least seven of 27 visited sites, with more frequent check-ins common at the beginning of the program and tapering off as participants were engaged in training or with employment.
- Commonly reported challenges related to case management included maintaining participant engagement, addressing participant needs, and navigating the COVID-19 pandemic. Eighty-three percent of surveyed grantees from 2018 and 2019 said that engaging and retaining participants was somewhat or very challenging. Program staff from 15 of 27 visited sites also noted that it was difficult to keep participants engaged or motivated, especially given everything going on in their lives and the

world. Program staff from 20 of the 27 sites that were included in virtual visits wished their referral partners had more capacity to provide participants with mental health services, substance use disorder treatment, housing assistance, and transportation services. According to staff from at least eight of 27 visited sites, the pandemic was also a key challenge for case management because they shifted to virtual services with little planning time and had to support participants who had experienced additional traumas due to COVID-19.

The duration of service receipt varied by target population. Based on WIPS records of 2018 and 2019 grantees, 28 percent of RP participants exited within 2 quarters of program entry and 52 percent exited within 4 quarters of program entry. Twenty-eight percent had not exited the program by the end of 2021. The duration of service receipt differed for young adult and adult grant participants, with young adults being more likely to receive services for a longer period than adults.

- A large portion of RP participants received education or training services (72 percent of 2018 and 2019 grantee participants), but grantees did not all provide training to the same amount of their participants. According to the WIPS data, 15 percent of RP participants from 2018 and 2019 grantees received basic career services (such as assistance using online search engines and applications and tailoring resumes for job applications); another 13 percent received individualized career services; and 72 percent received education or training (and also often received individualized or basic career services). Data was not available to measure receipt of case management services. Despite the high share of RP participants who received education or training services, there was high variation among grantees in the rates of participant training receipt (Exhibit II.5). Nine percent of 2018 and 2019 grantees provided training to all their program participants, while a fourth of grantees provided training to less than half of their participants.

Exhibit II.5. Percentage of Reentry Project participants receiving training, by grantee

RP participants receiving training	Share of 2018 and 2019 RP grantees
100 percent	9%
90–99 percent	16%
75–89 percent	17%
50–74 percent	33%
49 percent and below	25%

Source: Workforce Integrated Performance System data, July 1, 2018–December 31, 2021 (N = 81).

Note: One grantee with missing WIPS data.

- **In practice, training services were concentrated into a few categories.** Exhibit II.6 provides more information on the various types of training services RP participants received, training completion rates, and a breakdown of these services and training completion by adult and young adult participants. Occupational skills training was the most common type of training received (43.3 percent). According to site visit data, the most common occupational skills trainings offered were access to industry-recognized credentials, such as Occupational Safety and Health Administration (OSHA) certifications, forklift certificates, and certified nursing assistant credentials. Participants could also earn certifications in health care.

Exhibit II.6. Education and training services received by Reentry Project participants, by target population

Service type	All participants (N = 17,361)	Adult participants (N = 9,098)	Young adult participants (N = 8,263)
Training received			
Occupational skills training	43.3%	60.9%	24.0%
Registered apprenticeship program	1.3%	1.8%	0.8%
Skill upgrading	2.1%	2.8%	1.4%
On-the-job training	2.3%	3.0%	1.4%
Incumbent worker training	1.1%	1.4%	0.7%
Customized training	1.0%	1.5%	0.4%
Training completed, among those who started			
Any training	84.3%	89.7%	76.8%
All trainings started	80.3%	85.5%	73.1%

Source: Workforce Integrated Performance System (WIPS) data, July 1, 2018–December 31, 2021 (N = 17,361).

Note: The section of the exhibit on training completed is based on the number of participants who entered training, not all participants (N = 12,118).

- Grantees offered work-based learning (WBL) experiences, but a small share of participants received them. As identified through the grantee survey, 2018 and 2019 RP grantees most frequently offered WBL through apprenticeships (82 percent) and on-the-job training (59 percent). Across these grantees, the types of WBL opportunities and their length varied from light-touch job shadowing to more intensive apprenticeships. Despite the high percentage of grantees offering WBL, the actual percentage of participants who received these services was much lower which may be related to the timing of these grants and COVID-19 pandemic. For example, according to WIPS data shown in Exhibit II.6 only 1.3 percent of participants received registered apprenticeship programming and only 2.3 percent of participants received on-the-job training. Of the participants that did receive training, 80.3 percent completed all trainings they started. It is important to note that the WIPS data do not include other less intensive forms of WBL such as facility tours, job shadowing, and internships, although interviews with RP sites reported offering these types of WBL activities to reentry populations.
- Despite the availability of education and training opportunities, RP grantees identified challenges providing or connecting participants to these opportunities. Over half of surveyed 2018 and 2019 grantees (54 percent) reported some challenges providing or giving participants access to high-quality education-related activities. Respondents from 14 of the 27 sites involved in virtual visits noted that the length of educational programs often disincentivized participants from completing them (note that the actual length of these programs varied greatly as they ranged from shorter opportunities like preparing for a high school equivalency exam to enrolling in a college course). According to the site visit respondents, participants' financial constraints exacerbated this problem, as they needed to earn money while enrolled in classes. The COVID-19 pandemic also appeared to limit sites' abilities to offer education, training, and WBL opportunities to their participants. For example, the pandemic shifted some occupational skill training programs online. This posed challenges for participants from at least 10 of 27 visited grantees who did not have the proper equipment to access the training and for others who became exhausted by virtual engagements. Other program staff reported that employers reduced or

eliminated WBL opportunities during the pandemic due to their protocols for hosting people in their facilities. This may mean that the intensity of education and training services declined over the course of the study, which could have implications for detecting the impact of the program.

- When working to place participants in employment, RP program staff offered intensive job search support, job placement, and job retention assistance. As 16 of the 27 visited sites reported, job placement assistance was an important employment-focused service. The sites stressed that their programs build up to employment, with all the previous steps—assessment, goal setting, and training—leading up to job placement. RP staff then helped support career exploration, connected participants directly to employers for application and interviews, and provided ongoing support following placement. Six visited sites also shared their strategy for providing intensive support to teach participants how to search for jobs with the goal of building independence and their job search skills.
- Commonly reported challenges working with participants included meeting their basic needs and engaging participants in programming. When asked about the biggest participant-level challenges faced during implementation, respondents from 22 of the 27 visited sites spoke about participants' unmet basic needs. In particular, they reported only limited access to stable housing (12 sites), mental health and trauma-based services (12 sites), and transportation (11 sites). Site visit respondents also highlighted challenges with participant engagement. RP program staff from 16 of the 27 visited sites described how it could be difficult to keep participants engaged. Staff members from 12 of these 16 sites found it particularly challenging to keep the attention of young adults and sustain their motivation. They described the young adult population as not wanting to work, not yet thinking about the type of life stability that more education and training promised, not completing training once begun, and generally having a short-term mindset about personal change.
- Commonly reported successes working with participants included: helping shift participants' mindsets, connecting them to education and training, and helping them prepare for and find employment. Partners and participants from 11 sites noted that their most important program successes pertained to helping participants change their perspectives about what was possible for them given their past experiences. Interviewed respondents, including program staff and participants from 15 sites, indicated that some of their programs' greatest successes were helping participants to complete education and training services and to obtain degrees and certifications. Five sites mentioned the success of helping participants obtain a high school diploma or high school equivalency certification. Additionally, helping participants find and retain jobs was one of the greatest successes noted in interviews with 17 sites. Staff members from four sites also described helping participants find jobs with the potential for advancement.

B. Considerations for RP impact study

Interpreting results from an impact study requires understanding how participants are selected into the treatment group and the comparison group and the contrast in services that each of the two groups received. In this study, treatment group members enrolled in RP, while comparison group members received services under Wagner-Peyser, a public employment services program (as described in more detail in the next chapter). Chapter 3 provides additional information on the offerings and features of the

Wagner-Peyser program and the comparison group, including available services and the participants themselves.

As described earlier in this chapter, RP grantees typically enrolled participants through referrals from criminal justice system partners, community outreach efforts, and word-of-mouth referrals. RP grantees conducted screening activities with potential participants, such as conducting interviews with program staff to determine work readiness and willingness to commit to the program. Additionally, grantees serving young adult sometimes also assessed potential participants' education levels and prior work experience in considering their enrollment in the program. In comparison, Wagner-Peyser services are available broadly to job seekers looking for career services through the American Job Centers (AJC) system (English and Holcomb 2020).

Overall, the implementation study found that RP programs tended to offer more intensive services than is typical under Wagner-Peyser. Wagner-Peyser provides basic career services, often described as lighter touch. These services generally include job search and placement services; reemployment services for individuals receiving unemployment compensation; and access to the state's labor exchange, which includes open job orders (DOL 2024). Although available services vary across states, Wagner-Peyser participants are not typically able to access additional ongoing case management, education, training, or supportive services through the Wagner-Peyser program. Additionally, the Wagner-Peyser participants in the comparison group may have been more removed from the justice system, may not have been low-income, and may have been more motivated to find employment. Any dislocated workers would also necessarily have had at least once instance of successful employment.

In contrast, the RP services described in this chapter are more intensive than those offered through Wagner-Peyser and designed to address barriers that individuals with prior justice system involvement faced. However, these services may not have been as intensive as intended under the RP grant program model. Nearly all RP participants were offered case management services that were targeted to the needs of individuals with prior justice system involvement. However, in practice, few actually participated in structured employment experiences (WBL, apprenticeships, and career pathways). Despite this, according to the WIPS data the majority of RP participants served by 2018 and 2019 grantees (72 percent) accessed education or training services (Geckeler et al. 2023).

Finally, the implementation study also identified a potential decline in RP service intensity over the course of the grant. Site visit respondents described this decline as connected to the COVID-19 pandemic. Staff at nine visited sites discussed the process of switching from in-person to virtual service delivery as further diminishing service delivery capacity when other program aspects were already strained. As one staff member put it, "The effect was the loss of a sense of community." They attributed COVID-19 as limiting their ability to recruit participants, provide in-person case management, and offer the planned array of education and training options, such as in-person GED classes that shifted to self-guided online courses. Although changes to the provision of the Wagner-Peyser program services also occurred over the COVID-19 pandemic, it is plausible that because RP services are more intensive, they would have been more disrupted by the pandemic than the Wagner-Peyser services, thus reducing the service contrast between RP and Wagner-Peyser programming.

III. Impact Study Design

The impact study aims to estimate the causal effect of the RP program on participants' future criminal justice involvement, employment, and earnings. To do so, we compared RP participants to a matched group of similar individuals who sought out employment services via the Wagner-Peyser program. Intuitively, Wagner-Peyser represents an alternative to RP, providing lighter-touch services (for example, access to a computer and job postings website) to people who, like those who enroll in RP, request help in securing employment. In the absence of RP, many individuals with criminal justice backgrounds looking for employment assistance may very well go to American Job Centers or otherwise enroll in Wagner-Peyser. We constructed a comparison group of Wagner-Peyser participants that had prior involvement with the criminal justice system and shared other key background characteristics with RP participants.

This matching design has the advantage of estimating the impact of RP among all program participants whom we could match to a suitably similar group of Wagner-Peyser participants.[7] Our matched comparison group's labor market and criminal justice outcomes approximated what RP participants would have experienced had they not taken part in the program. By comparing these two groups' outcomes, we estimated the relative impact of the RP program on employment, earnings, and criminal recidivism.

We begin this chapter by describing how we developed the sample of states and RP grantees included in this impact study. We then describe our sample of RP participants and the nature of our comparison condition, a set of light-touch employment services provided through the Wagner-Peyser program. After discussing how we constructed our comparison group sample, we assess the similarity of our matched sample of RP and Wagner-Peyser participants. Finally, we describe the methods we used to estimate the impact of RP, the outcomes we evaluated, and the limitations of our design.

A. Selecting states and grantees for the impact study

RP grantees provided employment services to individuals with past criminal justice involvement, with the goal of improving their employability and earnings. Nationwide, DOL sponsored 116 grantees that served participants from 34 states plus the District of Columbia and Puerto Rico. For the impact study, we initially focused on 11 states where the four largest RP intermediary grantees operate.[8]

Although the study team attempted to include as many of these grantees and states as possible in the impact analysis, we faced two constraints that limited the scope of our final sample. First, our empirical analysis depended on access to person-level identifiers, such as names and Social Security numbers. We needed these identifiers to complete a person-level linkage between DOL program participation data contained in the WIPS—for both RP and Wagner-Peyser participants—and criminal justice and labor market administrative data from state justice agencies and the NDNH. For RP participants, we needed to collect names and dates of birth from RP grantees. For Wagner-Peyser participants, we needed to collect identifying information from individual state workforce agencies. To use study resources most efficiently,

[7] Although we explored the possibility of conducting a randomized control trial, in consultation with DOL, we determined that RP grantees would likely not achieve oversubscription to their programs, which was required to ethically randomize access to reentry services across otherwise eligible applicants.

[8] The 11 states included the six final states shown in Exhibit III.1, plus Minnesota, North Carolina, Ohio, Puerto Rico, and South Carolina.

we gathered identifying information from RP grantees that served the largest number of participants in the 11 selected study states and did not collect information from grantees serving relatively few participants in these states. In addition, we had to restrict our focus to the seven states where the state workforce agency was willing to share identifying information for Wagner-Peyser records.

The second constraint we encountered when selecting states for our study was the availability of statewide criminal justice records, and specifically state court records. These data are essential for the study. To construct our sample, we needed to know whether individuals faced criminal charges before program enrollment, though our preferred measurement of criminal recidivism depends on our ability to observe post-enrollment criminal convictions. However, not all states maintain and release identified court data that enable us to observe criminal charges. Therefore, the final sample only includes those states that agreed to provide both identifiers for Wagner-Peyser participants and individual-level criminal court data.[9]

Ultimately, we obtained both Wagner-Peyser participant identifiers and criminal charge data from six states—Alabama, Florida, New Jersey, New York, Oregon, and Pennsylvania.

Exhibit III.1. Reentry Project and Wagner-Peyser sample sizes, by state

State	RP participants	WP participants
Alabama	261	8,828
Florida	640	22,142
New Jersey	53	3,570
New York	1,101	25,619
Oregon	277	12,563
Pennsylvania	758	10,630
Total	**3,090**	**83,352**

Source: WIPS data matched to state criminal court records.

Note: Sample sizes include RP and Wagner-Peyser participants whom the study team matched to pre-program criminal charge data. RP participants include those enrolled in both intermediary and community-based organization grantees in the state for whom we could obtain identifying information.

RP = Reentry Project; WP = Wagner-Peyser.

B. Selecting RP participants for the impact study

Our sampled states contained 3,090 of the 18,740 total RP participants (Exhibit III.1) who enrolled between 2018 and 2021.[10] This sample generally resembled the full population of RP participants. In Exhibit III.2, we compare program group members from our selected states to the remaining RP participants nationally whom we were unable to include in the sample. Participants in our sample had similar distributions of ages, racial backgrounds, and gender as the population of RP participants not included in our impact

[9] We detail our approach to collecting criminal justice records in the Technical Appendix.

[10] As we discuss in further detail in the Technical Appendix, impact estimates are based on the subset of these RP participants for whom we observe employment and earnings data for our confirmatory outcomes and for whom we were able to identify a suitable matched comparison. The impact study included data on RP participants enrolled in 2017, 2018, or 2019 RP grantees, but excluded any 2017 RP grant participants who enrolled prior to 2018.

study sample. However, the study sample contained a smaller share of Hispanic RP participants than the full population, as well as a smaller share of participants with college (postsecondary) credits or degrees.

The impact study sample was also generally similar to the full population of RP participants in terms of the type of grantee in which they enrolled and the services they received, though participants in the study sample were more likely to have enrolled in a program run by an intermediary grantee rather than a CBO.

Exhibit III.2. Characteristics of Reentry Project participants in the impact study, compared to the full population of program participants

Characteristic	Study RP participants	All other RP participants
Demographic characteristics		
Age (years)		
18 to 24 years	49%	45%
25 to 29 years	14%	11%
30 to 39 years	21%	20%
40 to 49 years	9%	14%
50 to 59 years	5%	8%
60 years and older	1%	2%
Sex		
Female	21%	21%
Male	79%	79%
Race		
White	20%	25%
Black	73%	70%
Other/multiracial	6%	5%
Ethnicity		
Hispanic	14%	21%
Non-Hispanic	86%	79%
Education		
No HS completion	35%	37%
HS equivalent	26%	21%
HS graduate	34%	33%
Any postsecondary	5%	10%
Program enrollment and service receipt		
RP grantee type		
CBO	52%	63%
Intermediary	48%	37%
RP program type		
Adult	53%	55%
Young adult	47%	45%
Service receipt		
Received training or education services	70%	73%

Exhibit III.2 (*continued*)

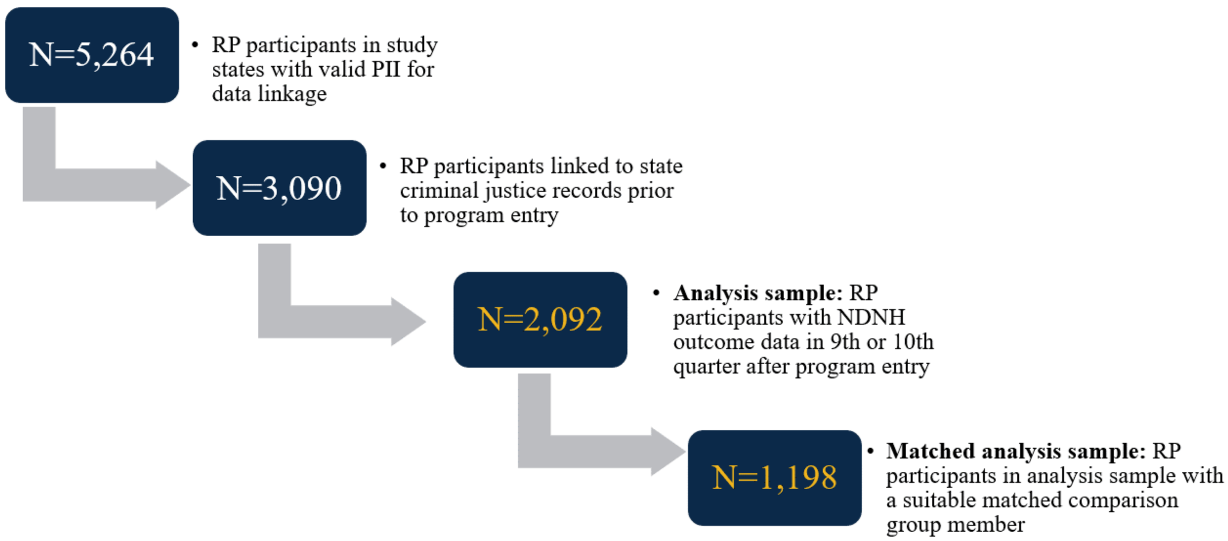
Characteristic	Study RP participants	All other RP participants
Received occupational skills training	41%	42%
Sample size	3,090	15,650

Source: WIPS data.

Note: Percentages may not sum to 100 within categories due to rounding.

This sample of RP participants served as the basis for our matching design, which we describe more formally in Section D. Exhibit III.3 provides a summary overview of the final analysis sample of RP participants and the sample inclusion criteria.

Exhibit III.3. RP sample inclusion criteria and sample sizes



NDNH = National Directory of New Hires; PII = personally identifiable information; RP = Reentry Project.

C. The Wagner-Peyser program

To estimate the impact of enrollment in RP on individual employment, earnings, and criminal recidivism, we constructed a comparison group consisting of participants in Wagner-Peyser employment services programs. Intuitively, Wagner-Peyser represents an alternative to RP, providing lighter-touch services to people who, like those who enroll in RP, request help in securing employment. In the absence of RP, many individuals with criminal justice involvement looking for assistance finding employment may very well go to AJC or otherwise enroll in Wagner-Peyser. We therefore argue that this comparison between RP and Wagner-Peyser participation captures both real variation in the intensity of services provided as well as a plausible counterfactual that approximates what our sample of "treated" RP participants would have done had they not taken part in the RP program.

