Department of Labor Evaluation Design Pre-Specification Plans

Background

The Department of Labor's Chief Evaluation Office is committed to upholding the department's Evaluation Policy principles of rigor, relevance, transparency, independence and ethics in independent evaluations. For all rigorous experimental studies and studies using methods described as quasi-experimental, CEO will publish Evaluation Design Pre-Specification Plans during the planning stages of evaluations to promote transparency, and replicability. It is important to note that changes may occur during the course of conducting research after the publication of Design Plans, and final evaluation products will clearly note where and why research altered from published plans.

This document provides a template that evaluators must use to meet the pre-specification practices articulated in OMB Memo M-20-12 *Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices*. OMB Memo M-20-12 calls for making an "evaluation's design and methods available before the evaluation is conducted and in sufficient detail to achieve rigor, transparency, and credibility by reducing risks associated with the adoption of inappropriate methods or selective reporting of findings, and instead promoting accountability for reporting methods and findings." The information reported must also provide sufficient information that final reporting could be assessed per the DOL Clearinghouse for Labor Evaluation and Research (CLEAR) evidence guidelines. Evaluators may also find it helpful to refer to their Office of Management and Budget's Paperwork Reduction Act (PRA) Information Collection Request requirements submissions.

Document Control

Table 1. Document Information

Title:	DOL Evaluation Design Pre-Specification Plan: America's Promise Job Driven Grant Program Evaluation
Evaluator	Mathematica (prime)/Social Policy Research Associates
Security Level:	Public; no access restrictions.
Contact Info:	chiefevaluationoffice@dol.gov

Table 2. Document History

Version	Date	Summary of Change
1		Initial version published.
2	8/2/2021	Reordering of items and addition of *Item 10 – References*.

Narrative

Instructions: Compile a narrative responding to each of the prompts in the items that follow. In each response, provide sufficient detail such that readers can determine the study's standards for rigor and independence, assess the credibility and objectivity of the findings, and replicate/reproduce the work.

Item 1 – Purpose, Research Questions and Hypotheses. *Briefly describe objective of the evaluation (its relevance). Include primary and secondary questions and hypotheses to be tested, including ancillary or exploratory questions.*

A skill gap between the qualifications of American workers and the needs of many American businesses continues to persist. U.S. firms annually sponsor hundreds of thousands of nonimmigrant H-1B visas to hire foreign workers into skilled positions (U.S. Department of State 2021). To reclaim some of these jobs for the American workforce, in January 2017 the Division of Strategic Investments within the Employment and Training Administration of the U.S. Department of Labor (DOL) awarded more than $110 million to 23 grantees for the America's Promise Job Driven Training Grants program. These four-year grants aim to support regional partnerships (involving workforce agencies, employers, industry representatives, training providers, community-based organizations, and economic development agencies) to identify the needs of specific industry sectors that rely on the H-1B visa program and implement career pathway programs that build the skills of the domestic workforce for middle- and high-skilled jobs in those sectors. At the same time, DOL's Chief Evaluation Office contracted with Mathematica and its partner, Social Policy Research Associates, to conduct rigorous implementation, outcomes, and impact studies of America's Promise.

The impact study will estimate the extent to which America's Promise affected participants' employment and earnings. We will also assess how outcomes vary by participant and partnership characteristics to understand who benefits most from the programs and identify implementation characteristics associated with improved outcomes.

To estimate the impact of participation in an America's Promise program, we will compare the outcomes of America's Promise participants to a group of job seekers with similar characteristics who received basic employment services funded by the Wagner-Peyser program. The Wagner-Peyser program, initially authorized by the Wagner-Peyser Act of 1933, seeks to "improve the functioning of the nation's labor market by linking job seekers with employers seeking workers. Wagner-Peyser services are provided nationwide through American Job Centers (DOL n.d.). Like America's Promise participants, Wagner-Peyser participants are individuals seeking support to find new employment or opportunities for advancement in the workforce, but in contrast to America's Promise participants who are eligible for intensive sector-based training programs,

Wagner-Peyser participants typically receive light-touch employment services such as job search assistance or job referrals.

For the impact study, we will analyze the research questions listed in Exhibit 1 using a sample of study participants from across the 12 America's Promise partnerships for which we have participant data (see Item 3 for additional details).

Exhibit 1. Research questions for cross-site impact study

Cross-site impact study	
Confirmatory research questions	
Compared with receipt of Wagner-Peyser services, what was the impact of participation in a program at one of 12 America's Promise partnerships on the following?	
C.1a	Employment rate in the fourth quarter after program enrollment
C.1b	Earnings in the second year (quarters 5–8) after program enrollment
C.1c	Employment rate in the eighth quarter after program enrollment
Exploratory research questions	
Compared with receipt of Wagner-Peyser services, what was the impact of participation in a program at one of 12 America's Promise partnerships on the following?	
C.2a	Quarterly employment rate and earnings for each of the eight quarters after program enrollment
C.2b	The rate at which individuals worked in a single job providing earnings greater than 200 percent of the federal poverty rate (for an individual) in the eighth quarter after program enrollment
C.2c	The rate at which individuals attained earnings equal to or greater than their earnings in the third quarter before program enrollment
C.2d	Total earnings in the two years following program enrollment
C.2e	The total number of jobs individuals worked in the two years following program enrollment and the number of jobs individuals worked in the eighth quarter after program enrollment
C.2f	Unemployment Insurance received in the two years following program enrollment
How did impacts on employment rates and earnings, estimated relative to those of Wagner-Peyser recipients, differ by the following subgroups?	
C.3a	Participants enrolled in America's Promise training programs targeting different industries
C.3b	Enrollment status (currently enrolled, previously enrolled, or not yet enrolled) when the COVID-19 pandemic began to affect the United States
C.3c	Participant's gender; race and ethnicity; and designation as unemployed, underemployed, or an incumbent worker as measured at program enrollment

To assess the impact of participation in programs provided by specific America's Promise partnerships, we will also estimate partnership-specific impacts. We will analyze the research questions in Exhibit 2.

Exhibit 2. Research questions for partnership-specific impact study

Partnership-specific impact study	
Confirmatory research question for each partnership	
P.1	What is the mean estimated impact of participation in a program at each partnership on earnings in the second year (quarters 5–8) following program enrollment, after using Bayesian adjustments to account for the impacts of participation in the other America's Promise programs?

