# 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: Impact Study of the Homeless Veterans' Reintegration Program Evaluation |
|---|---|
| Evaluator: | Mathematica (prime)/Urban Institute/Social Policy Research Associates |
| Security Level: | Public; no access restrictions. |
| Contact Info: | [chiefevaluationoffice@dol.gov](mailto: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.*

In 2018, nearly 38,000 veterans were homeless, which is defined as sleeping outside, residing in an emergency shelter, or living in a transitional housing program (U.S. Department of Housing and Urban Development 2018). Veterans face complex challenges—both unique to their military service and common to the broader population—that put them at risk for becoming homeless. Most veterans (69 percent) report that their greatest challenge is finding a job (Iraq and Afghanistan Veterans of America 2012). Because it can be difficult to find a new job, veterans making the transition from active duty to civilian life might be unable to afford housing. In addition, service-related trauma can lead to having a disability, mental health challenges, or substance abuse issues, which are risk factors for homelessness among the broader population (Balshem et al. 2011; Tsai and Rosenheck 2015). Homelessness and its associated outcomes, including lower family and housing stability and employment, poorer health, and more interactions with the criminal justice system (Caton et al. 2007; Culhane and Byrne 2010), are deep public concerns, especially given veterans' service to our country.

The U.S. Department of Labor's (DOL) Homeless Veterans' Reintegration Program (HVRP) seeks to help veterans experiencing homelessness find and retain meaningful employment. HVRP, initially authorized by the Stewart B. McKinney Homeless Assistance Act in 1987, has been providing one-year grants to local organizations to provide employment services and develop effective service systems. In June 2019, DOL's Veterans' Employment and Training Service (VETS) awarded 149 program year (PY) 2019 grants—composed of 51 new grants and 98 option year grants—to support more than 18,000 veterans. The following year, VETS awarded 157 PY 2020 grants, including 77 new grants and 80 option year grants, to support more than 21,000 veterans.

The DOL Chief Evaluation Office, in collaboration with VETS, contracted with Mathematica and its subcontractors, the Urban Institute and Social Policy Research Associates, to complete a mixed-methods evaluation of HVRP. The goal of this evaluation is to estimate how effective HVRP is in helping veterans experiencing homelessness find stable employment. The evaluation contains two studies: (1) the impact study, using a causal non-experimental design, and (2) the complementary implementation study, to help interpret the impact study findings. **This document describes the design of the impact study only.**

The primary research questions for the impact study are:

- Are HVRP participants in Program Years (PYs) 2019 and 2020 more likely to be employed two years after enrollment than similar non-participant veterans experiencing homelessness who received Wagner-Peyser services?

- Do HVRP participants in PYs 2019 and 2020 have higher earnings two years after enrollment than similar non-participant veterans experiencing homelessness who received Wagner-Peyser services?

Exploratory research questions include:

- Do HVRP participants in PYs 2019 and 2020 have better employment outcomes—that is, quarterly employment rates, quarterly earnings, and job tenure—over the entire two-year period after enrollment than similar non-participant veterans experiencing homelessness who received Wagner-Peyser services?

- Do the effects of HVRP on participants in PYs 2019 and 2020 differ for key subgroups defined by participant and local area characteristics?

- Which key HVRP features are most associated with beneficial program impacts?

***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.*

The conceptual framework of HVRP underpins the evaluation's integrated design, data collection, and analysis (Figure 2.1). Four sets of factors can affect a grantee's ability to improve the outcomes of veterans experiencing homelessness: (1) the grantee and its program model, (2) the partnerships the grantee develops, (3) the type and commitment of employers engaged, and (4) external regional structures. The framework also acknowledges veterans and the circumstances they bring to the program.