Critically, Wagner-Peyser provides basic career services, unlike RP. As outlined in DOL guidance, such services are "universally accessible and must be made available to all individuals seeking employment and training services in at least one comprehensive AJC per local area" (TEGL 19-16 2017). These services are described as light touch, and they often include the following:

- **Job search and placement services**, including counseling, labor market information, assessment, and referral to employers
- **Recruitment services** for employers to help fill vacancies
- **Reemployment services** for individuals receiving unemployment compensation
- **Access to the state's labor exchange**, which includes open job orders (DOL 2023)

Wagner-Peyser services can be grouped into three categories, in increasing order of intensity: basic career services, individualized career services, and training or education. As Spitzer et al. (2023) outline, in most states nearly all Wagner-Peyser participants receive basic career services; a significant minority receive individualized career services, while fewer than 10 percent typically receive training. In general, although available services vary across states, Wagner-Peyser participants are not typically able to access additional ongoing case management or supportive services through the Wagner-Peyser program. This light-touch model contrasts with the wraparound services provided through RP.

D. Constructing the comparison group

To construct our comparison group, we wanted to identify a set of Wagner-Peyser participants with observably similar characteristics to our RP participant sample. However, Wagner-Peyser serves a much larger and broader set of individuals than RP, and most Wagner-Peyser participants would not represent suitable matches for any of the RP participants in our sample.

To narrow down the list of prospective comparison group members, we conducted a first stage of matching, with the goal of tailoring our focus to Wagner-Peyser participants who shared key demographic characteristics with our sampled RP participants. Specifically, for each RP participant, we identified Wagner-Peyser participants who resided in the same county and who enrolled in a Wagner-Peyser program in the same year and quarter that the RP participant enrolled in their program. We then used coarsened exact matching (CEM) to match every RP participant to Wagner-Peyser participants with whom they shared demographic characteristics at enrollment, including age, gender, race/ethnicity, education level, employment status at program enrollment, receipt of dislocated worker services, English learner status, veteran status, and disability status. We matched without replacement, meaning that each prospective Wagner-Peyser comparison group member could match to, at most, one RP participant. This process left us with 254,553 Wagner-Peyser participants with observably similar geographic and demographic characteristics as our sampled RP participants.

We sent this list of first stage matched Wagner-Peyser participants to state criminal justice agencies for linking based on name, date of birth, and/or Social Security number. Using these data, we further restricted our sample to Wagner-Peyser participants for whom we observed past criminal charges in state court data. Again, unlike RP, Wagner-Peyser does not exclusively serve individuals with a pre-program history of criminal justice involvement; because of program eligibility requirements, most RP participants have been accused (if not convicted) of a criminal offense, and because justice-involved individuals face unique barriers to employment, we wanted to ensure that our comparison group shares this characteristic with our program group. Therefore, we excluded any Wagner-Peyser participants from our sample who did not have a criminal case filed against them between 2013 (the first year in which we observed criminal justice outcomes for our sample) and program enrollment, according to court records from the state in which they enrolled in Wagner-Peyser.[11] This restriction yielded a pool of 64,602 prospective Wagner-Peyser comparison group members.

For our final comparison group, we used an empirical approach that combines partial exact matching with caliper matching based on estimated propensity scores, as in Austin (2011) and Iacus et al. (2012). That is, we specified a set of criminal justice, DOL program characteristics, and person-level demographic characteristics that we wanted to match exactly between treated and prospective comparison individuals. We then estimated propensity scores—the probability that an individual in our sample enrolled in RP, as opposed to Wagner-Peyser—to identify the most similar treated and comparison group members within strata defined by our exact-match variables.

We chose variables for exact matching that correspond to characteristics known to be strongly correlated with labor market outcomes and those that we expected to generate the most comparable groups. We began by exactly matching young adults (ages 18 to 24) and adults (ages 25 and over), which reflects the fact that RP eligibility criteria differed for those groups. Beyond age group, we also exactly matched RP and Wagner-Peyser participants based on their state, quarter, and year of enrollment in RP or Wagner-Peyser, which together capture labor market conditions at the time of enrollment as well as the degree of pre- and post-program exposure to the COVID-19 pandemic.[12] We matched on other salient individual characteristics: the person's gender, and an indicator for whether they were employed at the time of program entry (from the WIPS). Finally, to ensure our program and comparison groups each had similar degrees of pre-program criminal justice involvement, we exact-matched on the following key justice-related variables: whether the person was convicted in their most recent criminal case, and whether they entered their employment services program within 3 quarters of a criminal case disposition or a release from state prison.[13]

We supplemented this exact-match component of our design with propensity scores that enabled us to further refine our comparison. For each RP participant and potential Wanger-Peyser comparison individual, we estimated the probability that they would enroll in RP, rather than Wagner-Peyser, based on an array of individual-level demographic characteristics and features of their pre-program criminal justice

[11] We applied a similar restriction to our program group of RP participants, such that our final sample only included individuals who faced criminal charges before enrolling in their employment services program. Note that although the WIPS contains an indicator for ex-offender status, the field is largely missing and we determined it would be more accurate to reference state court records.

[12] We explored the possibility of conducting an exact match on county of program enrollment, which would better capture local labor market conditions. However, we determined that small sample sizes at the county level resulted in worse-quality matches on average than those we achieved by matching at the state level. Estimating our propensity scores (which we discuss below) separately by county created a similar problem. As such, we chose to pool our sample by state and include county characteristics in the propensity score estimation.

[13] These latter variables capture the fact that adult eligibility criteria for RP depends on the time elapsed since an individual's release or probation sentencing. To the extent possible, we wanted to control for differences in time elapsed between a person's most recent criminal justice contact and their enrollment in the program, which might be correlated with their ability to secure a job.

involvement.[14] We considered various approaches to modeling these propensity scores, which we summarize in Section B of the Technical Appendix.

Ultimately, after assessing the relative performance of these approaches, we opted to use a least absolute shrinkage and selection operator (LASSO) regression. This approach helped winnow down our extensive list of potential covariates to focus just on those with the most explanatory power. We estimated our propensity score model separately for adults and young adults to capture differences in RP eligibility criteria across these groups. To arrive at our final analytic sample, we took the propensity scores estimated via LASSO and applied a caliper, a maximum distance in propensity scores between candidate matches. That is, we matched each RP participant to any candidate Wagner-Peyser participants who shared their exact-matched characteristics and whose propensity scores fell within a given "caliper" distance of their own. Our final comparison group consisted of the 16,032 Wagner-Peyser participants whom we successfully matched to at least one RP participant based on our exact-match and caliper match criteria. In our primary approach, and in all alternative matching methods that we explored, a large portion of RP participants in the analytic sample were dropped from the final matched sample due to not having quality Wagner-Peyser matches. We present sample sizes for each matched sample in the Technical Appendix.

By combining partial exact matching with caliper matching, we promoted better matches along the key pre-program variables selected, at the cost of shrinking the overall sample size. Nonetheless, as we discuss in the next section, our approach resulted in a well-balanced sample, and our matching criteria accounted for the justice backgrounds of our sample as well as salient demographic characteristics.

E. Sample balance

Our matched comparison design aimed to construct a sample of RP and Wagner-Peyser participants who, before program entry, had similar characteristics, such that the only observable difference between them was the employment program in which they participated. A more balanced sample, in which our treated and comparison individuals appeared relatively similar on average, would lend more support to a causal interpretation of our findings—that is, differences in post-program outcomes could more plausibly be attributed to RP, rather than any systematic pre-program differences in participant backgrounds. An imbalanced sample, by contrast, might indicate that we did not choose a suitable comparison group.

As with any matching design, we could not observe all possible features of our sample members, and thus cannot guarantee balance on every potential pre-program characteristic. Instead, by showing balance on a wide variety of observable characteristics, we implicitly assume balance along any other unobservable characteristics that might confound our estimates. For example, one might suspect that the average RP participant—who, in general, was offered services directly by grantees—might have less motivation compared to the average Wagner-Peyser participant, who had to seek out services for themselves.

[14] Although we preferred to include pre-program earnings and employment as part of our matched comparison design, we lacked the data to do so. The NDNH, from which we obtained individual labor market records, only includes two years of past data. Because our analysis began in 2021, we cannot observe pre-program earnings and employment for most of our sample, who enrolled in RP or Wagner-Peyser as early as 2018. In Chapter V, and in Section C of the Technical Appendix, we explore how including pre-program data affects our results among the subset of our sample for which we do observe at least two years of pre-enrollment labor market history.

Empirically, we have no way of measuring people's motivation; rather, we argue that because we achieved balance along characteristics likely correlated with motivation, such as education level, it is plausible that we also achieved balance along this unobservable dimension.

Beyond intangible characteristics like motivation, we note that we cannot observe—and thus cannot assess balance along—at least two pertinent aspects of individuals' backgrounds. We highlight these limitations here to provide context to our sample balance statistics, although we discuss both in detail in Section H (Limitations), and again in Chapter 4 when we interpret our final estimates.

- **Lack of data on reasons for RP eligibility**. As stated, RP eligibility depended in large part on participants' prior criminal justice involvement. Adults (ages 25 and over) must have been incarcerated or subject to supervised probation; most young adults (ages 18 to 24) could have had any prior criminal justice involvement, including juvenile justice contacts.[15] We do not have the juvenile justice records, local jail data, or criminal sentencing information that would enable us to infer reasons for eligibility for either our program or comparison groups. In practical terms, these missing data mean that we cannot restrict our sample of potential Wagner-Peyser comparison group members to just those with similar prior criminal justice involvement as RP participants. Therefore, we cannot assess whether program and comparison group members were incarcerated or served supervised probation—although we can observe characteristics of the most recent pre-program criminal case that might be correlated with sentencing, such as whether the case included a felony offense and whether the person was a repeat offender who had a history of other previous criminal cases.
- **Lack of data on pre-program earnings and employment.** Because the NDNH only holds two years of records at a given time, we did not have sufficient pre-program data on earnings and employment to use as part of our matching approach, a fact we laid out in our initial design report (DOL 2023). However, one might expect that, because RP participants were more likely to have been incarcerated or on probation in the lead up to their enrollment, compared to the average Wagner-Peyser participant, our program group would have lower earnings and a lower employment rate pre-program. In the absence of reliable data on either pre-program incarceration status or pre-program labor market outcomes, we could not explore the possibility that RP participants may have been on a different earnings trajectory than Wagner-Peyser comparison group members, which would bias our estimated differences in post-program outcomes.

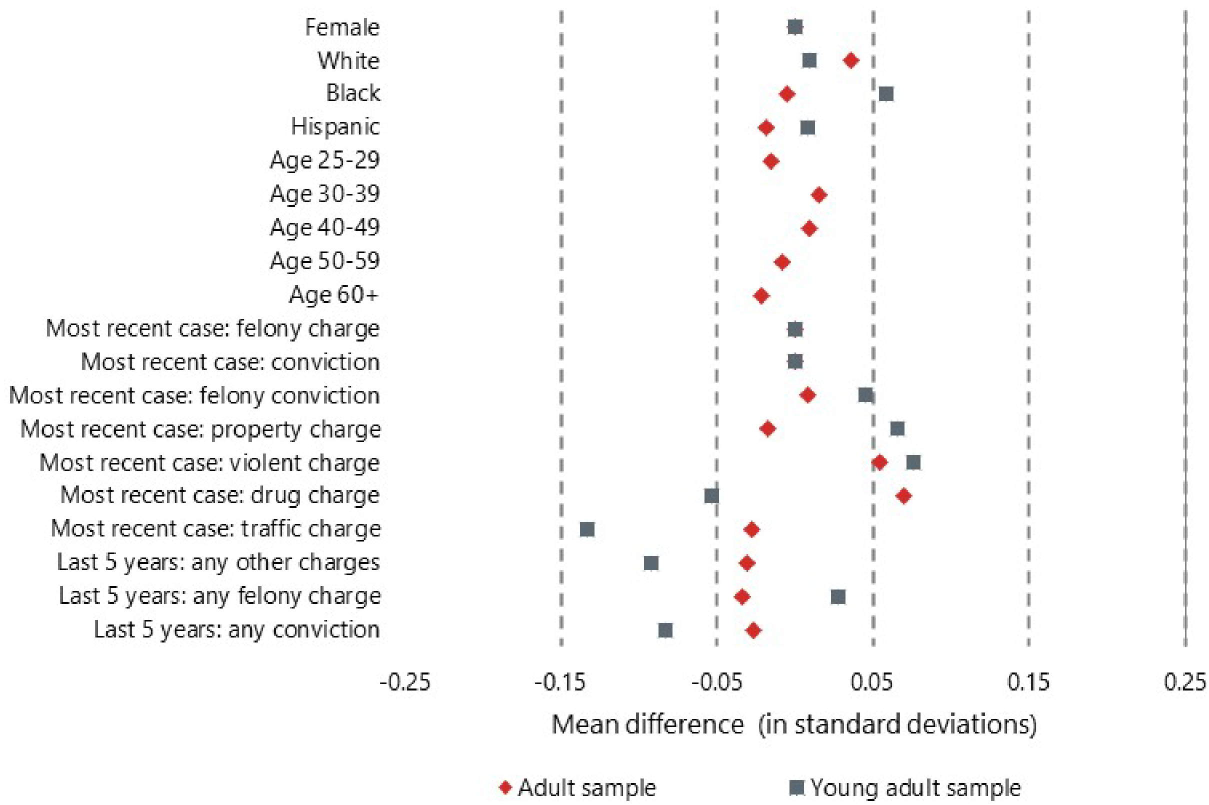
To assess the balance of our matched sample of RP and Wagner-Peyser participants along observable characteristics, we compared the features of the demographics and backgrounds of these two groups along with details of their pre-program criminal justice involvement. We then standardized the differences between these group means by expressing them in standard deviation units, a step that ensures the sample size does not influence our conclusions, as it would for other common statistical tests (Austin 2009). Following U.S. Department of Education's What Works Clearinghouse best practices, we considered any standardized mean differences between -0.25 and 0.25 to be acceptable and able to be controlled for

[15] Our implementation study found that some RP grantees had trouble meeting their enrollment targets and may have relaxed their eligibility rules during the COVID-19 pandemic. We cannot verify the extent to which individuals who did not meet the stated eligibility criteria enrolled in RP, but their presence in the data—and our inability to control for (in)eligibility status—marks another potential for mismatches between program and comparison group members.

as part of our regression estimation approach (What Works Clearinghouse 2022), which we discuss in the next section.

Exhibit III.4 presents the standardized mean differences in a selection of salient pre-program demographic and criminal justice background characteristics across our RP and constructed comparison Wagner-Peyser samples. We did not find any meaningful differences in these salient characteristics—nor in any other characteristics that we measured—across our program and comparison samples, with standardized differences all well below 0.25 standard deviation units.

Exhibit III.4. Standardized mean differences in characteristics of Reentry Project (RP) participants in the impact study relative to the matched Wagner-Peyser participants



Source: Workforce Integrated Performance System data matched to state criminal court records.

Note: The sample includes 664 adult RP participants and 534 young adult RP participants. All members of the young adult sample are between 18 and 24 years old, while none of the members of the adult sample are less than 25 years old.

F. Methods for estimating impacts

The goal of our impact study is to uncover the average effect of RP participation on those who took part in the program, also known as the average treatment effect on the treated (ATT). We estimated the ATT using a series of regression specifications, including as control variables the same demographic, program, and criminal justice history variables that we used to estimate our propensity score model. This doubly robust approach provides some assurance that our impact estimates do not capture effects stemming from imbalance in the observable pre-program characteristics shown in Exhibit III.2 (Funk et al. 2011).

We estimated these regression models using weighted least squares or, for binary outcomes such as employment status, weighted linear probability models. In these models, each program group member received an equal weight and each comparison group member received a weight proportional to the number of program group members they matched to.[16] Ultimately, of the 3,090 RP participants from the six states in our impact analysis sample, we successfully matched 1,198 to at least one Wagner-Peyser participant after applying these criteria.

We performed additional analyses to confirm that our choice of matching estimation approach did not drive our findings. As discussed above, to explore the sensitivity of our results to our choice of propensity score matching model, we re-estimated effects on our primary outcomes using three alternative approaches: Bayesian additive regression trees (BART), generalized boosted regression model (GBM), and a logistic regression with researcher-selected (rather than LASSO-selected) covariates. We also explored the sensitivity of our findings to an alternative nearest-neighbor—as opposed to caliper-based—approach to matching on propensity scores, whereby we matched each RP participant to the single Wagner-Peyser participant with the nearest propensity score and same exact-match characteristics. Lastly, we narrowed the caliper we used to screen prospective matches to 0.1 standard deviation units, which addresses potential concerns about our tolerance of matches with relatively large gaps in their propensity scores.

We further probed the sensitivity of our findings to a different regression framework and the inclusion of additional data to refine our comparison. First, to explore how our estimated effects on binary outcomes depended on our weighted linear probability model, we also estimated weighted logistic regressions. Second, we examined whether our inability to control for pre-program labor market outcomes biased our impact estimates. Specifically, although most of our sample enrolled in their training programs earlier than we can observe them in the NDNH, individuals who enrolled in RP or Wagner-Peyser in 2021 or later (PY 2021 Q2 or later) did have at least some observable pre-enrollment earnings reported in the NDNH. We used these data as part of our matching design, narrowing in on individuals who were on similar earnings trajectories before RP or Wagner-Peyser participation. Finally, we considered whether our impact estimates understated the potential effects of RP by excluding Wagner-Peyser comparison group members who received relatively intensive employment services, focusing on those who received light-touch services. This analysis helped identify how comparison group members who obtained more intensive services similar to those that RP provided influenced our results.

G. Outcomes

To evaluate our primary research questions, we tracked individuals' employment and earnings up to 10 quarters following their enrollment in RP or Wagner-Peyser, alongside several metrics of criminal recidivism. Our data on labor market outcomes come from the NDNH, while our measures of criminal recidivism come from state criminal justice records.

[16] Specifically, for each RP participant, we first calculate a parameter defined as 1 divided by the total number of matches found for that program group member. We then define the comparison group member weight as the sum of this parameter across all of their matches. For example, consider a comparison group member who was matched to two program group members, one with 10 matches and one with five. This comparison group member would be assigned a weight of 0.1 for the first match and 0.2 for the second match, for a total weight of 0.3.

The outcomes for our confirmatory research questions are indicators for whether an individual received a new criminal conviction in the 10 quarters after enrollment; a measure of their employment in the 9th and 10th quarters after program enrollment; and their average earnings during the 9th and 10th quarters after enrollment. To address our exploratory research questions, we evaluated the same outcomes as in our confirmatory analysis but focused on the 4th and 5th quarters following program enrollment. We also broadened our definition of criminal recidivism to include other outcomes, including arrest, incarceration, and the number of new criminal cases.

Exhibit III.5. Outcome measures for Reentry Project (RP) impact study confirmatory and exploratory research questions

Outcome	Description
Confirmatory outcomes	
Any criminal conviction during the 10 quarters following program enrollment	Indicator for whether the person had any criminal conviction during the 10 quarters following enrollment in RP or Wagner-Peyser
Employment in the 9th and 10th quarters following program enrollment	Measure of individual employment in the 9th and 10th quarters following enrollment; can equal 1 (employed in both quarters), 0.5 (employed in 1 quarter), or 0 (employed in neither quarter)
Average earnings in the 9th and 10th quarters following program enrollment	Average earnings reported in the 9th and 10th quarters following enrollment in RP or Wagner-Peyser
Exploratory outcomes	
Any arrest or incarceration in the 10 quarters following program enrollment	Indicators for whether the person had any arrest or period of incarceration during the 10 quarters following enrollment in RP or Wagner-Peyser
Employment in the 4th and 5th quarters following program enrollment	Measure of individual employment in the 4th and 5th quarters following enrollment; can equal 1 (employed in both quarters), 0.5 (employed in 1 quarter), or 0 (employed in neither quarter)
Earnings in the 4th and 5th quarters following program enrollment	Average earnings reported in the 4th and 5th quarters following enrollment in RP or Wagner-Peyser
Any conviction, arrest, or incarceration during the 5 quarters following program enrollment	Indicators for whether the person had any criminal conviction, arrest, or period of incarceration during the 10 quarters following enrollment in RP or Wagner-Peyser
Total new criminal cases over the 5 and 10 quarters following program enrollment	Total number of new criminal cases brought against the person during the 5 and 10 quarters following enrollment in RP or Wagner-Peyser
Total new criminal cases involving a felony offense over the 5 and 10 quarters following program enrollment	Total number of new criminal cases that involve at least one felony charge brought against the person during the 5 and 10 quarters following enrollment in RP or Wagner-Peyser

H. Limitations

As noted above, data availability limited the RP grantees and participants that we could include in our analysis to about 17 percent of all RP participants, all drawn from only six of the 34 states in which the program was operating nationwide. As such, the findings that we report in this study may not represent the impact of RP among all participants nationwide, given that states differ in labor market conditions and policy (particularly towards individuals with prior criminal convictions). Similarly, our study focused on people who took up employment services in the lead-up to and during the COVID-19 pandemic, which may influence our findings. The net effect of the pandemic on our results is unclear: the pandemic labor market may have increased the relative value of RP services, in which case we would overestimate the impact of the program, or it might have depressed the labor market benefits of the additional services RP provided, in which case we would underestimate the impact of the program. Moreover, our empirical approach controlled for time of program entry, which means we compared individuals who had the same exposure to the pandemic labor market, which should have mitigated any bias.