Partnership-specific impact study	
Exploratory research questions for each partnership	
Compared with receipt of Wagner-Peyser services and given the impacts of participation in the other America's Promise programs in the impact study, what is the probability that participation in the individual America's Promise program improved the following outcomes?	
P.2a	Employment in the fourth quarter following program enrollment
P.2b	Earnings in the fourth quarter following program enrollment
P.2c	Employment in the eighth quarter following program enrollment
P.2d	Earnings in the second year following program enrollment
P.2e	Earnings in the two years following program enrollment
P.2f	The rate at which individuals worked in a single job providing earnings greater than 200 percent of the federal poverty rate (for an individual) in the eighth quarter after program enrollment
P.2g	The rate at which individuals attained earnings equal to or greater than their earnings in the third quarter before program enrollment
Compared with receipt of Wagner-Peyser services and given the impacts of participation in the other America's Promise programs in the impact study, what is the probability that participation in the individual America's Promise program had the following impacts?	
P.3a	Improved employment by 5 percentage points or more in the fourth quarter following program enrollment
P.3b	Improved employment by 5 percentage points or more in the eighth quarter following program enrollment
P.3c	Improved earnings in the fourth quarter following program enrollment by $400 or more
P.3d	Improved earnings in the second year following program enrollment by $1,000 or more
P.3e	Reduced Unemployment Insurance received in the two years following program enrollment

Item 2 – Evaluation Design. Briefly describe the overall evaluation methodological approach, based on a logic model of the program or policy being evaluated. Briefly discuss the program of interest and the feasibility of the planned approach, including the process for developing credible control or comparison groups. Include any anticipated challenges that could result in changes in the methodological approach, and plans for how to address those challenges.

America's Promise grants

The America's Promise grant program encouraged regional partnerships to come together with a commitment—or a "promise"—to create a pipeline of trained workers to address regional labor market needs. The funding opportunity announcement (FOA) laid out the requirements associated with the grant, including eligible industries, eligible populations, required partners, and service delivery requirements (DOL 2016).

1. Eligible industries and populations

The target industries for the America's Promise grants include information technology (IT) and IT-related industries, advanced manufacturing, health care, financial services, and educational services. The grant was designed to fund education and training for "high-growth" jobs within each of these industries. According to the FOA, high-growth jobs are those that were projected to (1) add new jobs to the economy, (2) have job vacancies, (3) require workers to learn new skills

because of changes caused by technology and innovation, or (4) have an impact on the overall economy or on the growth of other industries and occupations.

The grant eligibility criteria required grantees to serve unemployed, underemployed, and incumbent workers interested in pursuing further education and training in these fields. America's Promise grantees were also encouraged to serve disadvantaged populations, which included people with low incomes, dislocated workers, underrepresented groups in the target industry (for example, women, racial minority groups), and those with barriers to employment.

2. Required partners

As detailed in the FOA, eligible America's Promise grantees included workforce development organizations, education and training providers, economic development agencies, or industry groups. The required regional partners were employer and industry representatives, workforce investment systems, economic development agencies, and education and training providers, including community and technical colleges as well as community-based organizations that offer job training. To ensure employers in the target industries were adequately involved, grantees were required to partner with a minimum of five employers or industry groups that represented at least five employers.

3. Grantee funding and service delivery requirements

In addition to funding partnership activities, America's Promise grant funds were intended to cover the costs of education and training in the target industries and occupations, including participants' tuitions and program fees. Within their regional partnerships, America's Promise grantees had to implement one or more of the following strategies: short-term or accelerated training, longer-term intensive training, or upskilling incumbent workers. Within each strategy, America's Promise partnerships could fund various work-based learning and classroom training activities, such as registered apprenticeships, on-the-job-training, paid work experience, paid internships, classroom training, distance learning, and competency-based programs.

4. Period of performance and COVID-19

The America's Promise grants were awarded in January 2017 and had an initial performance period of 48 months. The COVID-19 pandemic influenced implementation in the final grant years as states and jurisdictions began implementing stay-at-home orders in March and April 2020. The pandemic's negative impact on education and training was most notable in the advanced manufacturing and health care sectors that prioritize hands-on experience during training, whereas IT grant partnerships were well positioned to shift to virtual instruction (Bellotti et al. 2021).

Evaluating America's Promise

Although some promising research evidence suggests that sector-based training and career pathways approaches may be successful, there remains much to be learned about strategies for successfully implementing regional partnerships and the effectiveness of those efforts in combination with regional training strategies. The America's Promise Job Driven Grant Program Evaluation aims to fill some of the existing research gaps related to regional workforce partnerships and sector-based strategies through a rigorous mixed-method approach that evaluates the implementation, outcomes, and impacts of such approaches.

DOL's Chief Evaluation Office awarded the America's Promise Job Driven Grant Program Evaluation to Mathematica and its partner Social Policy Research Associates. The evaluation includes rigorous implementation, outcomes, and impact studies of America's Promise. The reports will cover the 23 America's Promise partnerships used America's Promise grants, which provide sector-based employment and training services and through regional partnerships.

A conceptual framework (Exhibit 3) guided the evaluation's design, data collection, and analysis and was refined throughout the study. The framework represents the complex array of factors that influenced the design and implementation of grant activities as well grantees' ability to achieve desired outputs and outcomes. It shows how key regional stakeholders came together under America's Promise to use federal, state, and local funding sources and develop partnerships to create a sustainable customer-centered service delivery system that achieves positive participant and employer outcomes. Within the framework, the integral America's Promise activities that focused on regional partnerships and sector-based participant services are showcased in the red boxes. Regional context and participant characteristics influence the design and implementation of grant activities as well as grantees' ability to achieve desired outputs and outcomes.