**Figure 2.1. Conceptual framework linking HVRP services to improved outcomes for veterans**
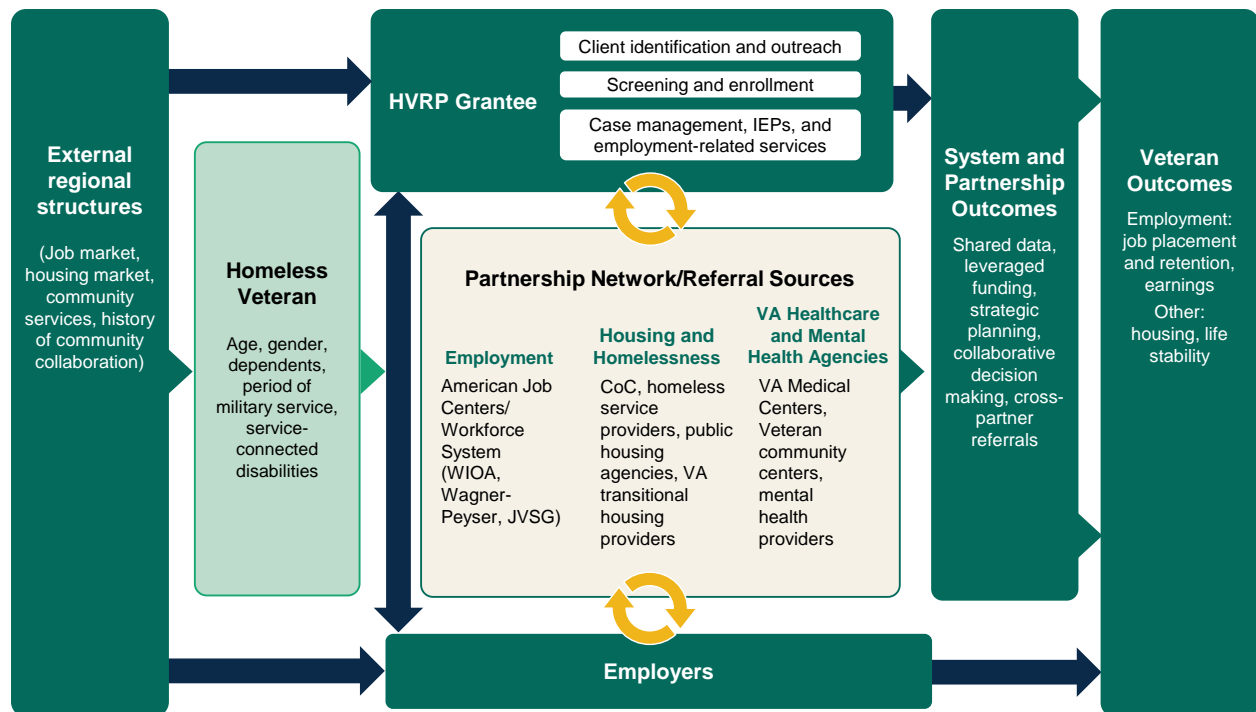


CoC = Continuum of Care; IEP = Individual Employment Plan; JVSG = Jobs for Veterans State Grant; VA = U.S. Department of Veterans Affairs; WIOA = Workforce Innovation and Opportunity Act.

Grantees aim to reintegrate veterans experiencing homelessness into the labor force by conducting outreach and recruitment; screening, assessing, and enrolling participants in the program; and providing case management services and employment, training, and supportive services tailored to participant needs. Grantees also make referrals, as needed, to partner organizations that will provide other employment services, help finding housing, and offer other critical health support services, as well as use relationships they have with employers for job placements.

A fully functioning service system is required to meet all the needs of veterans experiencing homelessness and make their reintegration easier. Such a system can help them find housing; receive medical and mental health treatment; find transportation to work, health care providers, and recreation facilities; and obtain meaningful employment. An HVRP grantee's success in building this system depends on its partnerships with service providers, including:

- **Employment-related partners.** The public workforce system through American Job Centers (AJCs) provides job-related services, including training, job search assistance, and Job Club workshops. In PY 2016, VETS instructed HVRP grantees to co-enroll their participants into a program funded by the Workforce Innovation and Opportunity Act, Wagner-Peyser Act Employment Services, or Jobs for Veterans State Grant at a local AJC (Veterans' Program Letter, no. 03-16).

- **Homelessness and housing providers.** Partners, such as a U.S. Department of Housing and Urban Development Continuum of Care and public and private homeless service providers, offer critical housing and homelessness services.

- **U.S. Department of Veterans Affairs (VA) health and mental health agencies.** VA medical centers, veteran community centers, and mental health agencies provide physical health care, behavioral health, trauma, and substance use treatment services. Links to these service providers are particularly important for veterans dealing with post-traumatic stress disorder and traumatic brain injury.

Employers are both a critical partner and a customer. Grantees recruit employers as partners that are open to hiring HVRP participants. Because employers then rely on HVRP to fill vacant positions, grantees must ensure that employers are satisfied with the placed veterans. Employers that become dissatisfied will be less open to hiring HVRP participants in the future.