A second set of limitations concerns our choice of comparison group. We stress that our estimates specifically capture the *relative* difference in outcomes for RP participants and Wagner-Peyser participants. Our results do not estimate the differences associated with receiving RP services relative to receiving no employment services at all. That is, we may have realized a greater contrast between the program and comparison conditions, and potentially larger differences in outcomes post-program, had we conducted a randomized control trial (RCT) or other design with a true control condition. Still, as we emphasized earlier in this section, we believe that Wagner-Peyser represents an informative and realistic counterfactual for what RP participants might have experienced had they not enrolled in RP. Indeed, individuals who are not selected into the treatment groups of many RCTs find training elsewhere and end up “treated” anyways, like Wagner-Peyser participants (Fortson et al. 2017). Critically, Wagner-Peyser participants exhibited sufficient motivation to pursue employment services, a key differentiator of RP participants from the population of justice-involved individuals and a principal reason why constructed our comparison group from Wagner-Peyser participants ..[17] Still, we stress that our results capture differences in outcomes between participants in two programs, and we explore the ramifications of our findings through that lens.

Our efforts to construct a truly comparable matched comparison group might have also been impaired by particular missing data elements. Although we aimed to construct a sample of Wagner-Peyser participants who appear similar to RP participants along observable dimensions, we may not capture differences in unobservable factors that could bias our estimates. Once again, we specifically highlight two critical pieces of missing data that, had they been available, might have improved the quality of our matches. Due to NDNH data retention practices, we could not obtain pre-program earnings and employment records for over 90 percent of our analytic sample. Without these pre-program labor market records, we cannot say for certain whether post-program differences in earnings and employment reflect preexisting gaps between RP participants and Wagner-Peyser participants, or whether those differences represent program effects. Similarly, although we collected detailed criminal justice records from our sampled states, we lack reliable data on juvenile offenses, local jail stays, and sentencing (to either probation or incarceration) that would enable us to accurately capture reasons for RP eligibility or the circumstances of recent justice involvement in our sample. As such, we might have compared individuals with observably similar justice contacts (for example, two people with misdemeanor convictions) that are actually quite distinct (for

[17] Unobserved motivation, as we noted earlier, may affect our findings. However, we note that we have no good reason to expect one group to have systematically more motivation than the other. For instance, one might argue that Wagner-Peyser participants may have had greater unobservable motivation because they sought out services rather than being recruited or compelled to enroll by the criminal justice system. On the other hand, RP participants could be more motivated because their programs may have implemented more explicit screening criteria, for example by assessing participants willingness to commit to program activities.

example, of the two people with misdemeanor convictions, one might have resulted in jail time, while the other received a fine).

These omissions pose a particular concern given our policy setting. By design, adult RP participants had recently experienced incarceration or supervised probation, both of which likely limited their employability and placed them at risk of re-offending. By contrast, Wagner-Peyser participants need not have been sentenced, much less incarcerated to participate in the program: our sample construction ensures that they faced criminal charges, but we have limited insight into the consequences of those charges. As such, we might expect that the average RP participant had more serious prior criminal justice involvement, lower earnings and probability of pre-program employment, and a higher likelihood of committing future criminal offenses than even observably similar Wagner-Peyser comparison group members. Although we cannot know for sure how our imperfect pre-program data affected our estimates, Chapter 4 offers descriptive analyses that point to the consequences of these missing data.

IV. Impacts on Convictions, Employment, and Earnings

This section presents our estimates of the impact that RP participation had on participants' future criminal convictions, employment, and earnings. We first specify the research questions that guided our analysis. We then discuss our impact estimates on recidivism, measured as new criminal convictions over the 10 quarters following program enrollment, before describing the estimated impact of RP on employment rates and quarterly earnings in the 9th and 10th quarters following enrollment. To help contextualize our results, we describe the implications of the empirical challenges we faced in constructing a suitable comparison group, as well as the potential consequences of limitations in the underlying data. We conclude by presenting subgroup analyses that highlight variation in estimated impacts across policy-relevant subpopulations of RP participants and discuss additional sensitivity checks we used to assess the robustness of our findings.

Key findings

- RP participants were 5.1 percentage points more likely to have a new criminal conviction in the 10 quarters after program entry compared to Wagner-Peyser matched comparison group members.
- In the 9th and 10th quarters after enrollment, RP participants were 4.1 percentage points less likely to be employed than comparison group members.
- RP participants earned $693 less in the 9th and 10th quarters after enrollment compared to matched Wagner-Peyser participants, who earned $2,937 on average during that period.
- Estimated impacts differed based on the severity of individuals' pre-program criminal justice involvement, with RP participants who had more serious prior justice involvement showing no statistically significant differences in outcomes compared to similar matched Wagner-Peyser participants.
- These patterns in impacts based on pre-program criminal justice involvement may reflect unobserved differences between RP participants and matched comparison group members in pre-program sentencing associated with recidivism.

A. Research questions

Our analysis centered on answering three primary questions about the effects of RP on participants' criminal justice and labor market outcomes. In Exhibit IV.1, we list these confirmatory research questions, as well as additional exploratory questions that we investigated to shed additional light on how RP affected participants. We pre-specified these questions in our design report (DOL 2023).

Exhibit IV.1. Impact study research questions

Question #	Research questions
Confirmatory research questions	
Compared with Wagner-Peyser employment services, what was the impact of RP on the following outcomes:	
C.1a	Probability of being convicted of a new offense over the 10 quarters after enrollment
C.1b	Employment rates in the 9th and 10th quarters after enrollment
C.1c	Average earnings in the 9th and 10th quarters after enrollment

Exhibit IV.1 (*continued*)

Question #	Research questions
Exploratory research questions	
Compared with Wagner-Peyser employment services, what was the impact of RP on the following outcomes:	
C.2a	Arrest and incarceration rates over the 10 quarters after enrollment
C.2b	Employment and earnings outcomes in the 4th and 5th quarters after enrollment
C.2c	Probability of being convicted, arrest rates, and incarceration rates over the 5 quarters after enrollment
C.2d	Frequency and severity of criminal justice outcomes over the 5 and 10 quarters after enrollment
How did the estimated impact of RP on future convictions, employment, and earnings differ across the following subgroups:	
C.3a	Adult and young adult participants
C.3b	Participants of different races and ethnicities
C.3c	Participants of different gender
C.3d	Participants with lower versus higher frequency of prior criminal justice involvement
C.3e	Participants served by grantees with different strategies or other characteristics uncovered by the implementation study
C.3f	Participants who received different types of services

Note: Any participant who enrolled in RP or Wagner-Peyser programs are considered to have participated, regardless of the services received.

Confirmatory research questions describe the primary analyses, which assessed whether RP participation affected principal outcome measures.

RP = Reentry Project.

B. Estimated impacts on future conviction, employment, and earnings

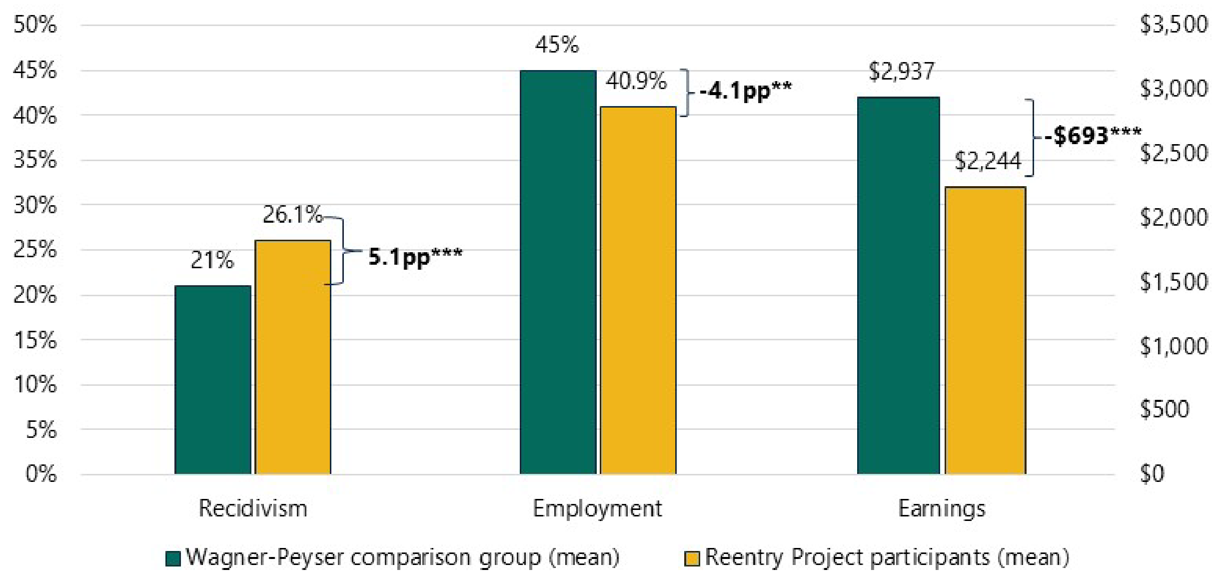
This study estimated the impact of RP on participants' criminal recidivism and labor market outcomes. We measured impacts along these margins 10 quarters after program enrollment, which allowed for sufficient time for program effects to materialize. To capture participant recidivism, we examined whether individuals had any conviction for a new criminal offense over the 10 quarters post-enrollment. We derived this outcome from criminal case data collected from state courts. To measure labor market outcomes, we calculated participants' employment and earnings during the 9th and 10th quarters post-enrollment, drawing on data from the NDNH. By compiling labor market outcomes over 2 quarters, we hoped to avoid capturing quarter-to-quarter fluctuations in labor market outcomes.

We first considered how recidivism rates differed between RP participants and Wagner-Peyser comparison group members to provide context for understanding our findings on labor market outcomes. The labor market prospects of individuals with recent criminal justice involvement hinge on their re-offending behavior. For instance, recidivism could have an incapacitation effect, removing people from the labor force and placing them in jail or prison. Future criminal convictions might also hurt individuals' employment prospects by adding to their criminal record, which could limit their attractiveness to employers, regardless of whether the new criminal charges lead to incarceration. Given that 40 percent of recently released individuals are re-incarcerated within a year (Bureau of Justice Statistics 2021), the close relationship between future conviction, employment, and earnings cannot be ignored in our sample, which consisted entirely of people who previously faced criminal charges and thus have an elevated risk of future conviction.

As discussed in Chapter 3, our ability to accurately measure the post-program impact of RP on criminal justice involvement depended on our ability to accurately measure pre-program criminal justice involvement. We confronted several limitations in this regard, which we describe in the last chapter. In particular, our inability to observe incarceration spells in local jails, probation status, or supervision status upon release—all of which are critical determinants of eligibility for RP services, particularly for adults—might have impaired our ability to realize ideal matches between program and comparison group members. Although we designed our matching approach to incorporate the array of criminal justice data at our disposal—including prior convictions, felony charges, and incarcerations in state prison—we could not assess, much less ensure, balance along dimensions that we could not observe, like prior jail stays and probation conditions. To the extent that underlying differences between these samples along these dimensions were correlated with program eligibility, prior criminal justice involvement, and future criminal justice involvement, these missing data might have affected our findings. We return to this point in Section C when we interpret our estimates.

Our primary findings appear in Exhibit IV.2. Overall, we found that RP participants were more likely to be convicted of a future criminal offense than Wagner-Peyser comparison group members and had lower employment and earnings in the 9th and 10th quarters post-enrollment. Our exploratory findings, which we present in Appendix Exhibit A.16, show qualitatively similar, though sometimes smaller, impacts on short-term outcomes and other measures of criminal recidivism.

Exhibit IV.2. Impacts of Reentry Project on recidivism, employment, and earnings



Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Notes: Employment is defined as having any earnings in a given quarter. Wagner-Peyser group means are unadjusted; Reentry Project group means are adjusted means equal to the Wagner-Peyser group mean plus the estimated impact. For a detailed description of estimation methods, please see the Technical Appendix.

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; WIPS = Workforce Integrated Performance System.

In the remainder of this section, we report our impact results for each of our primary outcomes—recidivism, employment, and earnings—across our whole sample as well as specifically for adults (ages 25 and up) and young adults (ages 18 to 24) in the program and comparison groups. As noted in Chapter 3, breaking down the sample this way reflects the fact that RP eligibility criteria differed for adults and young adults, which led us to conduct our matching process separately for each subsample. These findings by age group appear in Exhibit IV.3. In subsequent sections, we contextualize and discuss the caveats behind these estimates before describing results from subgroup analyses.

Recidivism. Relative to Wagner-Peyser comparison group members, we found that RP participants were 5.1 percentage points more likely to recidivate and have a new criminal conviction in the 10 quarters after program entry. Specifically, during this period, 21 percent of the comparison group were convicted of a new offense, while 26.1 percent of RP participants were convicted of a new offense. , This gap means that RP participants were 24 percent more likely of being convicted for a new offense over the 10 quarters after program entry than matched Wagner-Peyser comparison group members. Quantitatively, impacts for adults and young adults appeared similar: adult RP participants were 4.4 percentage points more likely to be convicted in the 10 quarters after program entry than matched comparison group members, while young adult RP participants were 5.1 percentage points more likely, though the effect on young adults was not statistically significant.

Employment. Our point estimates indicated that RP participants had employment rates 4.1 percentage points lower than the comparison group, 44 percent of whom were employed, on average, in the 9th and 10th quarters post-enrollment. We found no statistically significant differences in employment rates among adults. By contrast, young adult RP participants were 7.3 percentage points less likely to be employed in the 9th and 10th quarters after program entry than comparable Wagner-Peyser participants.

Earnings. RP participants had lower earnings compared to Wagner-Peyser comparison group members, who earned an already-low $2,937 on average across the 9th and 10th quarters after program enrollment. Specifically, RP participants earned $693 less, on average, during this period. Among young adults, RP participants earned $1,107 less, on average, than matched Wagner-Peyser participants; by contrast, adults RP participants earned $403 less, on average, than matched Wagner-Peyser participants.

Exhibit IV.3. Impact of Reentry Project on confirmatory outcomes

Outcome	Comparison mean	Impact estimates		
		Full sample	Adults	Young adults
Any new conviction during 10 quarters after enrollment	21%	5.1pp*** (1.5pp)	4.4pp*** (1.7pp)	5.1pp (2.6pp)
Avg employment in the 9th and 10th quarters after enrollment	45%	-4.1pp** (1.7pp)	-1.7pp (2.1pp)	-7.3pp*** (2.8pp)
Avg earnings in the 9th and 10th quarters after enrollment	$2,937	-$693*** ($144)	-$403** ($189)	-$1,107*** ($227)
RP sample size	N/A	1,198	664	534
WP sample size	16,032	16,032	14,718	1,314

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Exhibit IV.3 (*continued*)

Notes: Standard errors appear in parentheses below impact estimates. Employment is defined as having any earnings in a given quarter. For a detailed description of estimation methods, please see the Technical Appendix.

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; WP = Wagner-Peyser; RP = Reentry Project; WIPS = Workforce Integrated Performance System; NDNH = National Directory of New Hires.

C. Interpreting the impact estimates

Our impact estimates indicate that RP participants experienced worse criminal justice and labor market outcomes compared to Wagner-Peyser participants. These findings support potentially overlapping interpretations. On the one hand, RP might lead to poorer outcomes for participants, relative to Wagner-Peyser. On the other hand, our estimation approach might fail to address underlying differences that generate worse outcomes for RP participants. In this section, we weigh the plausibility of our findings by situating our impact estimates in the wider literature on post-incarceration employment services. We then explore potential sources of bias, highlighting limitations of our data and research design that might have shaped our results.

Contextualizing our findings in the literature. At least two past RCTs found substantial adverse impacts of reentry services programs on participant recidivism. Specifically, Wiegand and Sussell (2016) report that participation in RExO programs resulted in a 21 percent increase in the probability of future criminal convictions. Although the authors stress that this effect might be spurious, their estimate is smaller, but in line with our reported 24 percent difference in future convictions between RP and Wagner-Peyser participants. Similarly, D'Amico and Kim (2018) evaluated the Second Chance Act (SCA) and found similar rates of future conviction, but an overall increase in the number of new convictions, for program participants compared to control group members over the 30 months after program enrollment.

Both of these prior studies provide experimental evidence that reentry services might increase reoffending behavior, and thus these past findings suggest that it is plausible for Reentry Project participation to have had an adverse effect on recidivism, insofar as these examples provide experimental evidence that reentry services might increase reoffending behavior. Nonetheless, in the context of the existing evidence, our results are outliers. Cortina et al. (forthcoming) conducted a review of the evidence on the effectiveness of similar reentry service programs and documented mixed impacts of other reentry services programs—although typically, well-identified studies based on RCTs report either null effects of reentry services on employment and recidivism, or some benefits on these outcomes, at least in certain cases.

As such, although the notion that reentry services may not be effective at preventing future crime is not new (see, for example, Doleac 2019), our results still stand out from prior research. In particular, the magnitude of our estimates on recidivism are unusual: Wiegand and Sussell's (2016) RExO study reports the only comparably adverse effects on future conviction that we are aware of, and do so in an experimental context in which program and control group members would have had similar risks of recidivating. Moreover, the fact that we find substantially lower earnings for RP participants relative to our comparison group is, to our knowledge, unique in the literature. For instance, D'Amico and Kim (2018), who report mixed or adverse effects of reentry services on recidivism, still found that participants in SCA programs went on to earn over $1,000 more per quarter than control group members. Likewise, studies of transitional jobs programs, such as Valentine and Redcross (2015) and Cook et al. (2015), report short-term gains in employment-related outcomes, often attributable to program-provided positions. Although these employment effects dissipated over time, neither study found negative program impacts.

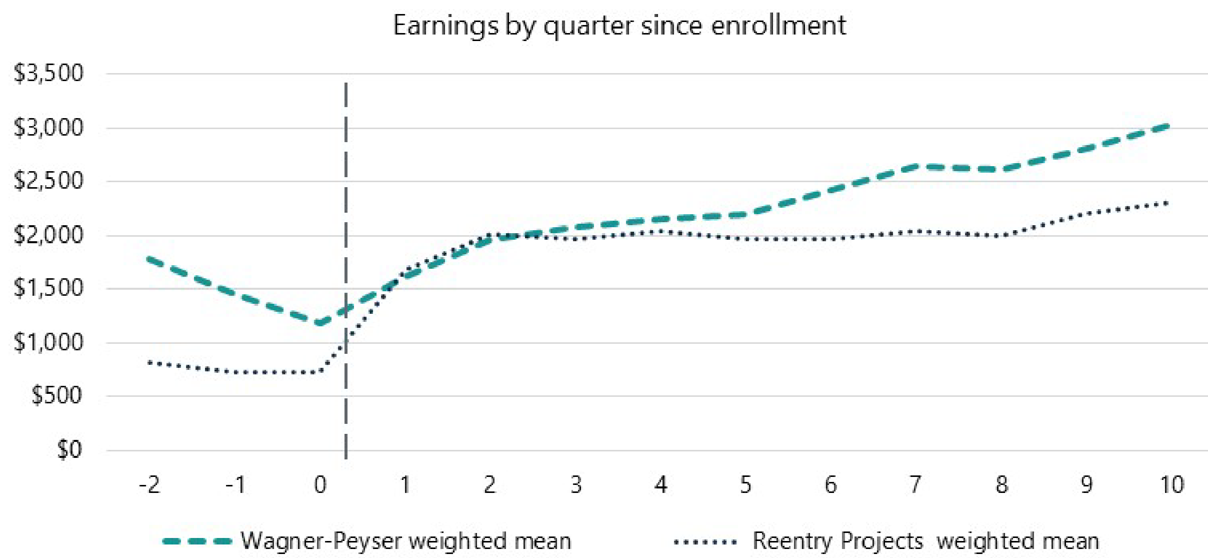
Potential bias from pre-program differences in labor market outcomes. Intuitively, our matched comparison design constructed a group of Wagner-Peyser participants with observably similar characteristics to our RP program group members. In Chapter 3, we noted that we achieved a high degree of balance on observable pre-program characteristics. For our matching framework to deliver unbiased estimates, we must implicitly assume that, because we produced balance along a wide array of characteristics that we could measure, we also achieved balance along characteristics that we could *not* measure—including, but not limited to, pre-program incarceration and probation status, as well as employment and earnings.