Exhibit 3. America's Promise evaluation conceptual framework
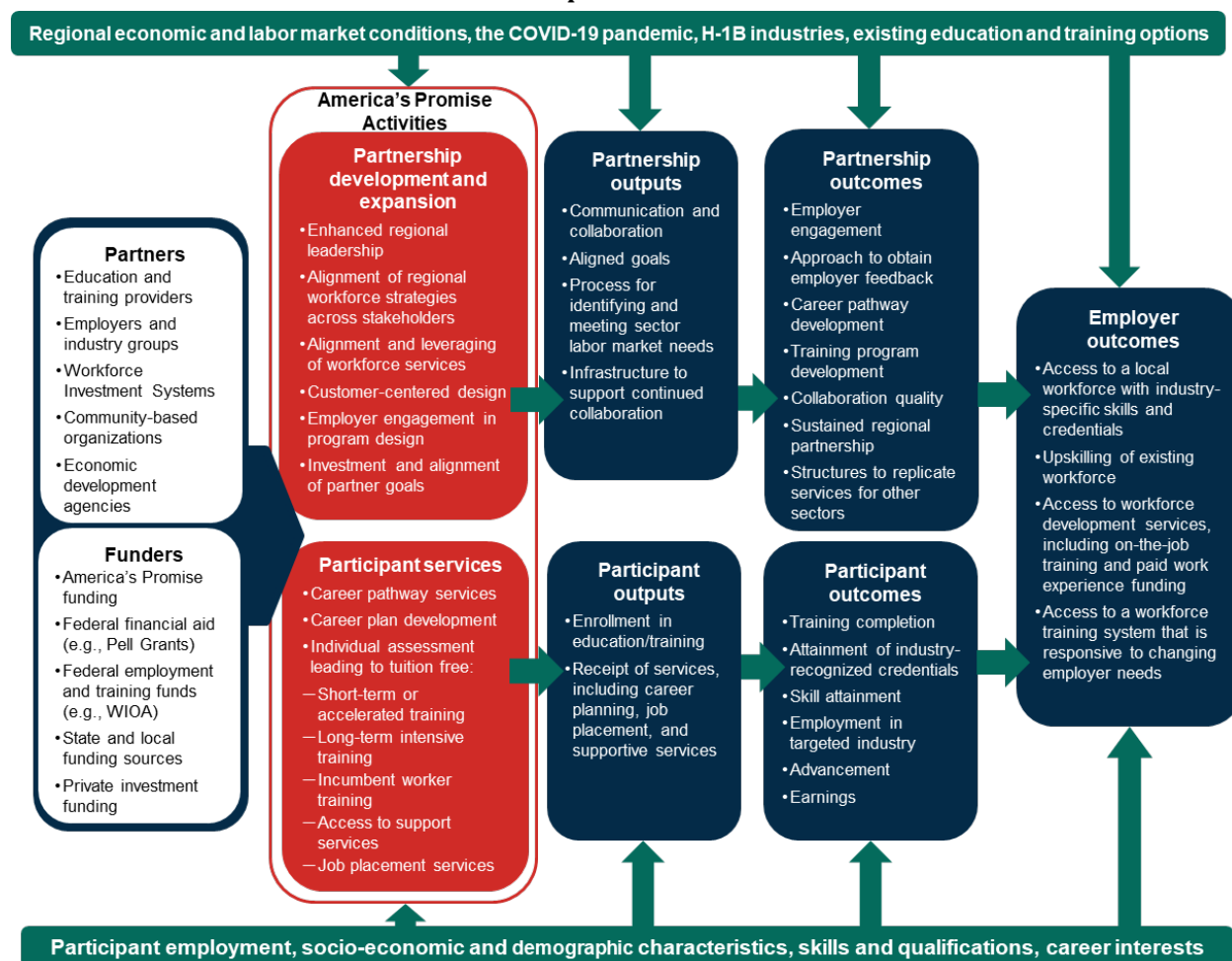
Source: Synthesis by authors based on review of relevant documents.

WIOA = Workforce Investment Opportunity Act.

Impact study design

The America's Promise impact study will include two major components: a cross-site examination of the impacts of America's Promise using data pooled across multiple partnerships and an examination of partnership-specific impacts. The impact study will include approximately 4,500[1] program year (PY) 2019 America's Promise participants from 12 partnerships and nine states. We will estimate the impact of America's Promise participation by comparing America's Promise participants to a comparison group drawn from approximately 581,000[1] Wagner-Peyser participants with similar characteristics.

---

[1] The final analytic sample size will be smaller and is dependent on results of the matching procedure and availability of employment and earnings data from the National Directory of New Hires (NDNH).

To estimate the impact of participation in an America's Promise program, we will compare the outcomes of America's Promise participants to a subset of Wagner-Peyser participants with similar characteristics and employment histories. The Workforce Integrated Performance System (WIPS) data include key demographic characteristics to use in the matching, including sex, age, education, and race and ethnicity for both America's Promise participants and Wagner-Peyser participants, but the WIPS does not include earnings data. We will obtain pre-program earnings from the National Directory of New Hires (NDNH) data under a memorandum of understanding between DOL and U.S. Department of Health and Human Services, Administration for Children and Families, Office of Child Support Enforcement (OCSE). Typically, researchers can submit Social Security numbers (SSNs) to OSCE to obtain information on employment and earnings for study participants. For this study, the WIPS data on America's Promise participants include SSNs, but the WIPS data for Wagner-Peyser participants include only workforce system identifiers. More than 1 million Wagner-Peyser participants are served each year, making it untenable to request personally identifiable information (PII) for all participants.

Therefore, we will use a two-staged matching design to identify a comparison group (Exhibit 4). We will begin by obtaining WIPS data on the full set of Wagner-Peyser participants. We will then conduct first-stage matching using participant characteristics from the WIPS data, identifying a set of potential matches. Next, we will submit the WIPS identifiers for this group to the states that agreed to participate in the impact study. These states will provide the study team with SSNs for the potential matches, which can also be matched to the WIPS data using a workforce system identifier. As the last step of the first-stage matching, we will submit the SSNs for these matches to the NDNH, from which we will obtain employment and earnings data.