Regional structural forces affect the system and the veterans. These forces may include local labor market and housing conditions, as well as the availability of supportive services through the community outside the partnership network.

The conceptual framework for HVRP ends with two sets of outcomes: those related to the system and partnerships, and those related to veterans. Grantees are funded to build or participate in systems that result in strong referral networks, leveraged funding, cross-partner referrals, and collaborative decision making. For veterans, primary metrics of success are participants' job placement, job retention, and other outcomes, such as housing and life stability.

The impact study will compare employment and earnings outcomes for HVRP participants to those for veterans experiencing homeless and not participating in HVRP. The design requires critical data on demographics and earnings before the intervention, as well as employment and earnings outcomes for the follow-up period. Informed by previous research (for example, Heinrich et al. 2013), the study team will select a set of potential comparison group members from individuals receiving Wagner-Peyser employment services, a group whose data are captured in the Workforce Integrated Performance System (WIPS). The team will then combine data from the WIPS with personally identifiable information (PII) obtained from state workforce agencies to match the treatment and comparison groups with National Directory of New Hires

(NDNH) wage records data from the Office of Child Support Enforcement (OCSE) at the U.S. Department of Health and Human Services. We will then select the final comparison group using pre-program NDNH wage data along with demographic information in the WIPS data to obtain well-balanced treatment and comparison groups based on pre-program characteristics. Finally, we will use NDNH post-program employment and earnings outcomes to estimate impacts.

We will perform five key steps as part of the impact study:

1. **Obtain WIPS data set from DOL.** The WIPS data set contains information on participants in DOL workforce programs and is used for performance reporting. The data the HVRP study team will obtain contains information on both HVRP participants who co-enroll in the Wagner-Peyser program and other veterans experiencing homeless who enrolled in Wagner-Peyser, and data flags exist to identify both groups. Accordingly, the study will estimate the added effects of HVRP services relative to other reemployment services provided at AJCs. Because the WIPS data only include co-enrolled HVRP participants and not all individuals enrolled in HVRP, the data cannot be used to obtain nationally representative impact estimates of HVRP services. However, the WIPS data yield convenience samples of sufficient size to obtain precise, informative impact estimates.

   One key element in the WIPS data is an indicator for whether a veteran experiencing homelessness was enrolled in HVRP (this will determine whether the individual is part of the treatment or comparison group). However, comparisons of HVRP grantees' Quarterly Performance Reports (QPRs) and WIPS data suggest that HVRP participation might be underreported in the WIPS. We will therefore use QPR data to improve the quality of this indicator (see Item 3 in this document).

2. **Recruit states to obtain PII on the sample.** The WIPS data set does not contain PII, but it does include a unique identifier that can be matched to Social Security numbers (SSNs) and names in states' workforce data systems. We will use PII from states to obtain the key outcome data for the study: NDNH data from OCSE, which contains quarterly employment and earnings submitted from state Unemployment Insurance systems. We will also use this PII to check the accuracy of the indicator for receipt of HVRP services in the WIPS data. In particular, we will match our sample to HVRP grantees' official reports of participants to identify comparison group members who were actually HVRP participants.

   We will seek PII for the universe of veterans experiencing homelessness in the WIPS data for each participating state. Our goal is to create data use agreements (DUAs) with up to 13 states, yielding a sample of approximately 1,000 HVRP participants and 2,000 Wagner-Peyser participants.

3. **Collect NDNH data.** After securing PII for the sample from the study states, we will request NDNH data from OCSE. Due to delays in the evaluation and obtaining agreements with states, each individual included in the analysis will have employment and earnings data for one to three quarters before, and one to eight quarters after, program enrollment, with variation based on state and quarter of program enrollment (see Item 8 in this document). (As noted in Item 3, these data will not be sufficient to answer

the study's primary research questions related to employment and earnings outcomes eight quarters after program entry.) The impact study will not include follow-up surveys with HVRP and comparison group members due to cost considerations.

4. **Select comparison group.** We will select a comparison group of non-HVRP veterans experiencing homelessness who received Wagner-Peyser services in PY 2019 and PY 2020 and had similar demographic characteristics and pre-program earnings as HVRP participants.