To underscore the implications of this assumption, we illustrate the average earnings of our RP and matched Wagner-Peyser comparison groups by quarter since (or leading up to) program enrollment. This analysis leverages the small share of individuals (162 RP participants out of the 1,198 in our matched sample) for whom we observed at least 1 quarter of pre-program earnings. As noted in Chapter 3 and laid out in our pre-specification plan, this small sample size precluded using these results as part of our primary analysis. Instead, we view this analysis as an exploratory exercise to examine potential pre-program differences in labor market outcomes that might generate bias, rather than as concrete findings.

With these caveats in mind, we present trends in earnings for our program and comparison groups in Exhibit IV.4. The dashed lines represent the weighted earnings of our program and comparison groups, where the weights are given by our matching procedure. Among the small share of our sample with pre-program earnings data, RP participants appeared to earn noticeably less than their Wagner-Peyser counterparts—about $610 less in the quarter just before enrollment.

This gap in pre-program earnings suggests that our matching design does not control for factors that might distinguish RP and Wagner-Peyser participants pre-enrollment. For example, this pattern might indicate that the comparison group members had greater employability, on average, than the RP program group members, due to factors like ability, work experience, or motivation. That RP participants appear to have lower earnings pre-program (again, within the small subsample for whom we can measure pre-program earnings) could also indicate that some were not in the labor force, potentially due to incarceration.

Exhibit IV.4. Average earnings by quarter since enrollment, for the subset of sample members with pre-program earnings



Source: NDNH data and state administrative court records matched to WIPS data.

Notes: Sample sizes vary by quarter due to earnings data availability and range from 103 RP participants in quarter -2 to 1,141 RP participants in quarter 10.

RP = Reentry Project; NDNH = National Directory of New Hires.

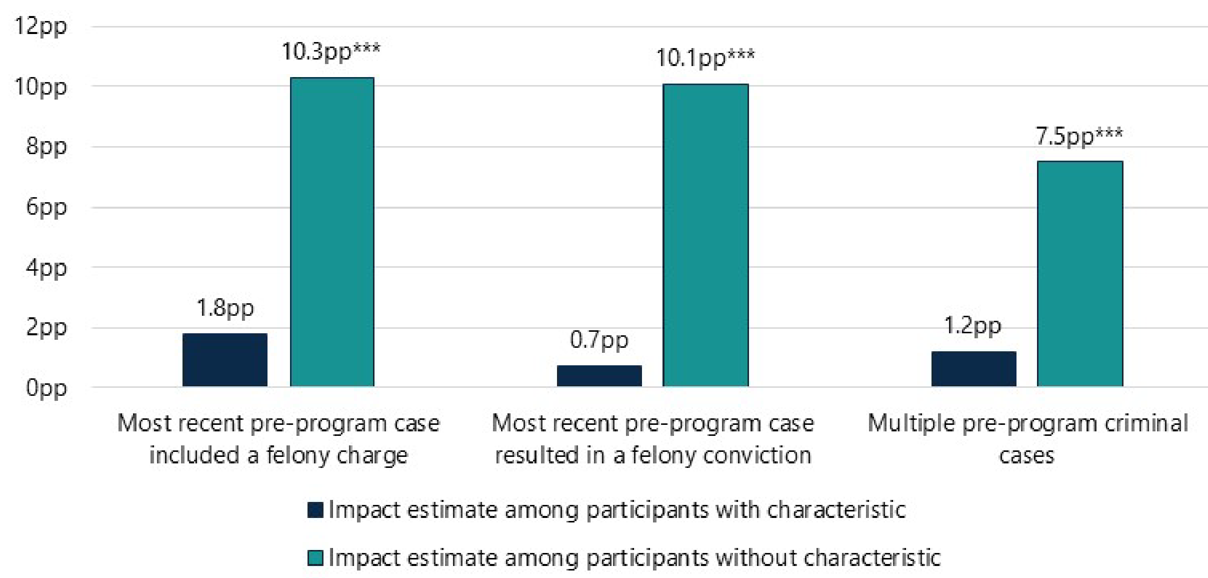
Potential bias from limited sentencing, incarceration data. As mentioned, by program design, most adult RP participants should have been released from incarceration before program entry or been under supervised probation shortly before enrollment; the majority of young adult participants should have had contact with the justice system, which could include juvenile offenses. We noted in Chapter 3 that we had limited sentencing and local jail data to accurately observe these statuses. In fact, we only observed pre-program incarceration records, derived from state prison data, for about half of adult RP participants in our analytic sample (all of whom we observed having faced criminal charges before enrollment). We likewise had limited information about probation status, nor any data on juvenile records, which are sealed by default. Put differently, although we know that many RP program participants must have experienced relatively serious contacts with the justice system before enrollment, in practice we cannot observe what about their prior justice involvement made them eligible for RP.

At the same time, we expect Wagner-Peyser participants to have had less serious prior justice contacts on average relative to their RP peers, because this program serves a wide range of participants and does not explicitly provide reentry services. However, even conditional on offense type and severity, which we can observe, we cannot distinguish those Wagner-Peyser participants with relatively minor sentences (for example, a conviction for a misdemeanor offense that led to a fine) from those with more serious sentences (for example, a conviction for a misdemeanor offense that led to jail time). Ideally, our matching design would focus on the latter Wagner-Peyser participants as candidate matches, because their sentences more closely align with what we expect RP participants to have experienced; in practice, however, we include both groups because we cannot tell them apart.

Consequently, our matching approach may have led us to compare RP and Wagner-Peyser participants with observably similar, but in reality, very distinct, prior justice involvement. Pre-enrollment, the average RP participant might have been more likely than matched comparison group members to be incarcerated—out of the labor force entirely and unable earn wage income—or under supervised probation, which might have limited their job prospects even as they remained in the labor force. Post-enrollment, RP participants may have been more likely to re-offend and less likely to secure gainful employment than the average comparison group member, given the relative seriousness of their prior justice involvement. In other words, lacking detailed data on pre-program criminal sentencing (and labor market outcomes), we may have arrived at a matched sample in which RP participants had systematically lower earnings potential pre-enrollment, and a higher risk of recidivating post-enrollment, than matched Wagner-Peyser participants.

To better understand how these data limitations might affect our results, we focus on characteristics of individuals' recent criminal cases that are likely correlated with more serious sentences—in particular, whether a person faced a felony charge, had a felony conviction, or faced multiple criminal court cases pre-program. We conducted subgroup analyses focused on individuals in the program and comparison groups who share these characteristics. Intuitively, if our estimates were biased due to missing data on prior justice involvement, we would have expected to find smaller program impacts when we restricted our sample to RP and Wagner-Peyser participants with more serious criminal histories, since these groups may have had more similar unobserved determinants of future recidivism.

Exhibit IV.5. Impact of Reentry Project on recidivism by severity of prior justice involvement



Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: The sample includes all RP participants, both adults and young adults.

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; WIPS = Workforce Integrate Performance System.

Indeed, when we narrowed our attention to subgroups with more serious justice involvement—who may have been more likely to experience similarly restrictive sentences—we observed much smaller and insignificant program impacts. As Exhibit IV.5 shows, we find that differences in post-program recidivism appear concentrated among participants with relatively minor past criminal justice involvement. That is, when we focused on RP and Wagner-Peyser participants with the greatest likelihood of pre-program incarceration or supervised probation, and the highest risk of post-program recidivism, we found relatively small—albeit, non-zero—differences between our program and comparison groups. Exhibit IV.6 indicates that this pattern bears out in our labor market outcomes as well.

Exhibit IV.6. Impact of Reentry Project (RP) on confirmatory outcomes by severity of prior justice involvement

Characteristic	RP sample size	Recidivism		Employment		Earnings	
		Comparison mean	Impact estimate	Comparison mean	Impact estimate	Comparison mean	Impact estimate
Most recent pre-program criminal case included:							
A felony charge	732	22%	1.8pp (1.9pp)	42%	-2.3pp (2.1pp)	$2,832	-$445** ($198)
No felony charges	466	19%	10.3pp*** (2.4pp)	49%	-6.9pp*** (2.7pp)	$3,103	-$1,083*** ($204)
Most recent pre-program criminal case resulted in:							
A felony conviction	653	22%	0.7pp (2.0pp)	41%	-1.4pp (2.3pp)	$2,813	-$503** ($211)
No felony convictions	545	19%	10.1pp*** (2.2pp)	49%	-7.2pp*** (2.4pp)	$3,079	-$915*** ($196)
Number of pre-program criminal cases:							
Multiple cases	449	30%	1.2pp (2.7pp)	37%	-1.9pp (2.7pp)	$2,063	-$161 ($205)
Exactly one case	749	15%	7.5pp*** (1.8pp)	50%	-5.4pp** (2.1pp)	$3,516	-$1,021*** ($193)

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates. The sample includes both adult and young adult RP participants who successfully matched to at least one Wagner-Peyser participant (1,198 of the 2,092 RP participants in our analytic sample).

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; WIPS = Workforce Integrated Performance System.

These findings support the conclusion that, when we narrowed our attention to these subgroups with more serious justice involvement, we may have improved our match quality, reducing the baseline differences in earnings prospects and the likelihood of recidivism between our program and comparison groups. Individuals with less serious and likely more heterogeneous prior justice involvement appear to have driven our topline results. This suggests that unobserved data on criminal justice backgrounds might have biased our main results, leading us to overstate adverse impacts on recidivism, employment, and earnings.

D. Subgroup analyses

To reflect on how our estimated impacts vary across RP participants who come from different backgrounds and who had different program experiences, we conducted a series of subgroup analyses. This exercise focused on understanding which policy-relevant subgroups exhibited particularly large differences in labor market and criminal justice outcomes between RP and Wagner-Peyser participants. Our subgroup findings appear in Exhibit IV.7. For succinctness, Exhibit IV.7 only presents estimates for our sample of all RP participants, both adult and young adult.

Race and ethnicity. We first broke down our findings by participant race and ethnicity. These subgroups speak to well-known racial and ethnic disparities in employment- and justice-related outcomes. We focused on non-Hispanic White individuals (about 19 percent of the RP sample), non-Hispanic Black individuals (58 percent of the RP sample), and Hispanic participants (16 percent of the RP sample). For completeness, we also include the handful of participants who are another race or ethnicity (7 percent of the RP sample).

We found that Black RP and Wagner-Peyser participants showed relatively large differences in recidivism and employment, compared to White and Hispanic participants. Black program group members were 6.7 percentage points more likely to recidivate, and 5.9 percentage points less likely to be employed in the 9th and 10th quarters after enrollment, relative to Black comparison group members. By contrast, Hispanic RP participants were no more or less likely to be convicted of an offense over the 10 quarters after enrollment than similar Wagner-Peyser participants. However, Hispanic RP participants did earn much less ($930) on average in the 9th and 10th quarters after enrollment. Finally, among White participants, we found no statistically significant impacts on any outcomes, although point estimates suggest higher recidivism rates and slightly lower earnings among the participants in this group. Overall, these mixed results point to more adverse impacts for Black RP participants, but generally less pronounced impacts for White and Hispanic participants.

Gender. We estimated differences in outcomes for men and women who participated in RP and Wagner-Peyser programs, with the goal of capturing the potential influence of gender disparities in earnings and employment. Point estimates indicated that female RP participants—who comprised only about 17 percent of RP participants—had much higher recidivism rates (16.1 percentage points) than female Wagner-Peyser participants over the 10 quarters after enrollment. This difference far exceeded the estimated difference in recidivism rates among male participants (2.8 percentage points). On the other hand, male RP participants were 5.4 percentage points *less* likely to be employed in the 9th and 10th quarters than comparison group members, while female RP participants were not statistically more or less likely to be employed than their counterparts in Wagner-Peyser. Both male and female RP participants had lower earnings than comparable Wagner-Peyser participants, although the impact estimate among females was not statistically significant. Taken together, these comparisons indicate that our observed negative differences in employment rates and earnings across our program and comparison groups were driven by male participants, while our estimated impacts on recidivism were driven by female participants.

Program grantee type. The companion implementation study of RP (Geckeler et al. 2023) observed variation in the consistency and intensity of reentry services programs across grantees. To examine how these program characteristics may have shaped participant outcomes, we broke down the RP sample to compare the experience of participants who enrolled in programs run by intermediary grantees against those for participants who enrolled in programs run by CBOs. As described in Chapter 2, intermediary grantees tended to implement a more standardized service delivery model, while CBOs developed more localized case management models based on the local context. About 44 percent of RP participants in our analytic sample enrolled in a program run by a CBO; the remaining 56 percent enrolled in a program run by an intermediary grantee.[18]

Our comparison of estimates by grantee type yielded mixed results. We found that participants from intermediary grantees were 8.8 percentage points more likely to recidivate than comparable Wagner-Peyser participants; this gap is larger than the 0.2 percentage-point difference in recidivism rates between RP participants whose programs were run by CBOs and comparison group members. On the other hand, participants from intermediary grantees stacked up more favorably against comparable Wagner-Peyser group members along labor market outcomes. For example, whereas participants from CBOs earned $947 less in the 9th and 10th quarters after enrollment, relative to the comparison group, participants from intermediary grantees earned $578 less than comparable Wagner-Peyser participants. These findings suggest that program grantee type might have affected participant outcomes, but not consistently for better or worse.

Type of services received. To further comment on how RP participants' experiences in the program shaped their outcomes, we evaluated our impact estimates specifically for those individuals who received different types of services. One might expect that participants who received more intensive services might have experienced more favorable post-program outcomes. Specifically, we assessed impacts among RP participants who received basic services (also referred to as light-touch), individualized career services only, or those who received training and educational services. We did not find evidence that people who received more intensive services had more favorable estimated impacts, even though a substantial majority (899 of 1,198 RP participants in our sample) received training or education services. Estimates for each of the three service tiers point to qualitatively similar gaps between RP and matched Wagner-Peyser participants, with potentially larger gaps in earnings for those receiving only basic services (who earned $1,143 less on average in the 9th and 10th quarters post-enrollment, against $465 and $683 lower earnings among those who received individualized and education and training services, respectively).

Enrollment during the COVID-19 pandemic. Finally, we parsed our sample by whether or not a person enrolled in their program (either RP or Wagner-Peyser) before or during the COVID-19 pandemic. This analysis speaks to how the disruption to the justice system and the labor market caused by the pandemic might have influenced participant outcomes. Because most RP participants in the impact study (907 out of 1,198) enrolled before the onset of the COVID-19 pandemic, we have limited statistical power to detect differences among the 291 program participants who enrolled during the pandemic. As such, although we found more statistically significant differences in outcomes among those who enrolled before the pandemic, the actual point estimates for these groups were very similar.

[18] In our pre-analysis plan, we indicated that we would examine impacts by grantee. However, our final sample size proved too small to conduct effective and well-powered grantee-specific analyses.

Given that the 10 quarters following enrollment for almost all individuals in our sample overlapped with the pandemic, the similarity of these estimates may not be surprising.

Exhibit IV.7. Impact of Reentry Project (RP) by subgroup

Characteristic	RP sample size	Recidivism: Comparison mean	Recidivism: Impact estimate	Employment: Comparison mean	Employment: Impact estimate	Earnings: Comparison mean	Earnings: Impact estimate
Race/ethnicity							
White	228	24%	4.1pp (3.4pp)	45%	-1.0pp (3.6pp)	$3,028	-$388 ($316)
Black	699	19%	6.7pp*** (1.9pp)	46%	-5.9pp*** (2.2pp)	$2,832	-$792*** ($177)
Hispanic	186	26%	0.1pp (4.2pp)	42%	-4.5pp (4.3pp)	$3,288	-$930** ($430)
Missing/other	85	16%	4.7pp (5.0pp)	45%	3.1pp (5.9pp)	$2,756	-$200 ($530)
Gender							
Female	207	10%	16.1pp*** (3.2pp)	49%	2.0pp (4.0pp)	$2,839	-$326 ($320)
Male	991	23%	2.8pp (1.7pp)	44%	-5.4pp*** (1.8pp)	$2,958	-$769*** ($159)
RP program sponsor							
CBO	523	21%	0.2pp (2.0pp)	45%	-6.5pp** (2.4pp)	$2,937	-$947*** ($190)
Intermediary grantee	675	21%	8.8pp*** (2.0pp)	45%	-3.6pp (2.2pp)	$2,937	-$578** ($191)
RP service tier							
Basic services	175	21%	5.0pp (3.5pp)	45%	-8.5pp** (3.6pp)	$2,937	-$1,143*** ($300)
Individualized services	124	21%	4.7pp (4.1pp)	45%	-1.9pp (4.2pp)	$2,937	-$465 ($338)
Training or education	899	21%	4.7pp*** (1.6pp)	45%	-3.8pp** (1.8pp)	$2,937	-$683*** ($154)
COVID-19 pandemic							
Enrolled pre-pandemic	907	20%	5.7pp*** (1.7pp)	44%	-3.1pp (1.9pp)	$2,854	-$629*** ($162)
Enrolled during pandemic	291	22%	2.7pp (3.2pp)	47%	-5.5pp (3.4pp)	$3,197	-$785** ($304)

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates. For subgroups based on RP program characteristics, there is a single sample mean for each outcome because each subgroup estimate includes the whole comparison group.

** p-value < 0.05

*** p-value < 0.01

CBO = community-based organizations; pp = percentage points; WIPS = Workforce Integrated Performance System; NDNH = National Directory of New Hires.

E. Sensitivity analyses

We conducted a variety of additional analyses to gauge the sensitivity of our reported findings to different analytic and modeling choices that we made. These checks evaluated alternative approaches to estimating our propensity scores and tolerances for matches based on those propensity scores, different regression estimators for evaluating impacts, and the consequences of restricting our comparison group to the subgroup of Wagner-Peyser participants who received the least-intensive services. In the Technical Appendix, we discuss these tests and their results in more detail. Qualitatively, though, these sensitivity analyses all revealed similar findings compared to our preferred approach.

V. Summary and Conclusions

The DOL's RP grant program operated during four grant cycles starting in 2016. Grants were intended to help organizations in high-crime communities implement comprehensive reentry programs designed to support justice system-involved young adults (ages 18 to 24) and older adults (ages 25 and up) successfully re-engage in their communities and avoid recidivism. Built on decades of work, DOL designed RP grants for both intermediary organizations that served large numbers of participants across multiple subgrantees and states, and smaller CBOs that served a smaller number of participants in a single location. Intermediary organizations received grants ranging from roughly $4 million to $4.5 million and CBOs received grants ranging from roughly half a million to $1.5 million. Although individual grantee organizations had substantial flexibility in designing their programs, DOL encouraged programs to employ evidence-informed or promising practices, especially around case management services (for example, youth positive development, motivational interviewing, cognitive behavioral therapy, trauma-informed care, etc.) and to build employment-focused services around major industries open to hiring justice system-involved individuals, including construction, culinary and hospitality programs, manufacturing, warehousing, and transportation. Grants were 36 to 39 months long and included a three-month planning period, a 24-month operational period, and a nine- to 12-month follow-up period.

A. Implementation evaluation summary

Mathematica and SPR conducted an evaluation of the RP grants, which included an implementation and impact study, both of which examined grantees from the 2018 and 2019 grant cycles. The implementation study included all 16 intermediary grantees and 68 CBO grantees in the 2018 and 2019 grant cycles and was based on the analysis of grantee survey data, virtual site visits to 27 sites (locations that could include CBO grantee locations or subgrantee locations of intermediary grantees), WIPS data, and document reviews. The implementation study findings, included in a prior report (Geckeler et al. 2023), highlight how RP grant programs reached large numbers of participants and provided them with wide-ranging and comprehensive employment services, even as grantees encountered challenges due to both historical events and the barriers to employment that justice-involved populations often face. More specifically:

- The 2018 and 2019 RP grantees enrolled 17,361 participants (9,098 adult participants and 8,263 young adult participants), who were predominantly male (88 percent), and identified as Black, Hispanic, Asian, or multi-racial (83.2 percent), with a little more than half (56 percent) having completed high school or having a high school equivalency degree.
- RP grantees struggled somewhat with enrollment, falling short of their enrollment goals, with 2019 grantees struggling more than 2018 grantees, and grantees serving young adults struggling more than grantees serving older adults. The COVID-19 pandemic and its subsequent repercussions on the economy explains some of the challenges, including impeding grantees' abilities to conduct outreach and to deliver services as well as the demands these changes to the economy placed on participants around employment, often raising the incomes earned by unskilled workers within communities and limiting the (short-term) need for training. Grantees also noted other challenges to implementation spurred by these historical circumstances, including those having do to with developing partnerships, retaining staff, and impediments to efforts to build their reputation and presence within communities.