Exhibit 4. Overview of quasi-experimental design
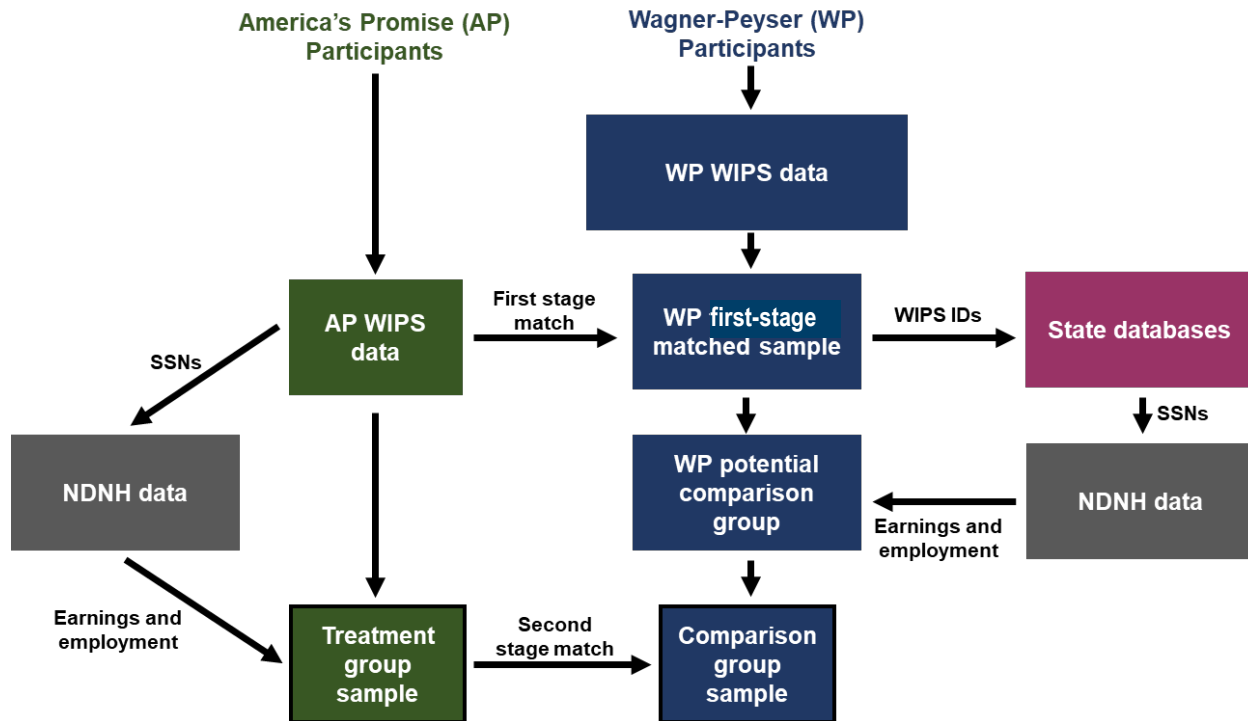
NDNH = National Directory of New Hires; SSNs = Social Security numbers; WIPS = Workforce Integrated Performance System.

Note: The first-stage matching model includes age at enrollment, gender, education level (grouped as no high school completion, high school diploma or an equivalent certificate, some postsecondary, and bachelor's degree or more), employment status at program enrollment, prior criminal justice involvement, receipt of dislocated worker services, English learner status, veteran status, disability status, and race and ethnicity (grouped as Hispanic of any race; non-Hispanic, Black; non-Hispanic, White; and other or multiracial). The second-stage matching model includes the demographic characteristics from the first-stage match, employment, earnings, receipt of Unemployment Insurance benefits in the three quarters prior to program enrollment, and local labor market characteristics.

We will use the first stage of the matching design to select the potential match pool. This group will include Wagner-Peyser participants who meet one of the following criteria: (1) the participant received services in PY 2019 in a county in an impact study partnership's coverage area or (2) the participant was matched, based on state and key demographic characteristics, to one or more America's Promise participants in an impact study state. The key demographic characteristics include 10 measures: age at enrollment, gender, education level (grouped as no high school completion, high school diploma or an equivalent certificate, some postsecondary, and bachelor's degree or more), employment status at program enrollment, prior criminal justice involvement, receipt of dislocated worker services, English learner status, veteran status, disability status, and race and ethnicity (grouped as Hispanic of any race; non-Hispanic, Black; non-Hispanic, White; and other or multiracial). At the end of the first-stage matching, the team

will have a single database containing pre-enrollment employment and earnings outcomes from the NDNH and service receipt and demographic characteristics from the WIPS, for both America's Promise participants and a subset of Wagner-Peyser participants within study states.

Second-stage matching will identify the final comparison group of Wagner-Peyser participants. These participants will be matched to America's Promise participants on demographic characteristics and pre-program employment and earnings. This sample will be selected using caliper matching on the likelihood of participation in America's Promise, which we will estimate using propensity scores (Lunt 2014). We will pool the group of America's Promise and Wagner-Peyser participants selected in the first stage of the matching procedure and estimate the probability that each individual participated in America's Promise (as opposed to Wagner-Peyser) based on observed demographic characteristics and pre-program employment and earnings information. To estimate these probabilities, called propensity scores, we will use all the demographic characteristics considered in the first-stage match; employment, earnings, and receipt of Unemployment Insurance (UI) in the three quarters prior to program enrollment; and labor market characteristics. One limitation of this approach is that the degree of similarity between the resulting matched samples is dependent on both propensity score model and the matching procedure. We will therefore consider multiple estimation approaches for the propensity scores as well as different matching strategies and determine which approach leads to the best balance in characteristics between America's Promise participants and Wagner-Peyser participants in the matched sample.

We will estimate propensity scores across partnerships using machine-learning methods designed to select the optimal comparison group based on all available data. We will use one of three machine learning methods that are designed to select predictors from a large number of covariates and their interactions (Chipman et al. 2010; Griffin et al. 2014; Belloni et al. 2014). We will first estimate propensity scores using each method and then run a "horse race" to determine which of the methods performs best in our sample. We will consider the following three methods for estimating propensity scores: (1) Bayesian additive regression trees (BART), (2) Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG), and (3) double-selection least absolute shrinkage and selection operator (LASSO). We will select our primary method for estimating the propensity score as the one that produces the lowest standardized mean difference in prognostic scores (Stuart et al. 2013). We will also assess covariate balance using t-tests for in-sample and out-of-sample covariate balance and comparing propensity score distributions using the Bhattacharyya coefficient (Bhattacharyya 1943).

Our planned primary strategy for matching based on the propensity scores is caliper matching (Lunt 2014). This strategy works by selecting all Wagner-Peyser participants within a given distance (the caliper) to form a comparison group for each America's Promise participant. Distance will be determined by individuals' propensity scores. Weights for the comparison group will be determined based on the number of America's Promise participant–specific comparison

groups for which a comparison group member is selected (Imbens and Wooldridge 2009). We will define the caliper as 10 percent or 20 percent of the standard deviation of the propensity score after transforming it to a logistic scale. To ensure this does not lead to particularly imbalanced groups, we will assess covariate balance and replace the caliper with a default of 0.1 if imbalance is a problem (Wang et al. 2013). We might choose to additionally match on a few select baseline variables to correct any remaining imbalance. We will also conduct sensitivity analyses using alternative matching strategies, including nearest-neighbor matching with replacement (Rosenbaum and Rubin 1985), matching within America's Promise service intake regions, and Bayesian causal forest (Hahn et al. 2020).