To summarize participant similarity, we will estimate propensity scores for sample members. We will use two key methods for estimating the propensity score. Generalized boosted regression models (GBM) will serve as our benchmark approach, and least absolute shrinkage and selection operator (LASSO) linear probability regression models will additionally be examined as a sensitivity analysis (McCaffrey et al. 2005; Tibshirani 1996). We chose these two strategies because they allow for flexibility while removing researcher decisions for model creation, which is important for this study given the unknown contexts of HVRP selection across counties. We selected GBM as our primary approach based on analysis of WIPS data from 2017.

We will use three common strategies for applying the propensity scores to estimation. The first strategy, and our benchmark approach, is inverse probability weighting (IPW; Horvitz and Thompson 1952). One of the primary benefits of IPW is that it allows for inclusion of a larger sample, which can decrease sampling variance and increase efficiency (Hirano et al. 2013). However, there are cases when weighting leads to large bias, particularly when the overlap of the two research samples is poor (Busso et al. 2014). Therefore, we will also use additional strategies as sensitivity analyses, including nearest-neighbor matching with replacement and caliper matching. Nearest-neighbor matching has the benefit of being simple, and it has been shown to be associated with smaller bias, on average, across a range of workforce studies attempting to replicate randomized controlled trials (Glazerman et al. 2003). That said, it has also been shown to perform poorly on other design-quality metrics, including overall error and large estimator variance (Huber et al. 2013; Busso et al. 2014). Caliper matching works by selecting all comparison group members within a given distance as representing the comparison group for each HVRP participant. Comparisons of outcomes are then made across groups while modeling the influence of covariates for each respective group (Imbens and Wooldridge 2009; Belloni et al. 2014). Caliper matching has been found to perform well under a range of circumstances by both Busso et al. (2014) and Huber et al. (2013).

Ideally, we would estimate a different propensity score model for each local area. But local area sample sizes are likely to be insufficient for estimating propensity score models that are local-area specific. Instead, we will estimate a single propensity score model using data pooled from across all HVRP areas included in the study (controlling for county-level characteristics), by state, but conduct matching using the propensity scores both within (preferably) and across (if needed) counties. We will assess the quality of both types of matches.

To gauge the success of our approach for generating balanced HVRP and weighted comparison group members on the observable matching variables, we will examine the

distribution of propensity score estimates for the two groups and assess the balance on the baseline covariates. We will focus on whether effect sizes for differences are small (less than 0.05 standard deviations) but also present two-tailed $p$-values resulting from $t$-tests.[1]

It is possible that our benchmark model performs poorly in terms of generating balanced HVRP and weighted comparison group members on the observable matching variables. We have conducted a proof of concept using PY 2017 WIPS data to mitigate this concern. The proof of concept demonstrates that using GBM to estimate propensity scores and IPW to estimate impacts will generate treatment and comparison groups that are well balanced based on observable characteristics. However, if there are issues with balance within the PY 2019 and PY 2020 data, we will explore using alternative methods (such as LASSO or nearest-neighbor matching) in our benchmark approach.

5. **Estimate impacts.** We will estimate impacts by comparing employment outcomes for the HVRP and comparison samples over time using the NDNH data and weighting based on the propensity score. To mitigate spurious significant impact findings that can occur due to multiple hypothesis testing across many outcomes, we have pre-specified the confirmatory (primary) outcomes as employment and earnings in the eighth quarter after random assignment.

---

[1] We acknowledge that the $t$-test may not be the best measure of similarity due to statistical significance being directly tied to sample size (Imbens and Ruben 2015) but will include it for its familiarity.

***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. 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 design 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.

**Workforce Integrated Performance System.** The WIPS is a centralized database that contains quarterly data on participants in workforce programs funded by DOL, including Wagner-Peyser employment services. 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. Key demographic characteristics we will use for the impact study include age, gender, race, ethnicity, disability status, education, timing of military separation, employment status at program enrollment, and English learner status. The WIPS also includes data on employment and training services received, which we can use to understand services the comparison group receives.

We will obtain PY 2019 and PY 2020 WIPS data for all veterans experiencing homelessness and participating in the Wagner-Peyser program. The WIPS data for Wagner-Peyser participants contain unique participant identifiers but do not contain SSNs or other PII that could be used to collect NDNH data. We are negotiating with states to provide SSNs based on the WIPS identifiers for PY 2019 and PY 2020.