- Participant retention was also challenging for grantees. Sites that the study team visited regularly cited being unable to meet certain basic needs of program participants as a retention challenge for them. Many of these sites discussed the challenges around keeping participants motivated and engaged, especially young adult participants. In response, staff noted how helping participants shift their mindset towards more positive ways of thinking and embracing the value of training and strategic thinking about long-term employment, and ensuring they were enrolled in multiple services was most useful in keeping them in the program, even though they were sometimes unable to compete with other employment opportunities.
- While moderate levels of participants received education and training services, very few received work-based learning such as apprenticeships and on-the-job training. Based on WIPS data, about 72 percent of RP participants from 2018 and 2019 received education or training services and about 43 percent received occupational skills training. At the same time, despite many programs offering work-based learning experiences, only 1.3 percent received registered apprenticeships and 2.3 percent received on-the-job training. The fact that relatively few RP participants received these work-based learning opportunities speaks to the difficulty of implementing these programs, which require a substantial grantee investment in developing employer relationships. While nearly all grantees offered a wide array of case management and employment readiness and placement services, the study team did not have the data needed to measure receipt of these services.

It is important to recap these implementation findings in light of the current impact study, which is the focus of this report.

B. Impact evaluation summary

The impact study compared participants from a subset of RP grantees to participants in the Wagner-Peyser program. The implementation study found that although grantees faced certain challenges, they were able to establish programs, enroll large numbers of participants, and deliver education and occupational skills training services to many participants. Although the Wagner-Peyser program provides employment-focused services to a large number of individuals, in comparison to RP, its services more narrowly focus on job search and placement services, access to information about jobs, support for individuals on unemployment, and developing employer partners to support hiring opportunities. Wagner-Peyser program services do not include intensive case management centered around addressing basic needs, staff who are focused on aiding individuals confronting barriers to employment regularly faced by justice system involved individuals, or any provision of education and training services. However, because Wagner-Peyser is one of the most readily available employment services to the public at large, it provides a good sense of the kind of support RP participants might have received had they not taken part in RP grant-funded programs.

The quasi-experimental impact study discussed in this report used a matched comparison design in which the study team compared the employment and criminal justice system outcomes of 1,198 RP grant program participants (the "program group") to the same outcomes for 16,032 Wagner-Peyser program participants (the "comparison group"). Program and comparison group members (the "study sample") included individuals served by either the RP grant or Wagner-Peyser programs between 2018 and 2021, and who received their respective program services in one of six states (Alabama, Florida, New Jersey, New York, Oregon, and Pennsylvania) where the four largest RP intermediary grantees operated and where state agencies were willing and able to provide sufficient administrative data. Study sample members were also selected based on whether they were observed in state criminal justice data with a criminal charge before enrollment in their respective programs and where individuals in both groups resembled one another in terms of their demographic characteristics and criminal justice histories.

After comparing the criminal justice system and employment outcomes of program group members to those of comparison group members, we found that RP participants had worse post-program criminal justice and labor market outcomes relative to the matched Wagner-Peyser comparison group. In particular:

- During the 10 quarters after enrollment, program group members were about 5 percentage points more likely to have a new criminal conviction than comparison group members.
- In the 9th and 10th quarters after exit, program group members were about 4 percentage points less likely to be employed than comparison group members.
- In the 9th and 10th quarters after enrollment, program group members earned $693 less than comparison group members, who earned an average of $2,937 per quarter.

C. Discussion

The impact study shows that program group members did not perform better than, and in some cases performed worse than, comparison group members in terms of new convictions, employment rates, and earnings. These findings support many potential interpretations, but we focus on two overarching narratives: (1) RP might truly lead to worse outcomes for participants, relative to Wagner-Peyser, and (2) our estimation approach might fail to address underlying differences between program and comparison group members that generate worse outcomes for program group members. We explore each of these interpretations in turn.

The first possible interpretation of the findings in this report is that the RP grant programs studied truly had a negative impact or no impact on participants' rate of new convictions, employment, and earnings, relative to Wagner-Peyser programs. If so, these results are disappointing given the investments and accomplishments grantees made in implementing their RP grant programs and the information they shared around the achievements of some participants, as illustrated in this report and the implementation study report. However, there is precedent for these results. Past research has found mixed evidence of reentry programs' effectiveness, due to variation in program models and implementation quality, and several existing studies of reentry programs similar to RP report unfavorable impacts on recidivism and employment.

Consistent with this interpretation is that many RP grantees experienced challenges in implementing their programs, primarily around enrollment and retention as well as around take-up of more intensive services. In addition, some of these challenges may have been exacerbated by the COVID-19 pandemic. Although RP programs do appear to have provided more intensive services than the typical Wagner-Peyser program, the mix of services received may not have been as intensive as those anticipated in the RP program logic model. Because RP participants in the impact analysis are not necessarily representative of all RP participants, this may also be the case specifically for program group members only.

A second possible interpretation of the findings in this report is that there are meaningful, unobserved, pre-program differences between the program and comparison groups which led to these results. Due to data limitations, our matching approach may have led us to compare RP and Wagner-Peyser participants with observably similar, but, in reality, distinct, prior justice involvement and employment backgrounds. Our impact study had to overcome two critical sources of missing data: (1) a lack of granular sentencing, incarceration, and probation records; and (2) a lack of pre-program employment and earnings information. Absent such data, there may be fundamental differences between the RP and Wagner-Peyser participants in this study even after creating a matched comparison group. In particular, because of RP program eligibility requirements, most adult RP participants likely had recently received serious criminal sentences involving incarceration or supervised probation; Wagner-Peyser has no such condition. We were able to restrict our pool of comparison group members to Wagner-Peyser participants with recent criminal charges with similar case characteristics (for example, whether the case included a felony offense or whether the person had a history of other criminal cases). However, we did not have the incarceration or probation data needed to limit the pool to individuals who would have met the RP eligibility criteria related to incarceration or supervised probation.

This may have resulted in a matched comparison group with very different background characteristics than the RP participants in our analytic sample. An analysis of variation in our impact estimates by severity of prior justice involvement supports this hypothesis. Although we cannot reliably observe incarceration or probation details, we can observe characteristics of individuals' recent criminal cases that are likely correlated with more serious sentences. Indeed, RP participants who had more serious prior justice involvement showed smaller differences in outcomes compared to similar matched Wagner-Peyser participants. That is, when we focused on RP and Wagner-Peyser participants with the greatest likelihood of pre-program incarceration or supervised probation, and the highest risk of post-program recidivism, we found relatively small and generally not statistically significant—albeit, non-zero—differences in outcomes. This pattern could suggest that missing data might have led us towards biased estimates.

In addition, the employment- and wage-related data used in this analysis were only available for a two-year period of time. Because of this limited observation period, we could not observe pre-program earnings or employment for the vast majority of the analysis sample. For the small set of individuals in which we could observe pre-program earnings, the results suggest that RP participants had lower pre-program earnings than the matched comparison group. This gap in pre-program earnings suggests that our matching design does not control for factors that might distinguish RP and Wagner-Peyser participants pre-enrollment. For example, this pattern might indicate that the comparison group members had greater employability, on average, than the RP program group members, due to factors like ability, work experience, or motivation. It could also indicate that a larger share of RP participants was not in the labor force, potentially due to incarceration, than the matched Wagner-Peyser sample.

An additional important caveat for all impact study findings is that they may not be generalizable. The impact study findings are specific to the participants in the six states included in the study, which were not randomly selected from the full set of states or grantees. They were selected based primarily on size and states in which there were agencies willing and able to participate in a study and were not necessarily representative of all RP grantees in terms of other factors like geography or program type. In addition, within these states and grantees, the RP participants included in the impact analysis were those with suitable pre-program and outcome data for whom we could find a matched comparison group member with similar observed characteristics. Other reentry programs with similar programming, or other RP grantees, may have different impacts on participant outcomes, and future policy should bear this consideration in mind when reflecting on the lessons learned from RP.

The results from this impact study provide several implications for policy and future research. Existing evidence on employment-focused reentry programs has reinforced the importance of intensive employment and training services paired with wraparound supports for improving labor market and recidivism outcomes for populations with prior justice involvement (Lacoe and Betesh 2019; Wiegand and Sussell 2016). These program components are central to the program logic model for the RP grants, but based on the implementation study, participants likely did not receive these services exactly as intended. These findings suggest that there might be returns to using programming and policy to address program implementation challenges experienced by grantees, helping to ensure they are able to offer the intended program model in its entirety, and that participants receive the type and dosage of services as intended. To evaluate the impact of the intended program components, future research should consider assessing implementation fidelity, prior to beginning an impact study, and using comprehensive measures of service receipt to test that implementation of the program model is occurring as intended before evaluating it. A mixed methods approach along these lines, combining insights from a qualitative implementation analysis with an impact evaluation, could also help shed light on the mechanisms that explain why a program did or did not improve participant outcomes.

The potential sources of bias in our impact estimates are unique to a matched comparison design but could be mitigated with additional data. In particular, more complete sentencing records that captured stays in local jails and probation outcomes, as well as more comprehensive pre-program and employment records, could have improved match quality in our study. These data would be essential for any future evaluation of a reentry program that leverages a matching design. Nevertheless, these data can be challenging and expensive to obtain due to being housed locally, at the city or county level of government, and not being accessible at the state or federal levels. In terms of employment data, state wage records or other data with longer histories than the NDNH data used in this study would need to be readily available. Outreach to multiple state agencies is burdensome but possible, but frequently state agencies are particularly restrictive and protective around the release of wage records that would be needed for this analysis. Data collection of wage data could be accomplished through completion of a participant-level survey, but these also involve substantial levels of effort and costs.

Alternatively, an RCT design, would obviate the need for additional pre-program administrative data, because treatment and control groups would both be formed from individuals eligible for the program and be identical on both observable and unobservable characteristics. However, RCTs are infeasible in many contexts. We explored the possibility of conducting an RCT for the RP grant evaluation but ultimately concluded that an experimental study design was not feasible because grantees would not attract sufficient applicants to allow us to randomize some to the control condition. i. Complementary approaches, such as piloting enhanced services, may allow for a randomized design in settings without oversubscription. Given the challenges in implementing an RCT design, running an RCT successfully may also require compensation for grantees to participate in such research and both clear requirements, and support in developing program plans to include RCT planning from the outset, for them to do so. Fortunately, DOL is continuing to support programs that have the potential to provide more robust services and evaluations that employ these types of designs and data collection efforts as part of its current research into the Pathway Home grant program and PROWD.

References

Abraham, Katharine G., John C. Haltiwanger, Kristin Sandusky, and James R. Spletzer. “Measuring the Gig Economy: Current Knowledge and Open Issues.” Working Paper no. w24950. Cambridge, MA: National Bureau of Economic Research, 2018.

Austin, Peter C. “An introduction to propensity score methods for reducing the effects of confounding in observational studies.” *Multivariate Behavioral Research*, 46(3), 2011, 399-424.

Barden, Bret, Randall Juras, CindyRedcross, Mary Farrell, and Dan Bloom. “New Perspectives on Creating Jobs: Final Impacts of the Next Generation of Subsidized Employment Programs.” Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018. https://clear.dol.gov/Study/New-perspectives-creating-jobs-Final-impacts-next-generation-subsidized-employment-programs.

Bellotti, Jeanne, Samina Sattar, Alix Gould-Werth, Jillian Berk, Ivette Gutierrez, Jill Stein, Hannah Betesh, et al. “Bringing the American Job Center into Jails: Implementation of the Linking to Employment Activities Pre-Release Grants.” Mathematica Policy Research, Social Policy Research Associates, 2018. https://www.dol.gov/agencies/eta/research/publications/implementation-linking-employment-activities-pre-release-leap.

Blakemore, Arthur E., Paul L. Burgess, Stuart A. Low, and Robert S. St. Louis. “Employer Tax Evasion in the Unemployment Insurance Program.” *Journal of Labor Economics*, vol. 14, no. 2, 1996, pp. 210–230.

Bronson, Jennifer, and Marcus Berzofsky. “Indicators of Mental Health Problems Reported by Prisoners and Jail Inmates, 2011-2012.” NCJ 250612. Bureau of Justice Statistics, U.S. Department of Justice, June 2017.

Carson, Ann. “Jail Inmates in 2020.” Bureau of Justice Statistics, U.S. Department of Justice, 2020. https://bjs.ojp.gov/content/pub/pdf/p20st.pdf.

Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. “BART: Bayesian Additive Regression Trees.” *Annals of Applied Statistics*, vol. 4, no. 1, 2010, pp. 266–298.

Cook, Philip J., Songman Kang, Anthony A. Braga, Jens Ludwig, and Mallory E. O’Brien. “An Experimental Evaluation of a Comprehensive Employment-Oriented Prisoner Re-Entry Program.” *Journal of Quantitative Criminology*, vol. 31, 2015, pp. 355–382.

Cortina, Hannah, Amy Maniola Allen, Christine Lindquist, Jillian Stein, and Jeanne Bellotti. “Charting the Path to Employment After Incarceration: A Research Synthesis of Employment-Focused Reentry Programs.” Mathematica, Research Triangle Institute. Report submitted to the U.S. Department of Labor. Forthcoming.

Council of State Governments Justice Center (CSGJC). “National Inventory of Collateral Consequences of Convictions.” Council of State Governments Justice Center, 2020. https://csgjusticecenter.org/publications/the-national-inventory-of-collateral-consequences-of-conviction/.

Czajka, John L., Ankita Patnaik, and Marian Negoita. “Data on Earnings: A Review of Resources for Research.” Mathematica, 2018. https://www.dol.gov/sites/dolgov/files/OASP/legacy/files/Data-on-Earnings-Report.pdf.

D’Amico, R., and H. Kim. “An Evaluation of Seven Second Chance Act Adult Demonstration Programs: Impact Findings at 30 Months.” Social Policy Research Associates, 2018. https://www.ojp.gov/pdffiles1/nij/grants/251702.pdf.

Denney, Andrew S., Richard Tewksbury, and Richard S. Jones. “Beyond Basic Needs: Social Support and Structure for Successful Offender Reentry.” *Journal of Quantitative Criminal Justice & Criminology*, 2014.

Doleac, Jennifer L. “Wrap-Around Services Don’t Improve Prisoner Reentry Outcomes.” *Journal of Policy Analysis and Management*, vol. 38, no. 2, 2019, pp. 508–514.

Employment and Training Administration. “Wagner-Peyser Act of 1933, as amended.” https://www.dol.gov/agencies/eta/american-job-centers/wagner-peyser. Accessed on August 7, 2024.

English, B., and P. Holcomb. “New Requirements for American Job Center Systems Regarding One-Stop Operators, Partnership Agreements, and Certification.” Mathematica, Social Policy Research Associates, 2020. https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/ETA_WIOAStudy_AJCsystems.pdf.

Finn, P. "Texas' Project RIO (Re-Integration of Offenders)." National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, 1998. https://www.ojp.gov/pdffiles/168637.pdf.

Fortson, Kenneth, Dana Rotz, Paul Burkander, Annalisa Mastri, Peter Schochet, Linda Rosenberg, Sheena McConnell, et al. "Providing Public Workforce Services to Job Seekers: 30-Month Impact Findings on the WIA Adult and Dislocated Worker Programs." Washington, DC: Mathematica Policy Research, January 2017. https://www.dol.gov/sites/dolgov/files/OASP/legacy/files/WIA-30mo-main-rpt.pdf.

Funk, Michele Jonsson, Daniel Westreich, Chris Wiesen, Til Stürmer, M. Alan Brookhart, and Marie Davidian. "Doubly Robust Estimation of Causal Effects." *American Journal of Epidemiology,* vol. 173, no. 7, 2011, pp. 761–767.

Geckeler, Christian, Leah Cadena-Igdalsky, Ivette Gutierrez, Madeleine Levin, Sergio Martinez, Jessica Muñoz, Anne Paprocki, et al. "Reentry Projects Grant Evaluation." Office of the Assistant Secretary for Policy, U.S. Department of Labor. 2023. https://www.dol.gov/agencies/oasp/evaluation/completedstudies/Reentry-Projects-Grant-Evaluation.

Holzer, Harry, Steven Raphael, and Michael Stoll. "The Effect of an Applicant's Criminal History on Employer Hiring Practices and Screening Decisions: Evidence from Los Angeles." Ann Arbor, MI: National Poverty Center, 2004.

Hotz, Joseph V., and John K. Scholz. "Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data." Institute for Research on Poverty, University of Wisconsin—Madison, 2001.

Houseman, Susan N. "Why Employers Use Flexible Staffing Arrangements: Evidence from an Establishment Survey." *Industrial and Labor Relations Review*, vol. 55, no. 1, October 2001, pp. 149–170.

Jacobs, Erin, and Brice Western. "Report on the Evaluation of the ComALERT Prisoner Reentry Program." 2007. https://scholar.harvard.edu/files/brucewestern/files/report_1009071.pdf.

Katz, Lawrence F., and Alan B. Krueger. "The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015." Working Paper no. w22667. Cambridge, MA: National Bureau of Economic Research, 2016.

Katz, Lawrence F., and Alan B. Krueger. "Understanding Trends in Alternative Work Arrangements in the United States." Working Paper no. w25425. Cambridge, MA: National Bureau of Economic Research, 2019.

Kornfeld, Robert, and Howard S. Bloom. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics*, vol. 17, no. 1, January 1999, pp. 168–197.

Lacoe, J., and H. Betesh. "Supporting Reentry Employment and Success: A Summary of the Evidence for Adults and Young Adults." Mathematica, 2019. https://www.dol.gov/agencies/eta/research/publications/supporting-reentry-employment-and-success-summary-evidence.

Looney, Adam, and Nicholas Turner. "Work and Opportunity Before and After Incarceration." Brookings Institution, 2018. https://pdfs.semanticscholar.org/399b/5d1747e721fdb63a5837296619528d361de6.pdf.

McCaffrey, Daniel F., Greg Ridgeway, and Andrew R. Morral. "Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies." *Psychological Methods*, vol. 9, no. 4, 2004, pp. 403–425.

Pager, Devah. "The Mark of a Criminal Record." *American Journal of Sociology,* vol. 108, March 2003, pp. 937–975.

Ramakers, Anke, Paul Nieuwbeerta, Johan Van Wilsem, and Anja Dirkzwager. "Not Just Any Job Will Do: A Study on Employment Characteristics and Recidivism Risks After Release." *International Journal of Offender Therapy and Comparative Criminology,* vol. 61, no. 16, 2017, pp. 1795–1818.

Raphael, Steven. "The New Scarlet Letter? Negotiating the U.S. Labor Market with a Criminal Record." W.E. Upjohn Institute for Employment Research, 2014.

Redcross, Cindy, Megan Millenky, Timothy Rudd, and Valerie Levshin. "More Than a Job: Final Results from the Evaluation of the Center for Employment Opportunities (CEO) Transitional Jobs Program." MDRC, 2012. https://clear.dol.gov/study/more-job-final-results-evaluation-center-employment-opportunities-ceo-transitional-job-program.

Shiferaw, Leah, and Dan Thal. "Digging Deeper into What Works: What Combinations of Services Work, and for Whom?" OPRE Report #2022-161. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022. https://www.acf.hhs.gov/opre/report/digging-deeper-what-works-what-services-improve-labor-market-outcomes-and-whom.

Spitzer, Ariella, Brittany English, Breyon Williams, Daniel Thal, Arielle Marks-Anglin, Chris Weiss, Jeanne Bellotti, Jillian Berk. "The Impact of Regional Sectoral Training Partnerships: Findings from America's Promise." Report submitted to the U.S. Department of Labor. Washington, DC. Mathematica, 2023. Available at https://www.dol.gov/agencies/eta/research/publications/impact-regional-sectoral-training-partnerships-findings-americas#:~:text=Employment%20and%20earnings%20trajectories%20after%20enrollment%20differed%20by,after%20enrollment%2C%20both%20of%20which%20were%20statistically%20significant.%E2%80%A2.

Stuart, Elizabeth A., Brian K. Lee, and Finbarr P. Leacy. "Prognostic Score–Based Balance Measures can be a Useful Diagnostic for Propensity Score Methods in Comparative Effectiveness Research." *Journal of Clinical Epidemiology*, vol. 66, no. 8, 2013, pp. S84–S90.

The United States Commission on Civil Rights. "Collateral Consequences: The Crossroads of Punishment, Redemption, and the Effects on Communities." June 2019. https://www.usccr.gov/files/pubs/2019/06-13-Collateral-Consequences.pdf.