To estimate the cross-partnership impact of participation in America's Promise on employment and earnings outcomes, we will pool all participants receiving services from the 12 partnerships who have been matched with at least one Wagner-Peyser participant. We will estimate a regression controlling for individuals' demographic characteristics and employment and earnings history. We will estimate partnership-specific effects by using Bayesian analysis to bring together information on the partnership-specific estimate of the program's impact and the estimated effects of other partnerships' programs. To generate Bayesian impact estimates for each of the 10 partnerships located in one of nine study states, we will first estimate partnership-specific estimates using the methods described for the cross-partnership study. We will then use the impact estimates for the other partnerships to fit a prior distribution of treatment effects (see Item 6 for additional detail).

Item 3 – Evaluation Data. *Describe data sources, the key outcomes and primary constructs of interest (including the level of measurement, such as individual, industry, firm or geographic area), and how they will be measured, including any variables that will be examined in existing administrative datasets. Describe any demographic data points, such as age, gender, race and ethnicity, etc., that will be available, and whether they may be meaningfully analyzed based on anticipated observations (including anticipated sample size or number of observations available after linking observation units across datasets, if merging administrative or other data sources). Include information about how the collected data will be verified or verifiable, and how it will accurately capture the intended information to address the questions of interest.*

The impact study will rely primarily on two data sources: (1) the WIPS, maintained by DOL's Employment and Training Administration, and (2) the NDNH, maintained by OCSE, Administration for Children and Families, U.S. Department of Health and Human Services. WIPS data include participant background information and service receipt data. We will use the WIPS data to measure background characteristics for both the treatment and comparison groups, which will be crucial for the study's matching and for defining subgroups for analysis. NDNH data include information on employment and earnings, both before and after program enrollment. We will use the pre-enrollment NDNH data for the study's matching design and the post-enrollment NDNH data to measure the employment and earnings outcomes of interest.

Nine states where America's Promise program participants reside agreed to provide the study team with data: Florida, Kansas, Michigan, Oregon, Rhode Island, Tennessee, Virginia, Washington, and West Virginia. Our main impact analyses will include information on the individuals served by these partnerships who were residents of the states participating in our study. However, we will not estimate partnership-specific impacts for the two partnerships without a grantee in a participating state because individuals residing in a state different from that of a grantee are likely not representative of individuals served by the partnership. We anticipate including approximately 4,500 America's Promise participants and 581,000 Wagner-Peyser participants in the impact analysis sample.

Workforce Integrated Performance System (WIPS)

The WIPS is a centralized database that contains quarterly data on participants in workforce programs funded by DOL, including America's Promise and Wagner-Peyser employment services.[2] It was created in 2016 as a way to have standardized data on all programs and participants. The WIPS data contain participant characteristics, including demographic information and some prior employment data. Key demographic characteristics we will use for the impact study include age at enrollment, gender, education level, employment status at program enrollment, prior criminal justice involvement, receipt of dislocated worker services, English learner status, veteran status, disability status, and race/ethnicity.

We obtained PY 2019 WIPS data for America's Promise and Wagner-Peyser participants. These data include SSNs for America's Promise participants but not Wagner-Peyser participants. For the America's Promise sample, we submitted the SSNs to NDNH to obtain participants' employment and earnings data. The WIPS data we obtained for Wagner-Peyser participants contain unique participant identifiers but do not contain SSNs or other PII that could be used to collect NDNH data. To obtain participants' SSNs, we negotiated with states to provide the SSNs based on the WIPS identifiers for PY 2019 only. It was not feasible to collect this information for all Wagner-Peyser participants in impact study states. Therefore, we identified a subset of Wagner-Peyser participants and asked participating states to share only those SSNs. The subset selected was identified as having similar demographic characteristics to America's Promise participants using a matching algorithm, which is described in Item 2.

National Directory of New Hires (NDNH)

NDNH data from OCSE at the U.S. Department of Health and Human Services contain information on quarterly earnings and UI benefits, submitted from state UI systems and the federal government's employment records (Solomon-Fears 2011).[3] We can obtain NDNH data for America's Promise participants by SSN and for selected Wagner-Peyser participants by first

---

[2] More information about the WIPS data is available at https://www.dol.gov/agencies/eta/performance/wips.
[3] More information about the NDNH data is available at https://www.acf.hhs.gov/css/training-technical-assistance/overview-national-directory-new-hires.

name, last name, and SSN. Our impact study will be restricted to PY 2019 and states that agree to provide individuals' SSNs. We will obtain data covering the period of at least three quarters prior to program enrollment to at least eight quarters following program enrollment.

NDNH data cover most wage and salary employment, as well as receipt of UI. They also include unique employer identifiers, so tenure with an employer can be measured. However, the NDNH has limitations. The NDNH data do not cover all types of jobs and industries. In particular, NDNH data do not cover self-employed workers, railroad employees, workers in service for relatives, most agricultural labor, some domestic service workers, and part-time employees of nonprofit organizations (U.S. Departments of Labor, Commerce, Education, and Health and Human Services 2014). NDNH data also exclude workers whose employers do not report their earnings to their UI agency, even in the formal sector, because of the prevalence of flexible staffing arrangements or illegally neglecting to report (Abraham et al. 2018). Additionally, NDNH data do not cover workers who are casually employed, such as day laborers or part-time helpers, and exclude most work that is part of the gig economy (Abraham et al. 2018; Katz and Kruger 2016).