One key element in the WIPS data is an indicator for whether a veteran experiencing homelessness was enrolled in HVRP (this will determine whether the individual is part of the treatment or comparison group). However, comparisons of QPRs and WIPS data suggest that HVRP participation might be underreported in the WIPS. To improve the accuracy of this indicator, we are using two additional data sources for the study: (1) QPRs from grantees that have partial name, gender, and race information for all HVRP participants and (2) full names of all veterans experiencing homelessness in the WIPS from states. By creating partial matches using these two data sources on names and limited characteristics, we will be able to identify potential comparison group members who were actually HVRP participants and adjust their group assignment before implementing the design.

**National Directory of New Hires.** NDNH data from OCSE contain information on quarterly earnings and Unemployment Insurance benefits, submitted from states' systems and the federal government's employment records (Solomon-Fears 2011). NDNH data cover most wage and salary employment and unemployment receipt. These data also include unique employer identifiers, allowing us to measure tenure with an employer. Due to delays in the project schedule and recruitment of states, we will obtain data from the NDNH for one to three quarters before and one to eight quarters after program enrollment for the current evaluation project on the current evaluation timeline. Availability of data will vary by state and quarter of program enrollment, with the average participant having data available for about four quarters after program enrollment.

The outcomes for the impact study will include quarterly measures of employment and earnings, as well as measures of job tenure. We will construct employment and earnings measures by quarter, by year, and over the entire follow-up period. In addition, we will request scrambled employer IDs in the NDNH data so that we can create measures of job tenure at firms across different quarters. Given delays in the early stages of the evaluation, we do not anticipate being able to obtain data on the study's confirmatory outcomes under the current evaluation contract.

***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.*

Nonresponse will not be an issue for this analysis, as all data used are from administrative sources. However, some baseline data are missing within these administrative data sources. To include the information available in baseline data without excluding individuals from the sample, we will use imputation, conducted independently for the HVRP and comparison groups. We will use a chained stochastic regression approach where variables are imputed using information from other variables (Rubin 1987; Raghunathan et al. 2001). The chained equation method runs a series of regression models that temporarily fill in missing values of variables when predicting other ones. This updating process continues until the change to the newly predicted values are below a pre-specified stopping criterion. We will then use predictive mean matching to impute missing observations. This method works, for example, by filling in a person's missing education level by (1) identifying a group of veterans with similar predicted education values to those of the person with the missing value and (2) using the actual education level of a randomly selected person in that group as the imputed value. This method is valid under the assumption that data are missing at random, conditional on the variables included in the imputation model.

In addition, our analysis will be restricted to HVRP programs in states that are 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 HVRP 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.*

The HVRP impact study's approach relies upon obtaining PII for Wagner-Peyser program participants from states. Because all data are administrative, this is the key selection mechanism determining the HVRP participants who will be included in the study.

We reached out to states in five stages. In particular, we (1) developed relevant materials, (2) determined which states to prioritize, (3) identified appropriate points of contact in each targeted state, (4) conducted outreach to invite states to participate in the study and share data, and (5) conducted legal negotiations and reviewed data use agreements (DUAs). Because the two projects had similar data needs, we reached out jointly for the HVRP and America's Promise Job Driven Grant Program (APG) Evaluations.

1. **State outreach 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. **Prioritization of states.** We determined the order that states were contacted based on the numbers of APG and HVRP participants reported in grantee QPRs as well as the number of veterans experiencing homelessness in each state's WIPS data for PY 2017. We reached out to a small set of 6 states in May 2019 to test our materials and approach, then continued to add states in waves through January 2020 until we reached 33 states.

3. **Identification of state contacts.** Identifying the correct point of contact for our request posed a significant challenge in many states. To identify contacts, we leveraged our team's experiences, recommendations from DOL staff, and public records searches.

4. **State outreach and responses to our request.** As of the end of April 2021, 11 of the 33 states that we approached ultimately approved DUAs for data sharing;[2] 3 were still in negotiations at the time this plan was drafted.

---

[2] One of these states did not include any HVRP grantees but was included in the outreach for the APG evaluation.

5. **Legal negotiation and review of DUAs.** For states open to considering our request, we began DUA negotiations using either the template our team developed or a state-supplied DUA.