U.S. Department of Justice. "Recidivism of Prisoners Released in 24 States in 2008: A 10-Year Follow-Up Period (2008–2018)." 2021. https://bjs.ojp.gov/library/publications/recidivism-prisoners-released-24-states-2008-10-year-follow-period-2008-2018.

U.S. Department of Labor. "Employment and Training Administration Grants Awarded." n.d. https://www.dol.gov/agencies/eta/grants/awards. Accessed September 9, 2022.

U.S. Department of Labor. "Partners for Reentry Opportunities in Workforce Development Fact Sheet." 2022. https://www.dol.gov/sites/dolgov/files/ETA/youth/pdfs/2022-10_PROWD%20Grants%20Fact%20Sheet.pdf.

U.S. Department of Labor. "DOL Evaluation Design Pre-Specification Plan: Reentry Programs." Issue Brief. 2023.

U.S. Department of Labor, Employment and Training Administration. "Notice of Availability of Funds and Funding Opportunity Announcement For: Reentry Projects (RP)." 2017. https://www.dol.gov/sites/dolgov/files/ETA/grants/pdfs/FOA-ETA-17-02.pdf.

U.S. Department of Labor, Employment and Training Administration. "Notice of Availability of Funds and Funding Opportunity Announcement For: Reentry Projects (RP)." 2018. https://www.dol.gov/sites/dolgov/files/ETA/grants/pdfs/FOA-ETA-18-02.pdf.

U.S. Department of Labor, Employment and Training Administration. "Notice of Availability of Funds and Funding Opportunity Announcement For: Reentry Projects (RP-3)." 2019. https://www.dol.gov/sites/dolgov/files/ETA/grants/pdfs/FOA-ETA-19-01.pdf.

Valentine, Erin Jacobs, and Cindy Redcross. "Transitional Jobs After Release from Prison: Effects on Employment and Recidivism." *IZA Journal of Labor Policy*, vol. 4, no. 1, 2015, pp. 16.

What Works Clearinghouse. What Works Clearinghouse Procedures and Standards Handbook, version 5.0. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). 2022. https://ies.ed.gov/ncee/wwc/Handbooks.

Wiegand, Andrew, and Jesse Sussell. "Evaluation of the Re-Integration of Ex-Offenders (RExO) Program: Final Impact Report." Social Policy Research Associates, 2016. https://clear.dol.gov/Study/Evaluation-Re-Integration-Ex-Offenders-RExO-program-Final-impact-reports-Wiegand-Sussell-2016.

Zeng, Zhen, and Todd Minton. "Jail Inmates in 2019." Bureau of Justice Statistics, 2021.

Zhang, Zhongheng, Hwa Jung Kim, Guillaume Lonjon, and Yibing Zhu. "Balance Diagnostics After Propensity Score Matching." *Annals of Translational Medicine*, vol. 7, no. 1, 2019.

Technical Appendix

This appendix presents supplementary details of our technical approach for the impact evaluation of the Reentry Project (RP) grants. We used a matched comparison design that compared recidivism, earnings, and employment outcomes between RP participants and a matched comparison group of individuals enrolled in Wagner-Peyser services with similar observable characteristics. Section A provides a detailed description of the data sources used for the impact study, as well as our process for collecting these data and linking individuals across data sources. Section B describes the methods used to construct a matched comparison sample of Wagner-Peyser participants, and Section C presents the methods used for impact estimation. Section D provides supplemental tables to accompany the final report.

A. Data sources and linkages

1. Workforce Integrated Performance System

The Workforce Integrated Performance System (WIPS) is a national database that contains data on participants in workforce programs funded by the U.S. Department of Labor (DOL), including Wagner-Peyser employment services and the RP grants. The WIPS contains data on individual-level demographic characteristics, including age, gender, race, ethnicity, disability status, education, employment status at program enrollment, and English learner status. The WIPS also includes data on employment and training services received through RP and Wagner-Peyser programs.

Both grantees (for RP participants) and state workforce agencies (for Wagner-Peyser program participants) collect these data consistently and submit them quarterly to the WIPS.[19] DOL has a validation procedure to standardize the allowable values of data elements in the submitted files. We obtained WIPS data from program year (PY) 2018 through the second quarter of program year 2021 (PY 2021 Q2) for all RP and Wagner-Peyser participants. Because program years start in the third quarter of each calendar year (that is, July 1), these data cover people who received services through RP and Wagner-Peyser between July 2018 and December 2021. The completeness and consistency of WIPS data varies across states; our research design, centered around a matching approach, compares individuals within states, which will help limit the potential bias from these discrepancies. However, we note that match quality and the comprehensiveness of our sample might differ across states and might be more reliable in states with more complete data.[20]

We used WIPS data at three stages of the impact study's design. First, we used personally identifiable information (PII, namely, Social Security numbers, names, and dates of birth) from the WIPS as part of the broader data collection process to obtain employment, earnings, and criminal justice data from a range of

[19] The full list of data elements included in the WIPS is available at https://www.dol.gov/sites/dolgov/files/ETA/Performance/pdfs/ETA_9172_DOL_PIRL_1.18.18.pdf.

[20] RP grantees were just beginning to use the WIPS for data collection during the time period of interest for this study. Although this does not affect the matched comparison design or impact estimates in this report, there may have been potential inaccuracies in service receipt data collected through the WIPS for RP grant participants, which could have resulted in low quality service receipt data for RP participants. WIPS variables with high rates of missing data or other quality issues were not included in the analyses presented in this report. WIPS reporting requirements and systems were updated during the quarter ending on September 30, 2022.

sources (described in further detail below). Note that while RP grantees collected Social Security numbers from participants, individuals could refuse to provide that information to the grantee, in which case we would drop them from the analysis. Second, we used WIPS data to measure the program quarter of enrollment, state and county of residence, and background characteristics of individuals who enrolled in RP and Wagner-Peyser programs, which informed our matched comparison design. Third, we used WIPS data on demographic characteristics and service receipt to define subgroups for analysis.

Although WIPS data for both RP and Wagner-Peyser include a self-reported indicator for prior criminal justice involvement, this indicator substantially undercounts the number of justice system-involved individuals and so was not used in the analysis. Among the pool of RP and Wagner-Peyser participants matched to state criminal justice data, only 56 percent self-identified as having prior justice involvement.[21]

2. National Directory of New Hires

We obtained quarterly employment and earnings data from the National Directory of New Hires (NDNH), a database maintained by the Office of Child Support Services (OCSS) at the U.S. Department of Health and Human Services. We used these data to generate two constructs that serve as outcomes for the following primary research questions: (1) employment in the 9th and 10th quarters after program enrollment, and (2) average quarterly earnings in the 9th and 10th quarters after program enrollment. We also used these data to examine employment and earnings over the full period after enrollment as secondary outcome measures.

The NDNH contains approximately two years of past earnings and employment data at any given time. We obtained NDNH data by submitting a series of match files—consisting of a list of Social Security numbers (SSNs) and names—to OCSS. OCSS then held data for the individuals included in the request, containing data on the eight quarters available before the match file submission date. We submitted match files on a rolling basis with SSNs and names obtained from RP grantees and state workforce agencies as we received that data.

We submitted match files beginning in June 2022, and depending on the state, the NDNH data we received included complete earnings and employment records starting between 2020 Q3 and 2021 Q4, and ending in 2023 Q3. Because RP participants enrolled in programs between 2018 and 2020, the NDNH data did not cover the pre-enrollment period for most RP participants. Thus, we could not use employment and earnings data as pre-program variables in the matching process for the full impact study sample. Pre-program earnings are generally critical matching variables for impact evaluations of employment programs. However, in this case, many sample members (especially within our adult RP sample) had been incarcerated for a large portion of the relevant pre-program time period and thus do not have applicable earnings. Later in this appendix, we explore how our estimated impacts change when we include the pre-program earnings and employment data at our disposal in our matching design.

Limitations. NDNH data contain outcomes only for people with reportable earnings in covered jobs. Although these data cover most wage and salary employment, they do not cover all types of jobs and

[21] This differed between RP and Wagner-Peyser participants: 83 percent of RP participants and 55 percent of Wagner-Peyser participants that matched to state criminal justice data reported prior justice involvement in the WIPS.

industries. In particular, NDNH data only cover earnings submitted to unemployment insurance (UI) agencies. NDNH does not contain data on self-employed workers, most agricultural workers, some domestic service workers, or part-time employees of nonprofit organizations (Czajka et al. 2018). In the past, these sectors have made up about 10 percent of U.S. employment (Kornfeld and Bloom 1999; Hotz and Scholz 2001). NDNH data also omit workers whose employers do not report their earnings to their state UI agency, even in the formal sector (Abraham et al. 2018; Blakemore et al. 1996; Hotz and Scholz 2001; Houseman 2001; Katz and Krueger 2016, 2019). Additionally, NDNH data do not cover workers who are casually employed, such as day laborers, and exclude most work that is part of the gig economy (Abraham et al. 2018; Katz and Kruger 2016, 2019). Because we cannot distinguish between people who are truly unemployed and those who are employed, but do not have reportable earnings, we assume that anyone in the study sample who is not in the NDNH data during a given quarter is not employed and has no wage earnings in that quarter. Future research could potentially address these limitations of the NDNH data by augmenting these records with other sources of earnings or income data that cover self-employment and gig work, such as tax records.

3. Criminal justice data

Our impact study leveraged criminal justice data to achieve two goals. First, we used individuals' pre-program criminal justice involvement as part of our matching approach to identify suitably comparable program and comparison group members. Second, to fully understand the impacts of RP on its justice-involved participants, we used administrative criminal records to construct measures of participant recidivism. For each of the six states in our impact study sample, we aimed to collect three sets of criminal justice records: (1) criminal court data identifying criminal charges individuals faced and their dispositions; (2) arrest records identifying arrest events involving individuals in our sample; and (3) state incarceration records identifying the period(s) during which individuals were incarcerated in state prisons.

Not every state could provide all three data elements, as we discuss in depth throughout this section. Conditional on providing data, different state agencies used different data formats and included different data elements in their extracts. Our cleaning efforts standardized these datasets so we could use them to create the comparison group and analyze outcomes related to recidivism. The remainder of this section summarizes the criminal justice administrative data collection and cleaning process.

Criminal justice data collection process

For each state, we attempted to collect data covering our three key domains—criminal charges, arrests, and incarcerations in state prisons—for program group and potential comparison group participants. For some states, this process involved interacting with two separate agencies (typically where arrest and conviction data were housed by the same agency), and in other cases it involved reaching out to three agencies (one for each type of data). With each agency, the study team completed a research request and application process. After clearing any other approval hurdles (for example, staff background checks and security clearances, and the signing of any data use agreements), the study team provided the agency with key identifiers for each selected RP participant and potential Wagner-Peyser comparison group member. The agencies then returned the relevant records for a time period of at least two years before the earliest enrollment date for participants in that state up through the time of the data pull.

Although we pursued criminal charge, arrest, and state incarceration data from all six states in our sample, we encountered various obstacles that prevented us from collecting all three types of data from every state. These hurdles included the following:

- Some agencies were unable to comply with data requests due to administrative burden and or their own limited internal capacity. At the time we began collecting data, many agencies were adapting to new, pandemic-related working environments or dealing with many of the issues that befell correctional system agencies during the COVID-19 pandemic.
- Some agencies were unwilling or legally prevented from providing data. We received replies indicating that particular regulations or restrictions impaired states' ability to provide data.

Note that these limitations also precluded us from collecting criminal justice records from five states and territories that we had originally hoped to include in our study; the lack of justice data for these locations in part led us to remove them from the sample and focus on the six states that could provide adequate data. Exhibit A.1 shows the final list of states and criminal justice agencies that provided data.

Exhibit A.1. Sources of criminal justice data, by type

State	Criminal charge data	Arrest data	State prison data
Alabama	AL Administrative Office of the Courts	AL Administrative Office of the Courts	AL Dept. of Corrections
Florida	FL Dept. of Law Enforcement	FL Dept. of Law Enforcement	FL Dept. of Corrections
New Jersey	NJ Courts	No data provided	NJ Dept. of Corrections
New York	NY Division of Criminal Justice Services	NY Division of Criminal Justice Services	No data provided
Oregon	OR Criminal Justice Commission	No data provided	OR Dept. of Corrections
Pennsylvania	PA Administrative Office of the Courts	PA Administrative Office of the Courts	PA Dept. of Corrections

Criminal justice data cleaning

After receiving the data, we ran preliminary descriptive analyses to check the contents of the various datasets, double check the minimum and maximum date ranges of included events, and determine the share of individuals from the WIPS who matched with the various criminal justice datasets (see Exhibits A.2 to A.4). We then stripped PII from the datasets and all data were assigned a study ID. The study team then created analysis variables, including indicators for conviction, arrest, and incarceration.

1. **Criminal charge data.** Our ultimate goal was to use the criminal charge records we collected from state courts to identify any criminal cases that were filed, regardless of their eventual disposition, and the cases that specifically resulted in conviction. The raw data from most states included charge-level records, meaning that the data included information on individual offenses that the courts would then bundle into criminal cases. We ultimately aggregated these charge-level records to the "event" level, where events included multiple cases (and all included charges) disposed within the same calendar quarter.

Although the data remained at the charge level, we took two steps to classify the offenses an individual was accused of and their dispositions in the judicial process:

2. **Categorizing offenses by type:** Criminal offenses fall into a handful of categories based on the nature of the underlying crime. Criminal justice researchers typically summarized charges as either relating to property offenses (such as burglary or theft), violent offenses (such as assault), drug offenses (such as drug possession), traffic offenses (such as driving under the influence [DUI or DWI]), and "other" offenses, usually including violations of ordinances or other "crimes against society" (such as loitering or intoxication).

   To categorize offenses, we used the Text-based Offense Classification (TOC) tool, an online machine-learning tool created by the Criminal Justice Administrative Records System (CJARS) (Choi et al. 2023). Using this online tool helped us categorize offenses consistently, despite idiosyncrasies in the charge descriptions that different states provided. To use TOC, we submitted a plain, single-column file that contained a unique list of criminal charge descriptions found in the data. In the rare cases we had to manually categorize the criminal charges, we used the various outputs of TOC as reference to maintain consistency in categorizations. TOC—which CJARS trained on an array of similar text-based charge descriptions—would then identify the charge categorization most likely to fit the text descriptions that the state courts provided. Because TOC actually has finer classifications than we planned to use for this report, we amended the TOC output slightly to combine traffic and DWI offenses into a single "traffic" offense type, because sex offenses and non-DWI traffic offenses were both quite rare in our sample. We treated offenses that TOC could not categorize (only 1 percent of the total) as "other" offenses.

3. **Categorizing offenses by severity:** The charge data include fields indicating whether a given charge constitutes a felony (relatively severe charges) or a misdemeanor offense (relatively minor charges that rarely involve prison time). We compiled these fields into an indicator for whether a given charge involved a felony offense.

4. **Categorizing charge dispositions:** Each charge in state court records corresponds to a particular disposition, indicating whether the person involved was convicted or if the charge against them was dismissed without penalty. We defined a conviction to be any disposition in which the defendant was found or pled guilty, including dispositions like "convicted," "guilty plea," "guilty verdict," and "nolo contendere."

After constructing these charge-level variables, we then aggregated the data to the "event" level using disposition dates. These "events" approximate criminal cases—they represent all charges against a single individual that were disposed within a single quarter. Intuitively, this aggregation step groups together all charges that we expect the courts to have handled as a single unit, which avoids misconstruing charges that were processed and even disposed together as part of separate "cases." We created indicators for whether a criminal court event includes charges of each given type (property, violent, drug, traffic, and other) and at least one felony charge. We also defined an indicator for whether any charge contained in the event led to a conviction. We defined the event disposition date as the last date on which a charge contained in the event was disposed by the courts.

We matched these aggregate, event-level criminal court records to our WIPS data. Because of differences in the exact contents of state criminal justice data (for example, some states included charges that did not result in conviction, while others did not), as well as unobservable differences in how states conducted their matches, we anticipated some variation in match rates across states. We summarize the results of our match in Exhibit A.2. Note that for all of our criminal justice datasets, we expected a much higher match rate for RP participants than Wagner-Peyser participants, because the former program primarily aimed to recruit individuals with prior criminal justice contacts, while the latter is open to all job seekers.[22] Although we tended to observe lower match rates in states that we knew sealed cases that did not result in conviction (New Jersey, New York, and Pennsylvania), we also had a relatively low match rate in Florida, a state that we believe seals few records. This variation in match rates also relates to our design decisions to include only RP and Wagner-Peyser participants for whom we observed prior criminal charges (rather than those with observed arrests, but no criminal charges for example), and our choice to match exactly on participant state. These decisions should have helped reduce potential measurement error stemming from imperfect and inconsistent match quality.

Using our matched WIPS-court records data, we compared event disposition dates to program enrollment dates to determine whether a given court event was before or after a persons' enrollment in RP or Wagner-Peyser programs. We then grouped events into three categories based on their timing relative to program enrollment: (1) the most recent charges the person faced before enrolling in the program; (2) any other charges the person faced in the five years before program entry, besides the most recent ones; and (3) any charges over the 5 and 10 quarters after program enrollment. These groupings provided the basis for our principal pre-program and outcome variables that we discussed in the main text of this report.

Exhibit A.2. Summary of criminal court data by state

	RP			WP		
State	**Matched records**	**Total records submitted**	**Match rate**	**Matched records**	**Total records submitted**	**Match rate**
Alabama	377	629	59.9%	10,005	24,099	41.5%
Florida	680	1,148	59.2%	25,277	101,703	24.9%
New Jersey	74	148	50.0%	4,110	12,297	33.4%
New York	1,053	1,644	64.1%	27,922	80,978	34.5%
Oregon	287	428	67.1%	14,091	29,011	48.6%
Pennsylvania	798	1,267	63.0%	12,860	55,912	23.0%

Source: State criminal court records.

Note: Some RP and WP participants who matched to court records only did so for events occurring after program enrollment. These participants were excluded from the analysis sample as described further below.

RP = Reentry Project; WP = Wagner-Peyser.

[22] As we described further in Section B, we conducted a first round of matching to identify a group of Wagner-Peyser participants for whom we would obtain state criminal justice records. For this reason, these match rates do not necessarily reflect the level of justice involvement among all Wagner-Peyser participants. We sent state criminal justice agencies PII for all Wagner-Peyser participants with similar observable demographic characteristics as the RP RP participants in our sample, as well as any Wagner-Peyser participants who self-reported having prior criminal justice involvement in the WIPS.

Arrest data. Our arrest records closely resembled our criminal charge data, unsurprisingly, given that most originated from the same state agencies. The raw data contained offense-level arrest information, indicating which offense(s) the arresting officer believed an individual had committed. However, unlike the charge data, we had modest goals for using the arrest data that made our data cleaning relatively straightforward. As with our criminal charge data, we used the arrest dates present in the data to aggregate offense-level data to an "arrest event" level, including all arrests for a person within a given quarter. As we did with our charge data, we matched records to WIPS (the results of which appear in Exhibit A.3) to identify (1) the most recent arrest of a person before they enrolled in the program; (2) any other arrests the person experienced in the five years before program entry, besides the most recent one; and (3) any arrests over the 5 and 10 quarters after program enrollment.

Exhibit A.3. Summary of arrest data by state

	RP			WP		
State	Matched records	Total records submitted	Match rate	Matched records	Total records submitted	Match rate
Alabama	387	629	61.5%	10,301	24,099	42.7%
Florida	708	1,148	61.7%	28,775	101,703	28.3%
New York	1,097	1,644	66.7%	28,909	80,978	35.7%
Pennsylvania	713	1,267	56.3%	9,576	55,912	17.1%

Source: State arrest records.

RP = Reentry Project; WP = Wagner-Peyser.

State prison data. The incarceration data provided the start and end dates of each incarceration spell in a state correctional facility, not counting intermittent releases due to court appearances or hospitalizations. As with our arrest data, we took these incarceration spells and matched them to the WIPS to determine whether they occurred before or after a person enrolled in RP or Wagner-Peyser programs (we summarize the match results in Exhibit A.4). We then identified (1) the most recent state prison stay of a person before they enrolled in the program; (2) any other incarcerations in state prison the person experienced in the five years before program entry, besides the most recent one; and (3) any new incarceration spells over the 5 and 10 quarters after program enrollment.