Item 4 – Response rates and attrition. *Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. Describe potential selection or response rate issues and other potential sources of bias, and resulting limitations for analyses, including limitations related to the ability to examine specific subpopulations of interest (e.g. disaggregation by gender, ethnicity, race, etc.). For collections based on sampling, a specific justification must be provided for any collection that will not yield 'reliable' data that can be generalized to the universe or population of interest.*

All of the data we will use in the impact analysis are from administrative sources. Although survey nonresponse will not be an issue for this analysis, we have the potential of missing baseline and outcome data. To maximize the number of potential matches from Wagner-Peyser in the first stage of matching, we will impute missing values to the modal value within a state for characteristics such as disability status, veteran status, and dislocated worker status. Participants missing data for age, gender, or education level will be excluded from the matching process and therefore the analyses. For the second stage of matching, some participants might be missing baseline employment and earnings data due to unsuccessful matching with NDNH. Our primary analysis will exclude these individuals from the second-stage match, proceeding with a complete case analysis. We will test the sensitivity of our main results to imputing missing baseline employment and earnings data using a multiple imputation procedure (Ling et al. 2020).

Individuals might have missing outcome data if the participant's SSN is not matched to quarterly earnings data in the NDNH. This may reflect a lack of any reported earnings or an incorrect SSN. Our initial estimates suggest that approximately 97 percent of submitted SSNs were matched to at least one quarterly earnings record in the NDNH database. We will include all individuals who were matched to at least one quarter of NDNH data in our analyses and interpret missing quarters as quarters in which the individual had zero earnings. Based on the high match

rate, the deidentified nature of the NDNH data, and our inability to distinguish between zero earnings and missing data, we will remove from the sample any individuals who did not match to any earnings data in the NDNH. We will not adjust sample weights for this loss of 3 percent of the sample.

In addition, the impact analysis will be restricted to America's Promise programs in states that were willing to share PII on Wagner-Peyser participants with the study team. We will therefore be careful to note that the results from this study do not generalize to the full set of America's Promise grantees. We will conduct a descriptive analysis of how the characteristics of the impact study states compare to the full set of grantees.

Item 5 – Sampling and Power Analyses. Describe (including a numerical estimate) the sampling frame and any sampling or other respondent selection method to be used. Describe the procedures for the collection of information including statistical methodology for stratification and sample selection; estimation procedure; degree of accuracy needed for the purpose described in the justification; unusual problems requiring specialized sampling procedures. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection. Include clear description of groups to be studied or compared and anticipated sample sizes. Also outline power calculations that align with each hypothesis to be tested to clearly demonstrate sufficient sample to examine the primary research questions with the selected methodology.

We conducted outreach to state workforce agencies to obtain PII that is needed to link the WIPS and the NDNH data. Outreach was conducted in tandem with the team from the Homeless Veterans' Reintegration Program (HVRP) study, which required similar data from states. The sample frame for the outreach included 33 states with America's Promise participants or HVRP participants. Twenty-six of these states included potential America's Promise participants (based on grantee coverage areas).

For America's Promise, grantees are required to submit an SSN for each participant; this is used to link the WIPS and NDNH data. However, SSNs are not available for members of the impact study comparison group; the WIPS data include only identifiers used within the workforce system. We therefore conducted outreach to state agencies to obtain crosswalks between comparison group members' WIPS identifiers and their names and SSNs.

The outreach process included five stages:

1. **Developing relevant materials.** The study teams developed a common set of outreach materials for communicating with states, as well as a tracking tool to identify and record the points of contact at each state.

2. **Prioritizing states.** We determined the order in which we contacted states based on the numbers of America's Promise and HVRP participants reported in grantee quarterly performance reports, as well as the number of homeless veterans in each state's WIPS data for PY 2017. We began outreach with a small set of six states in May 2019 to test our materials and approach and then continued to add states in waves through January 2020 until we reached a total of 33 states, 26 of which included potential America's Promise participants (based on grantee coverage areas).

3. **Identifying appropriate points of contact in each target state.** Identifying the correct point of contact for our request posed a substantial challenge in many states. Wherever possible, study team members with experience working with states on similar data collection efforts identified points of contact based on those experiences. In cases where there was no obvious contact person, our team conducted public records searches for technical leaders within state departments that handled workforce data. In several cases, DOL staff (including staff from the national office and one regional federal project officer) were able to provide contacts that were responsive.

4. **Conducting outreach.** After we obtained valid contact information, we sent initial emails and scheduled phone calls with points of contact. We continued to pursue states for the study until they had agreed to participate, declined to participate, or stopped responding to email requests.

5. **Negotiating and reviewing data use agreements.** For states open to considering our request, we began negotiating data use agreements using either a template developed by our team or supplied by the state. State solicitors or contracts staff reviewed the materials and often engaged in several rounds of feedback and revisions.

In total, nine states where America's Promise program participants reside agreed to provide the study team with data: Florida, Kansas, Michigan, Oregon, Rhode Island, Tennessee, Virginia, Washington, and West Virginia. The nine states included 12 partnerships—10 had a grantee located in-state and two served some individuals residing in the participating states but had a grantee in another state.

Sample and expected statistical power. We anticipate including approximately 4,500 America's Promise participants and 581,000 Wagner-Peyser participants in the impact analysis sample. To ensure that we are likely to have adequate statistical power to detect a meaningful impact, we estimated minimum detectable impacts using available data on America's Promise enrollment. In particular, we estimate that the study will be able to detect an impact of 1.6 percentage points or more for quarterly employment and $253 or more for quarterly earnings. These estimates are based on the assumption of a 79 percent employment rate among the comparison group and a $6,629 standard deviation of earnings, which were drawn from the Workforce Investment Act Gold Standard Evaluation using the pooled samples of adults and dislocated workers (Fortson et al. 2017).

Item 6 – Analyses. *Outline key models, plans for tabulation, coefficients, tables and descriptive statistics. Outline methodological approaches for regressions and other analytical methods selected by research question and hypothesis. Cite relevant literature for models used or otherwise outline the basis for the specific analytic approach. Address any complex analytical techniques that will be used. Describe how the data will be prepared and analyzed. Specify what data will be removed from final reporting due to disclosure risks. Outline dummy variables, coefficients or table cells that will be included in final public reporting (as well as those that may be removed due to disclosure risk).*

The planned analyses for the impact study will focus on comparing employment and earnings outcomes of America's Promise program participants to the matched comparison group created using the methods from Item 2. We will estimate the impact of participation in America's Promise on a range of employment and earnings outcomes, as shown in Exhibit 5. To promote tractability, we will estimate the partnership-specific impacts for a subset of these outcomes.