**Changes in approach and challenges encountered during outreach.** Given that this was the first time an evaluation team attempted to match WIPS data with PII from state workforce agencies, we encountered a series of challenges that required us to adapt our approach over time and that highlight the limitations to using this approach, particularly with some states. Challenges generally fell into three categories: (1) many states cited that our request was in direct conflict with state law or privacy regulations; (2) some states did not understand why we were contacting them for data because the state itself was not a grantee for either the APG program or HVRP; and (3) changes to the studies' designs and timelines required us to revise DUAs that were actively being negotiated or include an addendum to ones already executed.

**Sample and expected statistical power.** For the full sample of PY 2019 and PY 2020 enrollees, we anticipate having WIPS data for approximately 3,200 individuals across 13 states.[3] Given this sample size and our expected use of QPR data to increase the number of Wagner-Peyser participants identified as receiving HVRP services (see Item 3), we expect our sample to include around 1,000 HVRP participants. These individuals will have received services from up to 43 HVRP grantees.

Assuming a design effect of 2.66 due to IPW (based on preliminary calculations), we estimate the study will yield a minimum detectable impact of 5.9 percentage points for quarterly employment and $583 for quarterly earnings. These minimum detectable impacts may be attainable as they pertain to the value added of HVRP participation versus Wagner-Peyser services, where HVRP services provide more intensive case management. Evidence on staff-assisted/intensive services versus less intensive services from both matching studies (Heinrich et al. 2013) and an experimental evaluation (Fortson et al. 2017) shows impacts on employment as high as 10 percent and impacts on earnings as large as $1,000. Therefore, these impacts are within the range of recent impacts found in the literature.

---

[3] In addition to WIPS data, we must have data on PII from states to include observations in our analysis. Across the 10 states participating in the HVRP evaluation, we have typically received PII on about 90 percent of individuals for whom we have requested it.

***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).*

## Impact estimation methods

The parameter of interest for the impact study will be the average treatment effect on the treated (ATT). This parameter is meant to represent impacts for the population of individuals who received HVRP services. This differs from two other commonly reported parameters for evaluations: the average treatment effect (ATE) and the intent-to-treat (ITT) impact. The ATE is meant to represent the impact for the full population of potential program participants, and it is not the focus of this study because we know that our HVRP sample is already partially selected by dual enrollment in the Wagner-Peyser program. The ITT impact is often the focus of randomized controlled trials because it represents the impact for those who were offered program services, but it is not relevant for this non-experimental design because we do not systematically know who was offered the opportunity to participate.

We will use weighted least squares to estimate ATT effects, with weights determined based on the propensity score approach (see Item 2). Regressions will control for age, gender, race, ethnicity, disability status, education, timing of military separation, pre-program employment and earnings, and other characteristics selected for inclusion in the propensity score model (yielding a doubly robust estimator, see Funk et al. 2011).

All impacts will be estimated using imputed baseline data that will introduce additional variation. To account for this, we will estimate standard errors that include variance from the initial imputation stage. This will be done using the multiple imputation and pooling approach of Rubin (1987), which provides a method for pooling impact and variance estimates across imputed samples.

There are some limitations to the approaches suggested here that we cannot address due to data limitations. Specifically, the limited data on employment outcomes prevent us from implementing some of the best practices when applying these techniques. For example, our data on employment history do not go back far enough to test the selection-on-observables assumption as suggested by Imbens (2015), where null impacts are tested on a time period before the intervention. For similar reasons, we are unable to implement individual-level difference-in-differences estimators that have been effective at removing additional bias (Heckman and Smith 1999; Mueser et al. 2007). Instead, this study will rely on the baseline data available in WIPS and exhaustive interactions of past employment on a shorter time horizon to compensate for the shorter panel.

**Estimates for subgroups**

The purpose of the exploratory subgroup analysis is to determine what works and for whom. We will analyze impacts for two types of subgroups: (1) those defined by baseline demographic and local area characteristics and (2) those defined by grantee characteristics (when possible).

**Subgroups defined by veteran and local area characteristics.** The first subgroup analysis will examine the extent to which HVRP impacts vary based on the baseline characteristics of veterans experiencing homelessness and their local areas. We will focus on five sets of subgroups based on age, gender, education level at enrollment, employment status before enrollment, and the local unemployment rate.