Exhibit A.4. Summary of state prison data by state

	RP			WP		
State	Matched records	Total records submitted	Match rate	Matched records	Total records submitted	Match rate
Alabama	171	629	27.2%	4,909	24,099	20.4%
Florida	320	1,148	27.9%	9,350	101,703	9.2%
New Jersey	21	148	14.2%	3,315	12,297	27.0%
Oregon	106	428	24.8%	3,248	29,011	11.2%
Pennsylvania	280	1,267	22.1%	5,701	55,912	10.2%

Source: State prison records.

RP = Reentry Project; WP = Wagner-Peyser.

Criminal justice data limitations

The data that our team collected provide an informative snapshot of criminal justice involvement among our sampled RP participants and comparison Wagner-Peyser participants. However, these data have notable omissions that affect our research design and interpretation of our findings, as we discuss in Chapters 3 and 4:[23]

- **Sealed and expunged records.** By default, court systems seal or expunge particular records from their files. These include juvenile records, which all states seal by default. Some states also remove charges that did not result in conviction, such as those dismissed by the prosecution. In practice, these omissions limited our ability to evaluate the true extent of individuals' involvement with the criminal justice system. Particularly for young adults, we might not have accurately observed criminal justice contacts over the five years leading up to program enrollment, because any juvenile arrests, charges, and incarceration would not have been reported. For both adults and young adults, we also likely undercount actual criminal charges faced, because several states in our sample (to our knowledge, New York, New Jersey, and Pennsylvania) automatically remove charges that did not result in conviction. Even within observed criminal court "events," we might tend to misreport the relative prevalence of different charge types, because we cannot reliably observe all charges filed against an individual. These missing records prompted us to restrict our sample to just those individuals for whom we observed pre-program criminal charges, as noted in Chapter 3.
- **Lack of local jail data.** Although we observe incarceration spells in state prisons, we do not observe incarcerations in local jails. It would have been impractical to collect these data because they are held by individual county authorities. However, without local jail data, we almost certainly undercounted incarcerations among our sample, both before and after program enrollment. Similarly, because we could not distinguish people who had no history of incarceration from those who had previously been incarcerated, but only in county jails, we may have inadvertently matched individuals who had and had not been incarcerated in local jails. Per our discussion in Chapter 4, these imperfect matches may have biased our final results.
- **Lack of reliable sentencing data.** Our criminal court data provide us with descriptions of charges that individuals faced (and led to conviction, for certain states). However, these data do not provide reliable information about the sentences imposed and experienced by people in our sample. In particular, we cannot observe whether individuals were sentenced to incarceration or probation, nor do we have suitable data to examine the sentences people actually served (for example, how long they spent in local jail before being released on parole). We therefore cannot observe whether a person met the eligibility criteria for adult RP programs—namely, we do not know whether they were sentenced to incarceration or supervised probation. Consequently, we cannot observe exactly why a person qualified for RP services, nor can we restrict our sample of candidate Wagner-Peyser comparison group members to those who would have been eligible for RP. As noted in Chapter 4, both factors may have biased our estimates.

[23] Generally, while the degree to which these data were missing did vary in some cases across states, we did not have complete data along these dimensions in any analysis state.

Beyond these specific instances of missing data that affect our research design, we faced more general limitations with criminal justice data, including the following:

- We only attempted to match individuals to criminal justice records from the state in which they enrolled in RP or Wagner-Peyser programs. If someone experienced justice system involvement in a state other than the one in which they had enrolled in the program, we would not observe those contacts. Thus, the data do not capture all justice system involvement, only involvement within the states where the participants were first enrolled.
- Arrest data are often incomplete. States have different reporting standards—for example, some states do not report arrests that do not lead to charges. We likely underreport arrests in most state data, which is one reason why we focus more on criminal charge data to construct recidivism outcomes and tailor our analytic sample.

4. Data linkage process

To link the three data sources used for the impact study, we used three types of PII: SSNs, full names, and dates of birth. Exhibit A.5 indicates the three types of PII, what they will be used for, and the sources of those data for RP and Wagner-Peyser participants.

To collect quarterly employment and earnings data from NDNH, we used names and SSNs. To link study sample members to criminal justice data, we used names and dates of birth, as well as SSNs for some agencies that requested it. SSNs for RP participants are available in the WIPS data. We requested names and dates of birth for RP participants from grantees, linking them to the unique identifiers that grantees submit to the WIPS (WIPS IDs). For Wagner-Peyser participants, we requested SSNs, names, and dates of birth from state workforce agencies, because SSNs are not recorded in the WIPS for Wagner-Peyser participants.[24]

Exhibit A.5. Sources of key data elements for NDNH and criminal justice linkage

PII data element	Needed to obtain	Data source	
		RP participants	WP participants
SSN	NDNH data	WIPS database	State workforce agencies
Name	NDNH and criminal justice data	Grantees	State workforce agencies
Date of birth	Criminal justice data	Grantees	State workforce agencies

Note: Names are not required for the NDNH match but are requested to improve the accuracy of the match. Names are required for linking to criminal justice data.

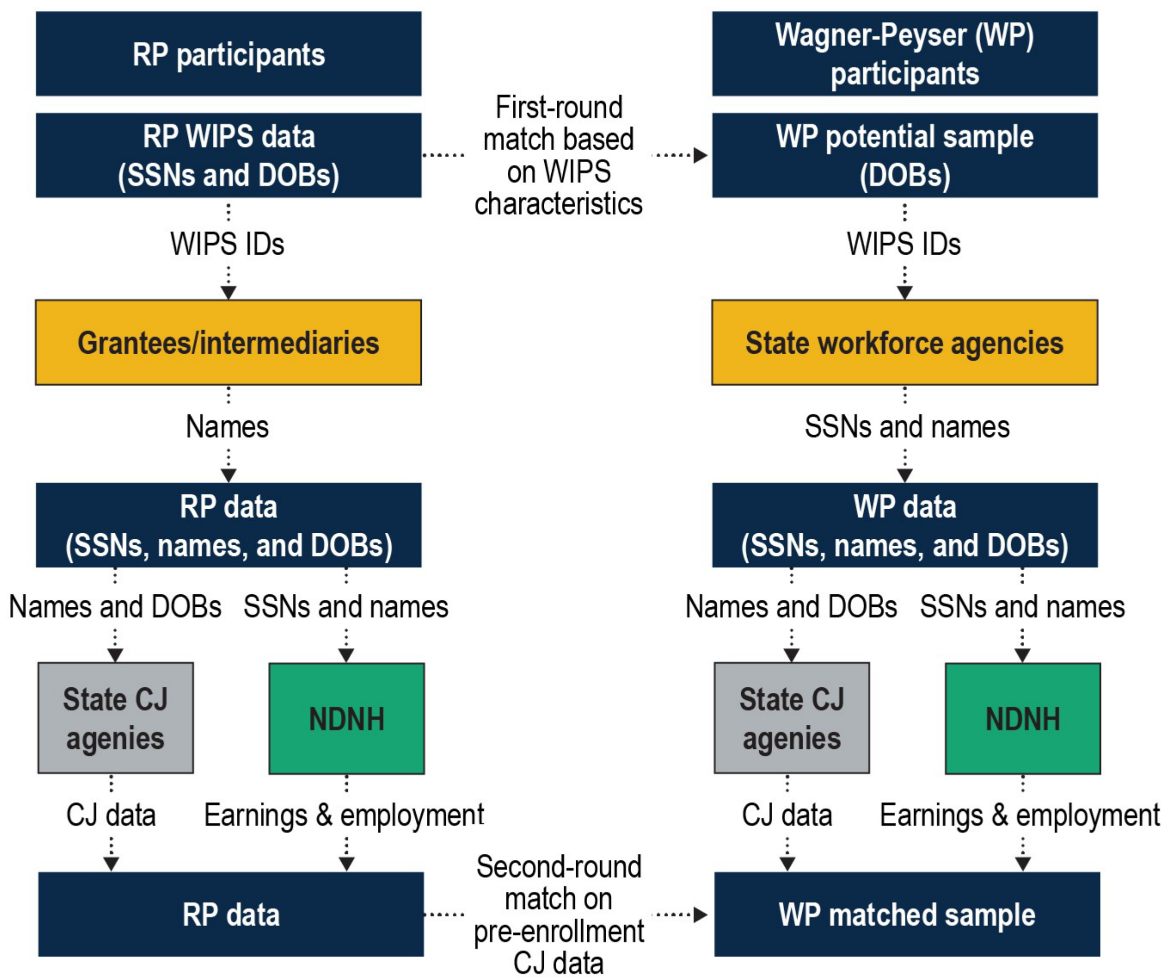
NDNH = National Directory of New Hires; PII = Personally Identifiable Information; RP = Reentry Project; SSN = Social Security Number; WIPS = Workforce Integrated Performance System; WP = Wagner-Peyser.

Exhibit A.6 illustrates the process we used to link various data sources for the program and comparison groups and the process we used to create a comparison sample. We collected data and created a

[24] Dates of birth are recorded in the WIPS for both RP and Wagner-Peyser participants, and we used these data to calculate participant age, a key matching variable. However, during the data collection process, it was not clear if we would have permission to redisclose dates of birth as reported in the WIPS to criminal justice agencies to link to criminal justice records. For this reason, we requested dates of birth from both RP grantees and state workforce agencies for criminal justice data collection.

comparison sample through a two-stage process. In the first stage, we obtained PII for RP participants from the WIPS and RP grantees. We then performed a first-round match (described in greater detail in Section B) to identify a broad pool of potential comparison group members using WIPS data. Using PII collected for both RP participants and potential comparison group members, we then collected criminal justice data from state agencies, as described above, to use pre-program criminal justice variables to construct a final comparison group.

Exhibit A.6. Data collection and linking process

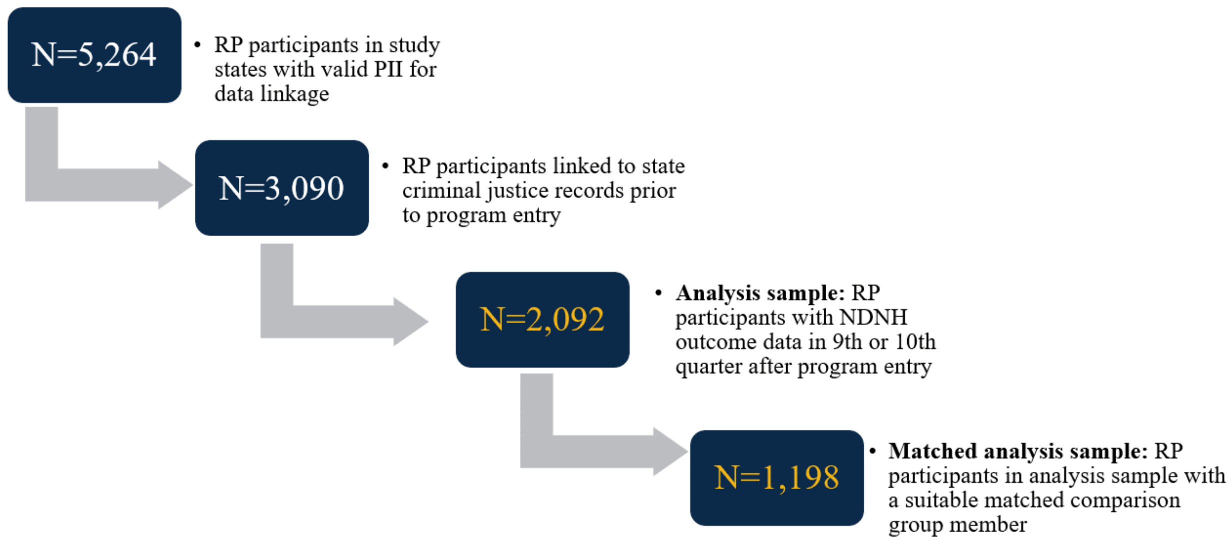


CJ = criminal justice; DOB = date of birth; NDNH = National Directory of New Hires; RP = Reentry Project; SSN = Social Security Number; WIPS = Workforce Integrated Performance System; WP = Wagner-Peyser.

Sample inclusion criteria for RP participants

As described in Chapter 3, within our six study states, the RP participants who were ultimately included in the analysis sample were those who met the following criteria: (1) the RP participant had name, date of birth, and a valid SSN for use in data linkage; (2) they were observed in criminal justice data before program entry; and (3) based on the timing of NDNH match file submission relative to their quarter of entry, the NDNH data coverage for the RP participant includes the 9th or 10th quarter after program entry. Each of these criteria was associated with sample loss, as summarized in Exhibit A.7. Our final analysis sample consisted of the 2,092 RP participants for whom we observed both pre-program criminal justice histories and our confirmatory 9th and 10th quarter outcomes. All impact analyses are restricted to the subset of this analysis sample for whom we could identify a suitable matched comparison group member, which was ultimately 1,198 RP participants. The next section describes our process for constructing the matched comparison sample of Wagner-Peyser participants.

Exhibit A.7. RP sample inclusion criteria and sample sizes



NDNH = National Directory of New Hires; PII = personally identifiable information; RP = Reentry Project.

B. Matched comparison design

We used a two-stage matching procedure to identify a matched comparison group for analysis. Due to limitations in access to the identifying information needed to link Wagner-Peyser participants to criminal justice and labor market outcome data, we used a preliminary first stage matching process to identify a subset of Wagner-Peyser participants for whom we would obtain state criminal justice records and NDNH earnings and employment information. After receiving these data, we then conducted a full matching procedure (the second stage) to identify a comparison group of Wagner-Peyser participants with similar observed characteristics to the RP study sample.

1. First stage: Identifying potential match pool

As described in Section A, full names and SSNs were not available in the WIPS data for Wagner-Peyser participants. Due to the large number of individuals enrolled in Wagner-Peyser, it was not feasible to collect this PII for all Wagner-Peyser participants in the impact study states. Therefore, to obtain state criminal justice data and NDNH earnings and employment information for the comparison group, we had to narrow down the size of the potential match pool. To do this, we conducted a first-stage matching using WIPS data to select a subset of Wagner-Peyser participants who were similar to RP participants based on their geographic location, demographic characteristics, and timing of enrollment in Wagner-Peyser.

We used coarsened exact matching (CEM) to match RP participants to Wagner-Peyser participants residing in the same state and enrolled in the same program quarter with whom they shared the closest overlap in demographic characteristics measured in the WIPS data. In addition to requiring an exact match on geographic location and program enrollment quarter, we constructed matching cells based on the following demographic characteristics: age at program entry, gender, education level, race and ethnicity, employment status at program entry, low-income status at program entry, English learner status, receipt of dislocated worker services, veteran status, and disability status.

We matched individuals without replacement. For a given RP participant, we first identified the pool of Wagner-Peyser participants residing in the same county with the same quarter of enrollment. We then selected all Wagner-Peyser participants in that cell who exactly matched on the key demographic characteristics described above. In addition to matching exactly on county, we conducted a second round of matching using demographic characteristics to identify additional Wagner-Peyser participants with the same state and program quarter of entry as each RP participant who were not already identified as a match in the first step. Finally, we selected all Wagner-Peyser participants flagged in the WIPS data as having prior criminal justice involvement who were not already selected as a match.

We aimed to reduce the size of the potential Wagner-Peyser participant match pool by excluding any Wagner-Peyser participants with large differences in their demographic backgrounds compared to the RP participant sample. We iterated through the matching process until the overall number of Wagner-Peyser participants identified was suitable in size. Exhibit A.8 summarizes the total number of Wagner-Peyser participants selected in the first-stage match.

Exhibit A.8. Wagner-Peyser matched comparison pool after first-round match, by state

State	RP	WP
AL	629	15,870
FL	1,148	89,518
NJ	148	7,514
NY	1,644	72,069
OR	428	21,085
PA	1,267	48,497
Total	**5,264**	**254,553**

Source: WIPS data.

RP = Reentry Project; WIPS = Workforce Integrated Performance System; WP = Wagner-Peyser.

We submitted the WIPS IDs for the Wagner-Peyser participants identified in this first round of matching to state workforce agencies to obtain SSNs, names, and dates of birth. As described in Section A, we used this PII to request matched records from state criminal justice agencies and the NDNH database. After receiving criminal justice data from each state, we limited the comparison group pool for second stage matching to the subset of Wagner-Peyser participants for whom we observed a criminal charge before program enrollment based on the state criminal justice records. This limited the pool to 83,352 Wagner-Peyser participants.

2. Second stage: Propensity score estimation and matching

In the second stage of matching, we identified the final matched comparison group of Wagner-Peyser participants for analysis. We first excluded any Wagner-Peyser and RP participants for whom we could not find any pre-program criminal charges in state court records. Among this sample of previously charged individuals, we matched Wagner-Peyser participants to RP participants on a range of variables measuring pre-program justice involvement as well as demographic and county characteristics.

We used a partial exact matching approach followed by propensity score matching with a caliper. Using a propensity score model enabled us to construct a summary measure of the likelihood of participation in RP that could incorporate the large number of variables available in the criminal justice and WIPS data.

Matching variables

As described in the report, the RP eligibility criteria related to criminal justice history differed for participants served by RP adult grantees (focused on individuals ages 25 and older) and those served by RP young adult grantees (focused on individuals ages 18 to 24). Adults participating in RP were required to have been previously incarcerated or released from prison or jail within 180 days of enrollment or currently under supervision. Young adults participating in RP were required to have been currently or previously involved with the justice system, including the juvenile justice system. To account for these differences in populations and potential differences in the enrollment processes, we performed matching separately for these two groups.

For both groups, we included a range of variables in the matching process to identify a comparison group with similar observable criminal justice backgrounds and other characteristics.

- **Exact match variables.** We required an exact match between RP and Wagner-Peyser participants, based on their state and quarter of enrollment, to identify a comparison group that likely faced similar labor market conditions and criminal justice contexts at the time of entry into their programs.[25] Requiring an exact match on program quarter of entry also ensures the intervention and comparison group have comparable levels of pre- and post-program exposure to the COVID-19 pandemic context. We required exact match on gender due to differences in criminal justice backgrounds, reentry contexts, and the differential impact of the COVID-19 pandemic by gender. We also required an exact match on self-reported employment status at entry, as measured in the WIPS. Lastly, we included key features of their most recent criminal case as exact match variables. Specifically, we included variables capturing whether the person was convicted, whether they faced a felony charge, and whether they entered the RP or Wagner-Peyser programs within three quarters of a criminal case disposition or a release from state prison.
- **Propensity score matching variables**. We included an extensive list of candidate variables for possible selection in the propensity score models. These variables captured information on demographic

[25] We explored the feasibility of conducting an exact match on county of residence. Due to the small sample sizes within a county, there was minimal overlap in covariates of interest between RP and Wagner-Peyser participants residing in the same county. This resulted in substantial sample loss when conducting matching within county. For this reason, we pooled our sample by state and included county characteristics in the propensity score model to capture variation in county contexts.

background, county characteristics, and criminal histories as observed in arrest, incarceration, and criminal court records. The demographic matching variables included age, disability status, race and ethnicity, and education level. County characteristics included population, urbanicity, and unemployment rate from the Census American Community Survey and the county crime rate and arrest rates per capita based on Uniform Crime Reporting (UCR) data. Criminal history matching variables included information on an individual's most recent criminal case before program entry, including the type of criminal charge (violent, drug, property, traffic, or DUI), the class (felony or misdemeanor), whether the person was convicted, and a measure of the severity of the convicted charge. We also included variables measuring the number of felony charges and convictions in the five years before program entry. In addition to this information on criminal cases, matching variables included information on observable incarceration and arrest histories, such as whether the person had an observable incarceration spell before program entry, the duration of their most recent incarceration spell, the total number of arrests in the five years before program entry, and the number of offenses associated with their most recent arrest before program entry.

Propensity score estimation approaches

We estimated propensity score models using the pool of RP participants and Wagner-Peyser participants selected in the first round of matching who also had a record of pre-program criminal charges. We estimated the probability that each of these individuals participated in RP (as opposed to Wagner-Peyser) using the observed characteristics described above. We conducted this estimation separately for RP adult and young adult grant participants. In all propensity score models, we included both the exact match and propensity score matching variables listed above as covariates.

We estimated propensity scores using several methods that are designed to select the optimal predictors of treatment from a large number of covariates and possible interactions between them.