Exhibit 5. Impact study outcomes

	Cross-partnership analysis	Partnership-specific analysis
Confirmatory		
Employment in the fourth quarter following program enrollment	X	X[a]
Employment in the eighth quarter following program enrollment	X	X[a]
Earnings in the second year (quarters 5–8) after program enrollment	X	X
Exploratory		
Earnings: quarterly following program enrollment	X	
Employment: quarterly following program enrollment	X	
Worked in a single job providing earnings greater than 200 percent of the federal poverty rate (for an individual) in the eighth quarter after program enrollment	X	X
The rate at which individuals who were not employed at program enrollment attained earnings equal to or greater than the earnings in the third quarter before program enrollment	X	X
Total earnings in the two years following program enrollment	X	X
The total number of jobs worked in the two years following program enrollment	X	
The number of jobs worked in the eighth quarter after program enrollment	X	
Receipt of Unemployment Insurance in the two years following program enrollment	X	X

[a] Exploratory analyses for the partnership-specific impacts.

Cross-partnership estimates

To estimate the cross-partnership impact of participation in America's Promise on employment and earnings outcomes, we will pool all participants receiving services from the 12 partnerships. We will estimate a regression controlling for individuals' demographic characteristics and employment and earnings history. We will use ordinary least squares will be used for continuous

outcomes (for example, earnings, number of jobs worked) and linear probability modeling for binary outcomes (for example, employment, receipt of UI). We will use the following regression model:

$$Y_{ip} = \alpha + \beta T_{ip} + \gamma X_{ip} + \delta_p + \varepsilon_{ip}.$$

Y_{ip} is the outcomes Y for individual i at partnership-state combination p. T_{ip} is an indicator for whether the individual i at partnership-state combination p received America's Promise services. X_{ip} is a set of covariates for individual i at partnership-state combination p,[4] and δ_p is a partnership-state fixed effect (that is, an indicator for living in a specific state and receiving services from a specific partner). We will also run sensitivity analyses controlling for match-specific fixed effects. Because we include controls for characteristics in both the propensity score and regression estimation, this approach is classified as a "doubly robust" strategy. Such approaches have been found to perform well under a range of circumstances, by both Busso et al. (2014) and Huber et al. (2015)—even when there is less overlap of the propensity score distribution. Results will be presented as standard impact estimates with the associated p-values. Each America's Promise participant will be included in the analysis sample with a weight of 1, and each selected comparison member will receive a weight equal to the number of times they are selected. Weights for the comparison group will then be normalized to sum to 1 (Imbens 2015). We will correct the standard errors for the variance that is introduced from the matching procedure (Abadie and Imbens 2008), but we will not consider the variance that is introduced from estimation of the propensity score, which might lead to standard errors that are either too big or too small (Abadie and Imbens 2016).

We will additionally analyze how the impact of participation in America's Promise varied by important subgroups of participants. Specifically, we will estimate the same model as the full sample, but additionally include an interaction effect, φ_S, for the interaction between treatment indicator T_{ip} and S_{ip}, an indicator for whether individual i in partnership-state combination p belongs to subgroup S.

$$Y_{ip} = \alpha + \beta T_{ip} + \varphi_S T_{ip} * S_{ip} + \gamma X_{ip} + \delta_p + \varepsilon_{ip}$$

We will estimate impacts separately by sector of training program; gender; race and ethnicity; designation as unemployed, underemployed, or an incumbent worker at program enrollment; and timing of enrollment and participation in training relative to the onset of the COVID-19 pandemic in the United States.

---

[4] We will include all the variables used in the second-stage matching procedure as covariates in the regression analysis.

Partnership-specific estimates

Although we will have data on approximately 4,500 America's Promise participants, most individual partnerships will contribute relatively few individuals to the analysis. Seven of the 12 partnerships contribute fewer than 200 enrollees each to the impact study, with the smallest sample consisting of 37 individuals. Such small samples can lead to very noisy estimates of partnership-specific effects and can lead to incorrect conclusions about the effectiveness of specific programs. Therefore, we will estimate partnership-specific effects by using Bayesian analysis (van de Schoot et al. 2021) to bring together information on the partnership-specific estimate of the program's impact and the estimated effects of the programs of other partnerships. The statistical technique we will use to estimate the partnership-specific impacts will allow us to show our best estimate of a program's effects, along with the probability that the true effect of the program is positive or above a certain threshold. To generate Bayesian impact estimates for each of the 10 partnerships located in one of nine study states, we will first estimate partnership-specific estimates using the methods described for the cross-partnership study. We will then use the impact estimates for the other partnerships to fit a prior distribution of treatment effects. We will then use Bayesian methods to incorporate the data from the partnership and estimate the posterior distribution of the partnership-specific impact estimate.

Item 7 – Expert and stakeholder inputs. *Include a description of a process for soliciting input and feedback through peer review, technical working groups, and/or other consultation from independent, unbiased experts.*

To assess the soundness of the evaluation design and the evaluation's findings, we will convene two technical working group (TWG) meetings with experts. The project's TWG included three individuals with expertise in impact study methodology: Avi Feller, University of California, Berkeley; Peter Mueser, University of Missouri; and Elizabeth Stuart, Johns Hopkins University. The first meeting, which took place on May 20, 2021, focused on the evaluation's impact study and matched comparison design. Before the meeting, experts provided written feedback on the proposed impact estimation methods, and the virtual discussion focused on key methodological issues including (1) method of estimating propensity scores for the second-stage matching, (2) method for using the propensity scores to compare the treatment and comparison groups, (3) multiple comparison testing, and (4) formal estimation of the sensitivity of results to unobserved factors.

The second meeting, which will occur in spring 2023, will discuss the evaluation's findings. We will consult individually with TWG members if specific needs arise during the analysis.

Item 8 – Timelines, Challenges and Changes*. Indicate where, when, and how data will be collected. Include, clear timelines and plans for releasing findings to relevant stakeholders and specify how departures from the plan, including changes related to timelines and methodological decisions, will be documented. Outline potential vulnerabilities to the timeline related to data collection or access and plans to mitigate risks. Provide the time schedule for the entire project, including beginning and ending dates of the collection of information, completion of report, publication dates, and other actions.*

Exhibit 6 outlines the timeline for the impact study. We provide further details on the processes for data collection under Items 3 and 5.