We will estimate impacts for veteran subgroups by modifying regression models to include terms formed by interacting subgroup indicators with the treatment status indicator and using *F*-tests to assess whether differences in impacts across subgroup levels are statistically significant. For example, to assess whether impacts are larger for male veterans than female veterans, we will construct an indicator variable that equals 1 for male veterans and 0 for female veterans. We will then interact this indicator with the treatment status indicator and include it as a covariate in the regression models.

A challenge with estimating subgroups one by one is that there could be considerable overlap across the subgroup categories, making the subgroup impact estimates difficult to interpret. To address this issue, we will explore using cluster analytic methods to create "typologies" of veterans experiencing homelessness that will group sample members with similar characteristics into the same "cluster." If this data reduction approach yields a manageable number of typologies (groups of sample members) that are policy relevant and interpretable, it is an efficient way to capture the range of independent subgroup effects.

**Subgroups defined by grantee characteristics.** An important study goal is to identify key HVRP features that are associated with more beneficial program impacts (to move beyond a "thumbs-up/thumbs-down" analysis). To achieve this goal, in the second set of subgroup analyses, we will attempt to examine impacts for subgroups defined by key HVRP features constructed using data from the grantee survey that will be administered to PY 2020 grantees as part of the implementation study. We will pursue this analysis only if we determine that we can reasonably map both PY 2019 and PY 2020 study participants to grantees that are still operating in PY 2020. Information on the grantee from which HVRP participants received services is available in the WIPS data, but the quality of the field is still unknown.

If possible, we will examine impacts for the following subgroups:

- **Partnerships.** HVRP grantees rely on partnerships with entities in their communities to ensure that their participants receive the services they need to be successful at work. We will use grantees' survey responses to develop a measure of partnership strength to determine whether grantees with strong partnerships are associated with larger HVRP effects than grantees with weak partnerships. Categorizations of strength can be determined after these data are collected and could relate to items such as co-enrollment with other employment or housing programs.

- **Case management.** Another key program element is case management. HVRP case managers can help participants access and navigate HVRP and partners' supports and services to help them overcome barriers that make work difficult. We will explore ways to develop a case management measure based on survey elements to determine whether more extensive case management (as defined by the number of meetings and caseloads) results in larger effects.

- **Relationship with employers.** Employers are a special type of partner, as HVRP grantees partner with employers and reach out to employers to develop job opportunities for their participants. Several items in the grantee survey will help us construct a measure of how grantees reach out to and partner with employers to determine whether special relationships with employers result in better outcomes for participants.

The complexity of the constructed measures for this subgroup analysis will depend on the extent of variation across grantees in terms of their service models and structure as well as data quality. The constructed measures could be binary categorizations (for example, whether intensive case management was offered) or more complex scales (for example, the intensity of case management along a continuum).

Impacts for subgroups defined by grantee characteristics will be estimated using methods similar to that for the subgroups defined by veteran characteristics. In addition, we will estimate multi-level models (using hierarchical linear modeling) that control for all grantee features at the same time to isolate the association between a particular feature and program impacts, holding constant the effects of other grantee features (including grantee-level measures of counterfactual service offerings).

***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 experts invited to the first meeting, which took place on July 12, 2019, included those on evaluation design and HVRP. This meeting allowed the experts to provide input on both the impact study's matched comparison group design using administrative data and the evaluation's implementation study. The second meeting, which will occur toward the end of the project, will discuss the evaluation's findings. In addition, over the course of the study, we may seek individual TWG members' input on additional design issues later in the project, such as on the comparison group created as part of the matched comparison design.

*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.*

The timeline for the impact study is outlined in Table 8.1. We provide further details on the processes for data collection under Items 3 and 5.

**Table 8.1. Timeline for HVRP impact study**

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

This schedule will enable us to collect data on participant outcomes for an average of four quarters following program enrollment. We will explore the possibility of conducting additional research to capture results over a longer follow-up period.

This schedule reflects several changes to the HVRP impact study design, based on changes in data availability over the course of the contract. The project's final report will document any further changes to our timelines or methods.

We do not anticipate any further issues in data collection, because almost all DUAs and processes are in place. Negotiations with three states are still ongoing, but lacking the data from these states will have a small effect on our sample size and not materially affect evaluation procedures or processes. If individual states are no longer willing to share data as previously agreed upon, the sample size for analysis will decrease but the overall project timeline should not be affected.