- **Generalized boosted regression model (GBM)**: a machine-learning approach that uses an algorithm to search over the set of provided covariates and select the interactions and data partitions that most predict participation (McCaffrey et al. 2004). In this nonparametric approach, the algorithm generates and includes interactions and higher-order terms of the covariates. We used the Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG) to implement the GBM method (Griffin et al. 2014). A benefit of GBM is that the algorithm automatically incorporates nonlinearities and interactions between covariates, and it is specifically designed to optimize covariate balance. This method has been shown to lead to impact estimates with lower bias and higher efficiency than other propensity score estimation methods (Griffin et al. 2014).
- **Bayesian additive regression trees (BART):** a machine-learning method that uses a Bayesian statistical model to iterate over a series of regression trees to identify covariates and interactions that best fit the data (Chipman et al. 2010). Like GBM, BART is nonparametric and incorporates nonlinearity and higher order interactions among the covariates included for selection in the model. One benefit of the BART approach is that its flexibility enables it to account for differences in the relationship between covariates and the propensity score across subgroups (for example, different relationships between the covariates and the propensity score across states).

- **Double-selection least absolute shrinkage and selection operator (LASSO):** a parametric machine-learning technique that selects covariates and interaction terms among a set of specified variables. The LASSO regression limits the number of covariates by penalizing each additional covariate added to the model. We used double-selection LASSO (Belloni et al. 2014), which selects covariates based on their ability to predict both the probability of treatment and the outcome. We focused on the confirmatory recidivism outcome of conviction in the 9th and 10th quarters after enrollment for the double-selection LASSO method. We required LASSO to select all exact matching covariates, and included all other propensity score matching variables as potential candidates for variable selection.

We also fit propensity score models using logistic regression with researcher-specified covariates, which served as a baseline comparison for the machine learning approaches.

Matching approach

After estimating the propensity scores, we identified the final matched comparison pool using a partial exact match followed by caliper matching. First, we constructed exact match strata based on the set of variables specified above. Within these strata, we used caliper matching to identify all Wagner-Peyser participants with a propensity score that falls within a specified distance (the caliper) to each RP participant. We conducted matching with replacement, meaning that a Wagner-Peyser participant could be selected as a matched comparison for more than one RP participant.

We specified a caliper of 0.2 times (20 percent of) the standard deviation of the logit of the propensity score, a width that has been shown in the literature to minimize the bias in the estimated treatment effect (Rosenbaum and Rubin 1985; Austin 2011; Wang et al. 2013). The size of the caliper determines the selectivity of the matching procedure—a caliper that is too wide may result in insufficiently balanced treatment and comparison group samples as it will include comparison group members that are not sufficiently similar on covariates to the treatment group. However, a caliper that is too narrow can result in excessive sample loss if there are not sufficient comparison group members within the caliper for certain RP participants, leading to imprecision in impact estimates. To address this tradeoff, we additionally assessed the covariate balance and sensitivity of our results under a narrower caliper of 0.1 times the standard deviation of the logit of the propensity score.

We constructed weights for analysis that were proportional to the number of times a Wagner-Peyser participant was selected as a match. We then rescaled the Wagner-Peyser weights to sum to the number of RP participants in the matching pool to recover impact estimates measuring the average treatment effect on the treated (ATT).

Selecting the estimation approach and caliper

The goal of the matching process is to construct treatment and comparison groups that are similar based on pre-program characteristics. We evaluated the range of propensity score estimation approaches and caliper widths to determine which resulted in a comparison group with the smallest differences in observed characteristics.

For each pre-program characteristic, we calculated the standardized mean difference between the matched RP and Wagner-Peyser samples, as well as the *p*-value resulting from t-tests for each difference.

However, due to the large number of candidate covariates (117 total), we assessed overall covariate balance across the methods using the prognostic score as a summary measure. The prognostic score has been shown in simulations to outperform selection based on comparisons of means across covariates (Stuart et al. 2013).

We calculated the prognostic score by estimating a regression model to predict an outcome (in our case, each of the three primary outcomes), using only the comparison group. We then predicted the average outcome for the treatment group using the coefficients estimated from the model using the comparison group only. Finally, we compared the mean predicted values for the two study groups (Zhang et al. 2019). A smaller difference in mean outcomes for a given propensity score model versus another model indicates that the model leads to better overall covariate balance, incorporating information on differences in means of the covariates between the study groups and how those covariates are associated with the outcome. We estimated prognostic scores using each of our three confirmatory outcomes: earnings and employment in the 9th and 10th quarters after enrollment and convictions over the 10 quarters after enrollment.

Exhibit A.9 presents the prognostic score differences for each of the propensity score estimation approaches and calipers we considered. Because we estimated propensity scores and conducted matching separately for the adult and young adult RP grant participants, we calculated prognostic scores separately for each group. Based on the results of our sample comparisons, we selected the LASSO approach for propensity score estimation with a 0.2 standard deviation caliper as our primary approach. This minimized the prognostic score difference for the confirmatory earnings outcome in the adult sample and performed well in the young adult sample. As discussed in Chapter 3 and presented in Appendix Section D, we conducted sensitivity analyses to determine the robustness of our impact estimates to alternative matching approaches.

The LASSO regression technique has the advantage of winnowing down our extensive list of pre-program variables to just those covariates that convey the most information. In this sense, the LASSO offers a degree of flexibility in that we can use this approach to refine our propensity score model for specific subgroups across which different characteristics might predict RP eligibility more strongly. In particular, we used LASSO to choose optimal covariates and estimate propensity scores separately for our adult and young adult samples, who, as discussed above, have very different eligibility criteria that might affect the strength of different predictors of program status. Ultimately, we used LASSO to specify two separate propensity score models—one for adults and one for young adults—which we then combined to arrive at our final propensity scores.

Exhibit A.9. Prognostic scores for candidate estimation approaches and calipers

Estimation method	Prognostic score outcome	Prognostic score standardized difference	
		RP Adult	RP Young Adult
Caliper=0.2			
BART	Earnings (9th and 10th quarter)	0.037	0.062
BART	Employment (9th and 10th quarter)	0.015	0.098
BART	Conviction (10 quarters)	0.007	0.067
Logit	Earnings (9th and 10th quarter)	0.026	0.000
Logit	Employment (9th and 10th quarter)	0.017	0.027
Logit	Conviction (10 quarters)	0.069	0.065
LASSO	Earnings (9th and 10th quarter)	0.003	0.028
LASSO	Employment (9th and 10th quarter)	0.005	0.040
LASSO	Conviction (10 quarters)	0.093	0.059
GBM	Earnings (9th and 10th quarter)	0.010	0.108
GBM	Employment (9th and 10th quarter)	0.005	0.080
GBM	Conviction (10 quarters)	0.048	0.006
Caliper=0.1			
BART	Earnings (9th and 10th quarter)	0.052	0.098
BART	Employment (9th and 10th quarter)	0.029	0.083
BART	Conviction (10 quarters)	0.003	0.071
Logit	Earnings (9th and 10th quarter)	0.018	0.000
Logit	Employment (9th and 10th quarter)	0.008	0.014
Logit	Conviction (10 quarters)	0.027	0.054
LASSO	Earnings (9th and 10th quarter)	0.013	0.052
LASSO	Employment (9th and 10th quarter)	0.004	0.012
LASSO	Conviction (10 quarters)	0.101	0.006
GBM	Earnings (9th and 10th quarter)	0.001	0.138
GBM	Employment (9th and 10th quarter)	0.015	0.136
GBM	Conviction (10 quarters)	0.011	0.109
BART	Earnings (9th and 10th quarter)	0.052	0.098

Source: State criminal justice data matched to NDNH and WIPS data.

Note: Caliper units are standard deviations of the propensity score on the logit scale.

BART = Bayesian additive regression trees; GBM = generalized boosted regression model; LASSO = least absolute shrinkage and selection operator; NDNH = National Directory of New Hires; RP = Reentry Project; WIPS = Workforce Integrated Performance System.

Matched sample balance

Using our primary approach of the double-selection LASSO method with a 0.2 caliper, our final matched sample consisted of 1,198 RP participants and 16,032 matched Wagner-Peyser participants. The reduction in sample size relative to the starting pool of RP participants in the analysis is a result of dropping any program group member without quality matches (in other words, there were no Wagner-Peyser participants in the same exact-match cell with a propensity score within the caliper range). Exhibit A.10

presents a summary overview of standardized mean differences across all candidate baseline covariates between the RP and final Wagner-Peyser matched comparison group.

Differences between the RP and matched Wagner-Peyser sample on observed characteristics were generally small, with an average standardized mean difference across covariates of 0.03 standard deviation units for the adult sample and 0.04 standard deviations for the young adult sample. Of the six county characteristics used for matching, RP and matched Wagner-Peyser participants had the smallest observed differences in county unemployment and poverty rate, and larger differences in crime rate per-capita, population, and urbanicity. Differences for key pre-program demographic and criminal justice background characteristics were all well below 0.25 standard deviations, the conventional benchmark for baseline equivalence used by the What Works Clearinghouse.

Exhibit A.10. Summary of sample balance, by covariate type

Measure	RP Adult				RP Youth			
	All covariates	WIPS	CJ	County	All covariates	WIPS	CJ	County
Baseline covariates tested	83	16	61	6	75	11	58	6
Average SMD across covariates	0.036	0.020	0.037	0.068	0.050	0.039	0.043	0.139
Percentage of covariates with SMD >|0.10|	6.0%	0.0%	4.9%	33.3%	8.0%	0.0%	5.2%	50.0%
Percentage of covariates with SMD >|0.25|	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	0.0%	33.3%
Percentage of SMDs with p<0.05	6.7%	6.3%	6.5%	50.0%	9.5%	0.0%	3.4%	50.0%
Employment prognostic score SMD	0.005	-	-	-	0.034	-	-	-
Earnings prognostic score SMD	0.003	-	-	-	0.028	-	-	-
Conviction prognostic score SMD	0.093	-	-	-	0.040	-	-	-

Note: WIPS covariates include individual-level demographic characteristics. County covariates include county-level characteristics including unemployment, poverty, and crime rates.

CJ= criminal justice; RP = Reentry Project; SMD = standardized mean difference; WIPS = Workforce Integrated Performance System.

C. Impact analysis

1. Impact estimation

We estimated impacts of RP program participation by comparing outcomes between RP participants and the matched Wagner-Peyser sample. To make this comparison, we used a weighted least squares regression of the following form:

$$Y_{ig} = \alpha + \beta T_i + \gamma X_{ig} + \delta_g + \varepsilon_{ig}$$

Y_{ig} is the outcome *Y* for individual *i* living in state *g*. T_i is an indicator for whether the individual received RP services. X_{ig} is a set of individual covariates, and δ_g is a state fixed effect (that is, an indicator for living in a specific state). We controlled for the same set of covariates in the impact estimation as we included in the propensity score estimation, described above. This approach produces what is called "doubly robust" impact estimates (Funk et al. 2011). For binary outcomes, we used weighted least squares estimation of the linear probability model.

We also assessed whether impacts varied across key subgroups of RP participants. For subgroup impacts, we estimated an analogous regression model, as presented above, with the inclusion of an interaction term between treatment status T and an indicator for belonging to a given subgroup S:

$$Y_{ig} = \alpha + \beta T_i + \theta T_i * S_i + \gamma X_{ig} + \delta_g + \varepsilon_{ig}$$

For all outcomes, we used weights equal to 1 for all RP participants. For the comparison group, we used matching weights, constructed as described in Section B, to adjust for the number of times a comparison group member was selected in the matching process. We estimated heteroskedasticity-robust standard errors in all regressions.

D. Sensitivity analyses

We conducted a variety of sensitivity analyses to gauge how our analytic choices might have shaped our results. Our goal was to address potential concerns that our main findings were spuriously driven by our technical approach.

- **Propensity score estimation**. Our main results stemmed from a matching approach that used a logit regression to estimate our propensity scores. We selected covariates to include in this model using LASSO. To explore how our results depended on this modeling technique, we replicated our analysis using alternative approaches: BART, GBM, and a logistic regression with researcher-specified covariates, rather than LASSO-selected covariates. All three approaches yielded similar findings as our preferred LASSO framework (Appendix Exhibit A.11).

Exhibit A.11. Impacts on confirmatory outcomes using alternative propensity score estimation approaches

Estimation approach	All RP	RP adult	RP youth
GBM			
Any new conviction during the 10 quarters after enrollment	4.6pp** (1.8pp)	3.7pp (1.9pp)	5.5pp (3.7pp)
Avg employment in the 9th and 10th quarters after enrollment	-5.7pp*** (2.2pp)	-3.1pp (2.4pp)	-11.2pp*** (3.9pp)
Avg earnings in the 9th and 10th quarters after enrollment	-$723*** ($183)	-$574*** ($218)	-$1,020*** ($326)
RP sample size	745	473	272
WP sample size	10,639	10,152	487
BART			
Any new conviction during 10 quarters after enrollment	6.0pp*** (1.7pp)	4.5pp** (1.8pp)	9.5pp*** (3.3pp)
Avg employment in the 9th and 10th quarters after enrollment	-6.6pp*** (2.1pp)	-2.1pp (2.3pp)	-15.9pp*** (3.8pp)
Avg earnings in the 9th and 10th quarters after enrollment	-$858*** ($173)	-$503*** ($194)	-$1,518*** ($296)
RP sample size	841	536	305
WP sample size	10,364	9,883	481

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates.

** *p*-value < 0.05

*** *p*-value < 0.01

BART = Bayesian additive regression trees; GBM = generalized boosted regression model; pp = percentage points; RP = Reentry Project; WP = Wagner-Peyser.

- **Selection of matched comparison group members**. We relied on a caliper match to identify similar comparison group members for each of our sampled program participants. This caliper—set at 0.2 standard deviations of the propensity score distribution—could influence our results by tolerating matches of relatively dissimilar observations. In the Section E, we show that when we use a more conservative caliper (of 0.10 standard deviations of the propensity score distribution), we recover similar impact estimates (Appendix Exhibit A.12). Likewise, when we switch from a caliper-based matching approach to a "nearest neighbor" framework—in which we match every program group member with exactly one comparison group member who has the closest propensity score—our estimates remained qualitatively similar (Appendix Exhibit A.13).

Exhibit A.12. Impacts on confirmatory outcomes using different caliper widths

Caliper width and outcome	All RP	RP adult	RP youth
Primary approach (Caliper=0.2)			
Any new conviction during the 10 quarters after enrollment	5.1pp*** (1.5pp)	4.4pp*** (1.7pp)	5.1pp (2.6pp)
Avg employment in the 9th and 10th quarters after enrollment	-4.1pp** (1.7pp)	-1.7pp (2.1pp)	-7.3pp*** (2.8pp)
Avg earnings in the 9th and 10th quarters after enrollment	-$693*** ($144)	-$403** ($189)	-$1,107*** ($227)
RP sample size	1,198	664	534
WP sample size	16,032	14,718	1,314
Alternate caliper (Caliper =0.1)			
Any new conviction during the 10 quarters after enrollment	5.3pp*** (1.6pp)	4.3pp** (1.9pp)	5.0pp (3.0pp)
Avg employment in the 9th and 10th quarters after enrollment	-3.0pp (1.8pp)	-2.4pp (2.2pp)	-3.9pp (3.2pp)
Avg earnings in the 9th and 10th quarters after enrollment	-$646*** ($158)	-$475** ($205)	-$813*** ($242)
RP sample size	1,009	577	432
WP sample size	10,870	10,048	822

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates.

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; RP = Reentry Project; WP = Wagner-Peyser.

Exhibit A.13. Impacts on confirmatory outcomes using nearest neighbor matching

Confirmatory outcome	All RP	RP adult	RP youth
Any new conviction during the 10 quarters after enrollment	5.7pp*** (1.7pp)	3.8pp (2.0pp)	8.6pp*** (2.8pp)
Avg employment in the 9th and 10th quarters after enrollment	-5.0pp** (1.9pp)	-1.4pp (2.4pp)	-10.1pp*** (3.1pp)
Avg earnings in the 9th and 10th quarters after enrollment	-$589*** ($175)	-$183 ($228)	-$1,169*** ($254)
RP sample size	1,410	805	605
WP sample size	1,198	749	449

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates.

** *p*-value < 0.05*** *p*-value < 0.01

- **Estimating impacts on binary outcomes.** By default, we used ordinary least squares (OLS) and its analogue, linear probability models (LPM), to estimate impacts. We prefer LPM to the best alternative—logistic regressions—because the resulting estimates are easy to interpret and typically similar to those that LPMs produce. To see how our estimates depended on our choice of LPM over logistic regressions, we re-estimated our models using logistic regressions. We found quantitatively similar effects using this logistic mode (Appendix Exhibit A.14).

Exhibit A.14. Impacts on conviction 10 quarters after exit using a logit model

Outcome	All RP	RP adult	RP youth
Any new conviction during the 10 quarters after enrollment	5.3pp*** (1.5pp)	5.3pp (2.8pp)	4.6pp*** (1.7pp)
RP sample size	1,192	521	663
WP sample size	16,022	1,297	14,717

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates.

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; RP = Reentry Project; WP = Wagner-Peyser.

- **Focusing on Wagner-Peyser participants who received light-touch services**. Our primary comparison group includes all Wagner-Peyser participants who faced criminal charges before the program and have a similar propensity score (within our specified caliper) to at least one RP participant. However, Wagner-Peyser includes a variety of services, some more intensive and similar to RP services than others. That is, we might report less-favorable impacts of RP because we include Wagner-Peyser participants who received similar or even more intensive supports. To amplify the contrast between our program and comparison group, we restricted our attention to Wagner-Peyser participants who received only light-touch services. As shown in Appendix Exhibit A.15, results were qualitatively similar to our main analysis findings. However, differences in average employment in the 9th and 10th quarters post-enrollment were larger when comparing RP participants to Wagner-Peyser participants who received light-touch services relative to the full sample.

Exhibit A.15. Impacts relative to Wagner-Peyser participants who received light touch services

Outcome	All RP	RP adult	RP youth
Any new conviction during the 10 quarters after enrollment	4.9pp** (2.2pp)	6.0pp (4.2pp)	4.8pp** (2.4pp)
Avg employment in the 9th and 10th quarters after enrollment	-8.1pp*** (2.5pp)	-11.9pp*** (4.3pp)	-5.8pp (3.2pp)
Avg earnings in the 9th and 10th quarters after enrollment	-$912*** ($207)	-$1,201*** ($301)	-$743** ($296)
RP sample size	643	281	362
WP sample size	4,508	338	4,170

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Note: Standard errors appear in parentheses below impact estimates.

** *p*-value < 0.05

*** *p*-value < 0.01

pp = percentage points; RP = Reentry Project; WP = Wagner-Peyser.

E. Supplemental tables

Below we present supplemental tables to accompany the final report.

1. Exploratory outcomes

Exhibit A.16. Impact of RP on exploratory outcomes

Outcome	Comparison mean	Impact estimates		
		Full sample	Adults	Young adults
Exploratory outcomes over the 10 quarters after enrollment				
Any new arrest over the10 quarters after enrollment	27%	4.8pp*** (1.5pp)	0.8pp (1.8pp)	8.5pp*** (2.7pp)
Any new incarceration spell over the 10 quarters after enrollment	5%	0.1pp (0.8pp)	-0.3pp (1.0pp)	0.3pp (1.3pp)
Any felony charge over the 10 quarters after enrollment	17%	4.0pp** (1.6pp)	1.9pp (1.9pp)	6.8pp** (2.8pp)
More than one criminal charge over the 10 quarters after enrollment	7%	0.9pp (0.9pp)	-0.1pp (1.1pp)	2.2pp (1.6pp)
Exploratory outcomes over the 4th and 5th quarters after enrollment				
Any new conviction over the 5 quarters after enrollment	11%	4.1pp*** (1.2pp)	4.7pp*** (1.4pp)	5.0pp** (2.0pp)
Any new arrest over the 5 quarters after enrollment	18%	3.1pp** (1.3pp)	-0.4pp (1.5pp)	6.2pp** (2.5pp)
Any new incarceration over the 5 quarters after enrollment	3%	0.2pp (0.6pp)	0.0pp (0.7pp)	0.1pp (0.1pp)
Avg employment in the 4th and 5th quarters after enrollment	44%	1.1pp (2.3pp)	3.0pp (2.9pp)	-2.0pp (3.8pp)
Avg earnings in the 4th and 5th quarters after enrollment	$2,274	-$361** ($163)	-$90 ($228)	-$702*** ($244)
RP sample size	N/A	1,198	664	534
WP sample size	16,032	16,032	14,718	1,314

Source: NDNH data and state administrative court records matched to WIPS data. Sample includes data from 2018–2023.

Notes: Standard errors appear in parentheses below impact estimates. Employment is defined as having any earnings in a given quarter. Sample sizes presented are for conviction outcomes measured over 10 quarters after enrollment.

** p-value < 0.05

*** p-value < 0.01

N/A = not available; pp = percentage points; RP = Reentry Project; WP = Wagner-Peyser.