Exhibit 6. Timeline for the impact study

Milestone	Timing
Project start	June 2017
Initial design report submitted to DOL	March 2021
Technical working group meeting on initial design report	May 2021
Final design report submitted to DOL	July 2021
Obtain WIPS data	April 2020 – February 2022
Obtain NDNH data	April 2020 – July 2022
Negotiate state DUAs	April 2019 – December 2020
Obtain PII from states	September 2020 – June 2021
Draft report delivered to DOL	March 2023
Final report delivered to DOL and released to stakeholders	September 2023

DOL = Department of Labor; DUAs = data use agreements; NDNH = National Directory of New Hires; PII = personally identifiable information; WIPS = Workforce Integrated Performance System.

This schedule will allow us to collect NDNH data ranging from three quarters before program entrance to eight quarters after program entrance.

Item 9 – Other relevant information. *Include any other information relevant to supporting the transparency and reproducibility of the study.*

We considered several alternative designs for this study. We first considered a randomized controlled trial (RCT) and a regression discontinuity design (RDD). Working with DOL, we determined that these approaches were infeasible, even though they are preferred for yielding unbiased impacts. An RCT was not feasible because the grantees did not anticipate sufficient oversubscription, and an RDD was not feasible because grantees were not using any eligibility criteria that aligned with the cutoffs needed for that design.

Item 10 – References. *Provide references and cite any relevant literature.*

Abadie, Alberto, and Guido W. Imbens. "Matching on the Estimated Propensity Score." *Econometrica,* vol. 84, no. 2, March 2016, pp. 781–807.

Abadie, Alberto, and Guido W. Imbens. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica,* vol. 76, no. 6, November 2008, pp. 1537–1557.

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.

Bhattacharyya, Anil. "On a Measure of Divergence Between Two Statistical Populations Defined by Their Probability Distributions." *Bulletin of the Calcutta Mathematical Society,* vol. 35, 1943, pp. 99–109.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives,* vol. 28, no. 2, 2014, pp. 29–50.

Bellotti, J., B. English, and A. Harrington. "Sector Training Strategies During the COVID-19 Pandemic: Issue Brief – Lessons from the America's Promise Partnerships." Princeton, NJ: Mathematica, September 2021.

Busso, Matias, John DiNardo, and Justin McCrary. "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators." *The Review of Economics and Statistics,* vol. 96, no. 5, December 2014, pp. 885–897.

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.

Fortson, Kenneth, Dana Rotz, Paul Burkander, Annalisa Mastri, Peter Schochet, Linda Rosenberg, Sheena McConnell, and Ronald D'Amico. "Providing Public Workforce Services to Job Seekers: 30-Month Impact Findings on the WIA Adult and Dislocated Worker Programs." Report submitted to the U.S. Department of Labor, Employment and Training Administration. Washington, DC: Mathematica Policy Research, 2017.

Griffin, Beth Ann, Greg Ridgeway, Andrew R. Morral, Lane F. Burgette, Craig Martin, Daniel Almirall, Rajeev Ramchand, Lisa H. Jaycox, and Daniel F. McCaffrey. "Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG)" Santa Monica, CA: RAND Corporation, 2014. Available at http://www.rand.org/statistics/twang.

Hahn, P. Richard, Jared S. Murray, and Carlos M. Carvalho. "Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with discussion)." *Bayesian Analysis,* vol. 15, no. 3, 2020, pp. 965–1056.

Huber, Martin, Michael Lechner, and Andreas Steinmayr. "Radius Matching on the Propensity Score with Bias Adjustment: Tuning Parameters and Finite Sample Behaviour." *Empirical Economics,* vol. 49, no. 1, 2015, pp. 1–31.

Imbens, Guido W. "Matching Methods in Practice: Three Examples." *Journal of Human Resources,* vol. 50, no. 2, 2015, pp. 373–419.

Imbens, Guido W., and Jeffrey M. Wooldridge. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature,* vol. 47, no. 1, 2009, pp. 5–86.

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.

Ling, Albee, Maria Montez-Rath, Maya Mathur, Kris Kapphahn, and Manisha, Desai. "How to Apply Multiple Imputation in Propensity Score Matching with Partially Observed Confounders: A Simulation Study and Practical Recommendations." *Journal of Modern Applied Statistical Methods,* vol. 19, no. 1, 2020.

Lunt, Mark. "Selecting an Appropriate Caliper Can Be Essential for Achieving Good Balance with Propensity Score Matching." *American Journal of Epidemiology,* vol. 179, no. 2, 2014, pp. 226–235.

Rosenbaum, Paul R., and Donald B. Rubin. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician,* vol. 39, no. 1, 1985, pp. 33–38.

Solomon-Fears, Carmen. "The National Directory of New Hires [electronic version]." Washington, DC: Congressional Research Service, 2011.

Stuart, Elizabeth., Brian Lee, and Finbarr Leacy. "Prognostic Score–Based Balance Measures for Propensity Score Methods in Comparative Effectiveness Research." *Journal of Clinical Epidemiology,* vol. 66, no. 8, 2013, pp. 84–90.

U.S. Department of Labor. "Notice of Availability of Funds and Funding Opportunity Announcement for: America's Promise Job Driven Grant Program." August 2016. Available at https://www.doleta.gov/Grants/pdf/FOA-ETA-16-12.pdf.

U.S. Department of State. "Worldwide H1B, H2A, and H2B Visa Issuances (FY 2013–2018)." 2021. Available at https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/graphs.html.

U.S. Departments of Labor, Commerce, Education, and Health and Human Services. "What Works in Job Training: A Synthesis of the Evidence." 2014. Available at https://www.dol.gov/sites/dolgov/files/OASP/legacy/files/jdt.pdf.

U.S. Department of Labor, Employment and Training Administration. "Wagner-Peyser Act Employment Service Results." Available at https://www.dol.gov/agencies/eta/performance/results/wagner-peyser

van de Schoot, Rens, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G. Tadesse, Marina Vannucci, et al. "Bayesian Statistics and Modelling." *Nature Reviews Methods Primers,* vol. 1, no. 1, 2021, pp. 1–26.

Wang, Yongji, Hongwei Cai, Chanjuan Li, Zhiwei Jiang, Ling Wang, Jiugang Song, and Jielai Xia. "Optimal Caliper Width for Propensity Score Matching of Three Treatment Groups: A Monte Carlo Study." *PloS One,* vol. 8, no. 12, 2013, e81045.