***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 experimental and regression discontinuity designs. Working with DOL, we determined that these approaches were infeasible, even though they are preferred for yielding unbiased impacts. We then designed a non-experimental study similar to that outlined here but using pre-program earnings data from the WIPS. That study design is summarized in detail in an internal design report submitted to DOL. In March 2020, the study team learned that we could not use pre-program earnings data from the WIPS. We therefore redesigned the study to use pre-program earnings data from the NDNH for matching.

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

Balshem, Howard, Vivian Christensen, and Anais Tuepker. "A Critical Review of the Literature Regarding Homelessness Among Veterans." Washington, DC: Department of Veterans Affairs, 2011. Available at https://www.hsrd.research.va.gov/publications/esp/homelessness.pdf. Accessed July 20, 2017.

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.

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, 2014, pp. 885–897.

Caton, C., C. Wilkins, and J. Anderson. "People Who Experience Long-Term Homelessness: Characteristics and Interventions." Paper presented at Toward Understanding Homelessness: The 2007 National Symposium on Homelessness Research, U.S. Department of Health and Human Services and U.S. Department of Housing and Urban Development, Washington, DC, 2007.

Culhane, Dennis, and Thomas Byrne. "Ending Chronic Homelessness: Cost-Effective Opportunities for Interagency Collaboration." Philadelphia, PA: University of Pennsylvania, March 2010.

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

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.

Glazerman, Steven, Dan M. Levy, and David Myers. "Non-Experimental Versus Experimental Estimates of Earnings Impacts." *The Annals of the American Academy of Political and Social Science*, vol. 589, no. 1, 2003, pp. 63–93.

Heckman, James J., and Jeffrey A. Smith. "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies." *The Economic Journal*, vol. 109, no. 457, 1999, pp. 313–348.

Heinrich, Carolyn, J., Peter R. Mueser, Kenneth R. Troske, Kyung-Seong Jeon, and Daver C. Kahvecioglu. "Do Public Employment and Training Programs Work?" *IZA Journal of Labor Economics*, vol. 2, no. 1, 2013, pp. 1–23.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, vol. 71, no. 4, 2003, pp. 1161–1189.

Horvitz, Daniel G., and Donovan J. Thompson. "A Generalization of Sampling Without Replacement from a Finite Universe." *Journal of the American Statistical Association*, vol. 47, no. 260, 1952, pp. 663–685.

Huber, Martin, Michael Lechner, and Conny Wunsch. "The Performance of Estimators Based on the Propensity Score." *Journal of Econometrics*, vol. 175, no. 1, 2013, pp. 1–21.

Imbens, Guido W. "Matching Methods in Practice: Three Examples." *The 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.

Imbens, Guido W., and Donald B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press, 2015.

Iraq and Afghanistan Veterans of America. "Veterans' Employment Challenges: Perceptions and Experiences of Transitioning from Military to Civilian Life." Newark, NJ: Prudential Inc., 2012. Available at https://eoa.oregonstate.edu/sites/eoa.oregonstate.edu/files/veterans_employment_challenges.pdf. Accessed June 18, 2019.

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.

Mueser, Peter R., Kenneth R. Troske, and Alexey Gorislavsky. "Using State Administrative Data to Measure Program Performance." *The Review of Economics and Statistics*, vol. 89, no. 4, 2007, pp. 761–783.

Raghunathan, Trivellore E., James M. Lepkowski, John van Hoewyk, and Peter Solenberger. "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models." *Survey Methodology*, vol. 27, no. 1, June 2001, pp. 85–96.

Rubin, Donald B. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley, 1987.

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

Tibshirani, Robert. "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, 1996, pp. 267–288.

Tsai, Jack, and Robert A. Rosenheck. "Risk Factors for Homelessness Among US Veterans." *Epidemiology Review*, vol. 37, no. 1, 2015, pp. 177–195.

U.S. Department of Housing and Urban Development. "HUD 2018 Continuum of Care Homeless Assistance Programs Homeless Populations and Subpopulations." Washington, DC: U.S. Department of Housing and Urban Development, November 13, 2018.

Veterans' Program Letter. "Subject: Enrollment of Homeless Veterans Program Participation into a Workforce Innovation and Opportunity Act (WIOA) Workforce Program at an American Job Center (AJC)." Washington, DC: U.S. Department of Labor, Assistant Secretary for Veterans' Employment and Training, No. 03-16, July 15, 2016.